

ON THE NATURE AND SCOPE OF HABITS AND MODEL-FREE CONTROL

EDITED BY: John A. Bargh, Wendy Wood and David Ellis Melnikoff
PUBLISHED IN: Frontiers in Psychology





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88966-300-2

DOI 10.3389/978-2-88966-300-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

ON THE NATURE AND SCOPE OF HABITS AND MODEL-FREE CONTROL

Topic Editors:

John A. Bargh, Yale University, United States

Wendy Wood, University of Southern California, Los Angeles, United States

David Ellis Melnikoff, Yale University, United States

N.B. This Research Topic was co-developed with David Melnikoff - a junior Topic Editor managing this article collection but not involved in editing manuscripts submitted to this Research Topic.

Citation: Bargh, J. A., Wood, W., Melnikoff, D. E., eds. (2020). On the Nature and Scope of Habits and Model-Free Control. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88966-300-2

Table of Contents

- 04** *Habit and Identity: Behavioral, Cognitive, Affective, and Motivational Facets of an Integrated Self*
Bas Verplanken and Jie Sui
- 15** *Habits, Quick and Easy: Perceived Complexity Moderates the Associations of Contextual Stability and Rewards With Behavioral Automaticity*
Kiran McCloskey and Blair T. Johnson
- 34** *Enhanced Avoidance Habits in Relation to History of Early-Life Stress*
Tara K. Patterson, Michelle G. Craske and Barbara J. Knowlton
- 47** *Expected Value of Control and the Motivational Control of Habitual Action*
Andreas B. Eder and David Dignath
- 59** *Habit Expression and Disruption as a Function of Attention-Deficit/Hyperactivity Disorder Symptomology*
Ahmet O. Ceceli, Giavanna Esposito and Elizabeth Tricomi
- 76** *Binary Theorizing Does Not Account for Action Control*
Bernhard Hommel
- 87** *Model-Based and Model-Free Social Cognition: Investigating the Role of Habit in Social Attitude Formation and Choice*
Leor M. Hackel, Jeffrey J. Berg, Björn R. Lindström and David M. Amodio
- 98** *On How Definitions of Habits Can Complicate Habit Research*
Jan De Houwer
- 107** *Hierarchical Action Control: Adaptive Collaboration Between Actions and Habits*
Bernard W. Balleine and Amir Dezfouli
- 120** *Model-Free RL or Action Sequences?*
Adam Morris and Fiery Cushman
- 142** *A Mobile Phone App for the Generation and Characterization of Motor Habits*
Paula Banca, Daniel McNamee, Thomas Piercy, Qiang Luo and Trevor W. Robbins
- 153** *The Law of Recency: An Episodic Stimulus-Response Retrieval Account of Habit Acquisition*
Carina G. Giesen, James R. Schmidt and Klaus Rothermund
- 170** *Ideomotor Action: Evidence for Automaticity in Learning, but Not Execution*
Dan Sun, Ruud Custers, Hans Marien and Henk Aarts
- 188** *How Sequential Interactive Processing Within Frontostriatal Loops Supports a Continuum of Habitual to Controlled Processing*
Randall C. O'Reilly, Ananta Nair, Jacob L. Russin and Seth A. Herd



Habit and Identity: Behavioral, Cognitive, Affective, and Motivational Facets of an Integrated Self

Bas Verplanken* and Jie Sui

Department of Psychology, University of Bath, Bath, United Kingdom

OPEN ACCESS

Edited by:

John A. Bargh,
Yale University, United States

Reviewed by:

Benjamin Gardner,
King's College London,
United Kingdom
Julian De Freitas,
Harvard University, United States

*Correspondence:

Bas Verplanken
b.verplanken@bath.ac.uk

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 03 February 2019

Accepted: 13 June 2019

Published: 10 July 2019

Citation:

Verplanken B and Sui J (2019)
Habit and Identity: Behavioral,
Cognitive, Affective, and Motivational
Facets of an Integrated Self.
Front. Psychol. 10:1504.
doi: 10.3389/fpsyg.2019.01504

Two studies investigated associations between habits and identity, in particular what people consider as their “true self.” Habit-identity associations were assessed by within-participant correlations between self-reported habit and associated true self ratings of 80 behaviors. The behaviors were instantiations of 10 basic values. In Study 1, significant correlations were observed between individual differences in the strength of habit-identity associations, measures of cognitive self-integration (prioritizing self-relevant information), self-esteem, and an orientation toward an ideal self. Study 2 further tested the assumption that habits are associated with identity if these relate to important goals or values. An experimental manipulation of value affirmation demonstrated that, compared to a control condition, habit-identity associations were stronger if participants explicitly generated the habit and true self ratings while indicating which values the behaviors would serve. Taken together, the results suggest that habits may serve to define who we are, in particular when these are considered in the context of self-related goals or central values. When habits relate to feelings of identity this comes with stronger cognitive self-integration, higher self-esteem, and a striving toward an ideal self. Linking habits to identity may sustain newly formed behaviors and may thus lead to more effective behavior change interventions.

Keywords: habit, identity, integrated self, true self, self-esteem, self-regulatory focus, value affirmation

INTRODUCTION

What determines our identity? A potential source of identity, which has received little attention in the literature on the self and the self-concept, is the array of our habits. A large portion of everyday behavior is habitual, that is, being performed frequently, often automatically, and in stable contexts (e.g., Verplanken and Aarts, 1999; Wood et al., 2002; Gardner, 2015; Verplanken, 2018). Habits vary in a number of ways. One is complexity; some habits involve simple acts, such as nail biting or checking the time, while others are part of more complex behaviors or

routines, such as donating to charity or exercising. Habits also vary in terms of involvement of other people. For instance, taking the car to work is a solitary activity, whereas calling your parents maintains a relationship. And habits vary in the extent to which they are important to us. We may not even be aware of the many *unimportant* habits, such as where you sit at the table or the way you tie your shoes. Other habits are more important, such as those which express an important value. An unanswered question is whether or when habits contribute to what we consider as our identity, and if this is the case, how these sources of identity are embedded in other self-related constructs and processes, such as beliefs about ourselves, self-esteem, and self-regulation.

Personal or self-identities can be considered as mental representation individuals hold about who they are, which include autobiographical memories, self-attributions, beliefs, motivations, recurrent thoughts, emotions, and self-perceptions. These narratives are constantly constructed and revised (e.g., Vignoles, 2011). Habits may become part of self-identities through various psychological processes. One such process may be the end result of enacted motivations, such as suggested in socio-cognitive models (e.g., Fishbein and Ajzen, 1975; Deci and Ryan, 1991; Rise et al., 2010). A strong motivation, anchored in self-identity, may instigate repeated action, which may then become a habit. Such habits may function as vehicles of self-control in accomplishing a goal: habits relieve an individual from having to deliberate and decide on actions and may thus promote the accomplishment of a goal (e.g., Galla and Duckworth, 2015). Another path to a habit-identity relation is through self-perception (e.g., Bem, 1972). Through the perception of our own frequently performed behaviors, we may infer that these are important to us and may thus be part of who we are (e.g., Neal et al., 2012; Wood and Rünger, 2016).

Empirical Evidence for a Habit-Identity Relation

What is the evidence for a habit-identity relationship? Some habits directly signify a particular identity. For instance, while the culture around smoking is rapidly changing, in some population segments this habit still stands for masculinity or being “cool” (e.g., Ng et al., 2007). Self-identity has been studied as a potential addition to the theory of planned behavior. This theory poses an intention to act as the primary determinant of behavior, which in turn is determined by an attitude, normative pressure, and perceived control of the behavior (e.g., Ajzen, 1991). In a meta-analysis on the role of self-identity in the theory of planned behavior, Rise et al. (2010) established that self-identity correlated 0.33 with past behavior, which has often been considered as a proxy for habit. A number of primary studies provided evidence for a habit-identity association. Charng et al. (1988) reported a 0.22 correlation between blood donation habit and a measure of identity as blood donor. Gardner et al. (2012) found a strong correlation between measures of binge drinking habit and binge drinking identity among university students ($r = 0.69$). Gardner and Lally (2013) found a strong correlation

between habit and intrinsic motivation for physical activity ($r = 0.64$). Gatersleben et al. (2014, Study 2) found that measures of environmental and frugal identities mediated between environmental values and pro-environmental behaviors ($\beta_s = 0.35$ and 0.28 , respectively). Lindgren et al. (2015) found a significant correlation between an implicit measure of drinking identity and drinking habit ($r = 0.36$). Verplanken and Roy (2016) found that an index of pro-environmental habits correlated significantly with biospheric values ($r = 0.31$), personal norms ($r = 0.45$), and personal involvement ($r = 0.30$), i.e., constructs that are closely related to self-identity. McCarthy et al. (2017) reported a strong correlation between a measure of health-conscious identity and an assessment of healthy eating habit ($r = 0.69$). Albini et al. (2018) found a significant correlation between personal importance and habit of consuming vegetables ($r = 0.49$). As the relationships mentioned above are correlational, the causal flow in the habit-identity relation is unknown and may well be bi-directional: a particular identity may instigate behavior and thus maintain a habit, while the self-perception of a habit may feed into self-identity (cf., Wood and Rünger, 2016).

Two perspectives outside the social psychological domain may be taken to support a link between habit and identity. The first comes from the area of moral development of the self as a core of personal identity. Developing a self-concept and self-identity comes with the development of a moral identity (e.g., Blasi, 1994). From an early age, we learn to “do the right thing” in a variety of situations. By repeating such moral actions, these may turn into moral habits and feed into a moral identity. Such habits may become what can be designated as “character,” “second nature dispositions,” or indeed, a moral identity (e.g., Aquino and Reed, 2002; Hulseley and Hampson, 2014; Ward and King, 2018). Second, an interesting view on the relationship of habit and identity from a philosophical perspective was put forward by Wagner and Northoff (2014). These authors discussed the difference between “personhood” and “personal identity.” Personhood refers to features that define a person at one specific point in time. However, as such features are fluid and impermanent, and in order to persist as the *same* person, that is, to have a personal identity, features need to remain stable. Wagner and Northoff (2014) thus considered habit as an explanatory construct, which links these different temporal dimensions to form a personal identity.

The empirical basis of a relationship between a habit and a self-identity is not unequivocal. For instance, Murtagh et al. (2012) reported nonsignificant correlations between a measure of identity and measures of past travel mode behaviors (r s varying between 0.02 and 0.07). Also, while in the Albini et al. (2018) study cited above personal importance and habit of consuming vegetables correlated significantly, no such correlation was present for consuming fruit ($r = 0.06$). There was neither evidence of a habit-identity relation in a comprehensive study into the nature of students’ everyday habits conducted by Wood et al. (2002), in which participants were asked to write hourly reports on their ongoing behaviors and experiences. If anything, in this study habits were associated

with negative self-evaluations and the relative *unimportance* of these behaviors for attaining personal goals.

Taken together, the studies and perspectives discussed above lead to two conclusions. The first is that there exists significant, and sometimes substantial, associations between measures of habit and measures of self-identity, and there are some arguments beyond social psychology for such a relationship. Second, such correlations are not being found across the board; there is a large variation between studies in the size of habit-identity correlations. This suggests that certain habits, but not all, relate to self-identity. We contend that prime candidates for such a role are habits that are related to important goals or values. Goals and values may be integrated in one's self-concept and are thus likely to be repeatedly enacted (e.g., Deci and Ryan, 1991; Sheldon and Elliot, 1999; Aarts and Dijksterhuis, 2000; Verplanken and Holland, 2002; Bardi and Schwartz, 2003; Hitlin, 2003; Gatersleben et al., 2014; Burkley et al., 2015). In addition, we anticipate that people differ in the strength of habit-identity associations. First, different people have different habits and may thus associate different habits with their self-identity, which may lead to variation between studies. Second, people may differ in the extent to which they identify habits as being relevant for one's identity in the first place.

The Integrated Self

An emerging theme in the literature on the self is the realization that some parts of the self are more essential than others, which has been referred to as real self (e.g., Rogers, 1961), authentic self (e.g., Koole and Kuhl, 2003; Johnson et al., 2004), or true self (Newman et al., 2014; Strohminger et al., 2017). At the heart of this concept lies the notion of an *integrated self*, that is, a high degree of connectedness within and between cognitive, affective, motivational, and behavioral systems. Kuhl et al. (2015) presented a neurobiological model, which explains the various functional characteristics of the integrated self, such as emotional and somatosensory connectedness, attention to self-relevant information, and self-positivity. The integrated self is holistic and incorporates a vast amount of autobiographical memory. It functions by means of high-level parallel-distributed processing, operating largely at implicit levels, and is thus able to integrate a large amount of self-related processes – cognitive, emotional, motivational, and volitional – simultaneously (Kuhl et al., 2015). We contend that self-perception of behaviors *per se* is not what connects them to the self but that behaviors become part of the integrated self if two conditions are fulfilled. One is that the behavior has become habitual, that is, being repeatedly and automatically executed and has thus become ingrained in the person's autobiographical memory. The second is that a behavior is related to an important goal or value. This is not the case for all habits and for all individuals.

Sui and Humphreys (2015) summarized a body of work that sheds light on properties of an integrated self in more detail at the neuro-cognitive level. As an indicator of the degree to which a person possesses an integrated self, these

researchers used perceptual matching tasks, which assess differences in reaction time and accuracy between matching self-related versus other-related stimuli (Sui et al., 2012). Larger differences indicate stronger self-prioritization effects (i.e., a stronger “self-bias”). Sui and colleagues demonstrated that self-referencing can have wide-ranging integrative effects with respect to perception, attention, memory, and decision making (e.g., Sui et al., 2012), which is thus interpreted as cognitive self-integration. This evidence suggests that self-referencing is not simply a narrative reflecting ongoing self-related processes. Rather, self-referencing actively modulates cognitive processes and acts as a “glue,” which binds different forms of information, for instance, between stimuli in perception and memory, or integrates different stages of information processing, such as in decision making. Sui et al. (2012) argued that self-referencing leads to robust self-prioritization effects in perception and cognition. Sui and Gu (2017) further put forward a neural framework of an integrated self where they argued that cognitive and affective aspects of the self-interact to influence behavior through the three neural networks – the ventral network including the ventral medial prefrontal cortex (vmPFC), the cognitive control, and the salience networks. Researchers have reported that inducing emotional valence can alter self-prioritization in face recognition. For example, when participants are asked to evaluate negative personality traits, there is a reduced advantage for processing self vs. others' faces (Ma and Han, 2010). Consistent with this, the self-prioritization effect in the perceptual matching task was disrupted in individuals with low mood (Sui et al., 2016), due to the breakdown of the integrated self (in this case, the intrinsic association between self and positive emotion) in depressed individuals. In short, the strength of self-prioritization (“self-bias”) observed in these perceptual matching tasks can be considered as a proxy for cognitive self-integration.

At the experiential level, a number of authors describe the “true self,” which arguably is a subjective experience of an integrated self (e.g., Newman et al., 2014; Strohminger and Nichols, 2014; De Freitas et al., 2018). The true self is what a person considers as one's authentic core and is experienced as inherently moral and good. Although the true self is in essence a belief a person holds about oneself and may thus be false or distorted, it has consequences for a person's cognitive and social functioning. For example, it has been reported that unfavorable self-related events are more likely to be forgotten (Hu et al., 2015). People also tend to attribute positive outcomes to themselves relative to other people while linking negative outcomes to others, thus demonstrating biased causal attributions in social evaluation (Greenwald, 1980), or to influence the environment (e.g., Newman et al., 2014). Finally, moral values that make up someone's true self may serve as benchmarks to judge others' moral value status (e.g., Newman et al., 2014). Thus, the true self has the potential to evoke feelings of self-worth and a sense of meaning in life (e.g., Schlegel et al., 2009) and to protect the self from negative perspectives (Sedikides and Green, 2009). Particular habits, then, may be seen as instantiations of the accomplishment

of goals or values associated with the true self and may thus become incorporated in one's self-identity.

The Present Studies

The present studies aimed at investigating the relationships between the degree to which individuals associate habits with their true self and how this relates to cognitive, affective, and motivational aspects of the self. Variation in habit-identity associations was assessed by presenting participants with 80 behaviors, and asking two ratings for each of those behaviors, i.e., self-reported habit and how much the activity reflects their true self. For each participant, a correlation was calculated between these two ratings across the 80 behaviors, which thus served as a measure of habit-identity associations. In Study 1, this association measure was correlated with the measures of cognitive self-integration obtained by the perceptual matching paradigm as developed by Sui et al. (2012). In addition, the study contained assessments of self-esteem as an affective component of the self and chronic self-regulatory focus style (i.e., "promotion" and "prevention"; Higgins, 1998) as a motivational aspect of the self. A promotion style is an orientation toward hopes, aspirations, and your ideal self. A prevention style is an orientation toward safety and responsibilities and fulfills what you think ought to be done. Positive correlations were expected between habit-identity associations, cognitive self-integration, self-esteem, and a promotion-style self-regulatory focus. Study 2 focused in more detail on the habit-identity association measure. This study aimed at demonstrating that habit-identity associations are stronger if these are being generated in the context of goals and values compared to a more concrete context.

STUDY 1

Method

Participants and Procedure

The study was conducted in a laboratory at the authors' university. A power analysis was conducted prior to this study. In a previous study among 67 participants, admittedly older than in the present study, a mid-range correlation of 0.36 ($p < 0.003$) was found between cognitive self-integration in the perceptual matching task used in the present study and a self-report measure of personal distance (Sui and Humphreys, 2017). Together with setting an α of 0.05, two-sided testing, accepting a power of 0.80, and aiming at detecting medium effect size correlations ($r \approx 0.30$), a sample size of approximately 85 was required. A total of 90 participants were recruited from the university's student population. There were 29 males and 61 females. Their mean age was 21 years ($SD = 2.67$). All participants had normal or corrected-to-normal vision. Informed consent was obtained from all participants according to procedures approved by the authors' departmental ethics committee (IRB).

Participants worked individually and visually separated. They first carried out the perceptual matching task, which assessed cognitive self-integration. This was followed by a questionnaire,

which contained the habit and identity ratings and assessments of self-esteem and self-regulatory focus. A session took 30–40 min. Participants were paid £5.00 for their contribution.

Measures

Cognitive Self-Integration

Cognitive self-integration was measured by assessing the strength of self-prioritization ("self-bias") in a perceptual matching task (Sui et al., 2012). Participants were first asked to name one of their best friends. They then selected a gender-matched stranger from a common name list not corresponding to anyone they knew. The named friend and stranger were then used in the perceptual matching task, where they were instructed to associate three geometric shapes (triangle, circle, square) with labels indicating the self ("You"), the named best friend ("Friend"), and the named stranger ("Stranger"), respectively. The assignment of the particular shapes to the three labels was counterbalanced across individuals. The self-prioritization scores were calculated using the performance scores of "You" and "Stranger." The reason "Friend" was included in the task was to make it sufficiently challenging so as to avoid ceiling effects.

After the association instruction, participants conducted the shape/label matching task. Participants were asked to judge whether or not simultaneously presented shape/label pairs (e.g., a circle/"You") matched according to the associations they had been instructed to make. Each trial started with a central fixation cross for 500 ms, followed by a shape/label pair at the center of the screen for 100 ms. A shape (triangle, circle, or square) with $3.5 \times 3.5^\circ$ of visual angle appeared above a white central fixation cross with $0.8 \times 0.8^\circ$ of visual angle. One of three labels ("You," "Friend," or "Stranger") covering $1.76/2.52^\circ \times 1.76^\circ$ of visual angle was displayed below the fixation cross. All stimuli in white were displayed on a gray background. E-prime software version 2.0 was used to present the stimuli and to record responses. The experiment was run on a PC with a 22-in monitor ($1,920 \times 1,080$ pixels) at 60 Hz.

Half of the shape/label pairs conformed to the association instruction and should thus be responded to as "match" trials; on the remaining trials, the shapes and labels were re-paired to form "mismatch" trials. For mismatch trials, a shape was paired with one of the other labels (e.g., a circle/"Stranger," in our example). The next frame was a 1,000 ms blank field. Participants were encouraged to make a "match" or "mismatch" response as quickly and accurately as possible within this 1,000 ms interval by pressing one of two keys on the keyboard with the index or middle finger of the right hand. The order of response keys was counterbalanced across participants. A feedback message ("correct," "incorrect," or "too slow") was then given in the center of the screen for 500 ms. Participants were informed of their overall accuracy at the end of each block. There were three blocks of 60 trials following 12 practice trials. Thus, there were 30 match and 30 mismatch trials in each block.

Self-bias scores were calculated for reaction times (RT) and accuracy, respectively, for correct responses on match

shape-label trials. Only correct responses longer than 200 ms were included. All participants had accuracy scores >0.55 (i.e., 5% or more above chance level). Self-bias on RT was inferred from the difference in RT for the self against the stranger condition, divided by the sum of the two conditions and multiplied by 100 [i.e., $100 \times [(stranger - self)/(self + stranger)]$]. Self-bias on accuracy was indexed by the difference in performance for the self against the stranger condition divided by the sum of the two conditions [i.e., $(self - stranger)/(self + stranger)$]. Larger scores of both measures indicated a stronger self-bias and thus were taken as stronger cognitive self-integration.

Habit-Identity Associations

Participants were presented with 80 behaviors, which were chosen to cover 10 value-related motivation areas (cf., Schwartz, 1992; Bardi and Schwartz, 2003): hedonism (e.g., “Enjoy a movie”), stimulation (e.g., “Do something exciting”), self-direction (e.g., “Find something out by yourself”), universalism (e.g., “Buy ecological products”), benevolence (e.g., “Donate to charity”), conformity (e.g., “Wear what’s in fashion”), tradition (e.g., “Attend family occasions”), security (e.g., “Make sure your door is locked”), power (e.g., “Make your voice be heard”), and achievement (e.g., “Study during the weekend”). Participants were asked to provide two ratings for each of the behaviors. The first rating was the self-reported frequency of performing the behavior (“How frequently do you do this activity”), which was considered as a proxy for habit strength. Responses were given on a 5-point scale ranging from “never” (1) to “always” (5). The second rating concerned the extent to which the behavior reflected participants’ true self. The instruction was to indicate “how much this activity is something that reflects *who you really are* as a person (your “true self”).” Responses were given on a 5-point scale ranging from “not at all” (1) to “very much” (5). For each individual participant, a correlation was calculated between the frequency and true self ratings across the 80 behaviors. These within-participant correlations were considered as a measure of individual differences in habit-identity associations.

Self-Esteem

Self-esteem was assessed by the 10-item Self-Esteem Scale (Rosenberg, 1965). Sample items are “I feel I have a number

of good qualities” and “I wish I could have more respect for myself” (reverse-coded). Responses were given on 5-point scales ranging from “disagree” (1) to “agree” (5). Scores were coded such that higher numbers indicate higher self-esteem. Cronbach’s α was 0.85.

Self-Regulatory Focus

Individual differences in self-regulatory focus were assessed by the 18-item Promotion/Prevention Scale (Lockwood et al., 2002). The scale contains two subscales measuring a promotion and a prevention self-regulatory orientation, respectively. Examples of promotion orientation items are “I frequently imagine how I will achieve my hopes and aspirations” and “My major goal right now is to achieve my ambitions.” Examples of prevention orientation items are “I’m anxious that I will fall short of my responsibilities and obligations” and “My major goal right now is to avoid becoming a failure.” Responses were given on 7-point scales ranging from “not at all true of me” (1) to “very true of me” (7). Scores were coded such that higher numbers indicate a strong promotion or prevention focus. Cronbach’s α s were 0.87 and 0.73 for the promotion and prevention orientation subscales, respectively. The correlation between the two subscales was 0.42, $p < 0.001$. In order to investigate the unique variances of each subscale, uncorrelated factor scores for each subscale from a Varimax rotated factor analysis were used in the further analyses.

Results and Discussion

The within-participant habit-identity correlations ranged from -0.19 to 0.89 , suggesting substantial individual differences in habit-identity associations. The median correlation was 0.46 . In the subsequent analyses, the habit-identity correlations were Fisher-Z transformed, although the results were nearly identical when untransformed correlations were used.

In **Table 1**, means, standard deviations, and correlations between the study variables are presented. In **Figure 1**, the corresponding scatterplots of eight key correlations are shown. The results suggest that the degree to which individuals associated habits with self-identity correlated statistically significantly with both self-bias measures as well as with self-esteem and a promotion self-regulatory orientation. In addition, the self-bias measures correlated statistically significantly with self-esteem and a promotion orientation.

TABLE 1 | Means, standard deviations, and correlations (Study 1).

Variable	<i>M</i>	<i>SD</i>	2	3	4	5	6
1. Habit-identity associations ¹	0.46	0.36	0.34***	0.34***	0.45***	0.68***	0.06
2. Self-bias on RT	6.23	5.34		0.51***	0.22*	0.40***	0.09
3. Self-bias on accuracy	0.11	0.12			0.21*	0.41***	0.05
4. Self-esteem	3.60	0.71				0.52***	-0.34***
5. Promotion orientation ²	0.00	1.00					0.00
6. Prevention orientation ²	0.00	1.00					

Note: $N = 90$. * = $p < 0.05$; *** = $p < 0.001$.

¹Within-participant Fisher-Z transformed correlations.

²Factor scores from a Varimax rotated solution. The means and standard deviations of the promotion and prevention raw scores were 5.13 (1.01) and 4.39 (0.88), respectively.

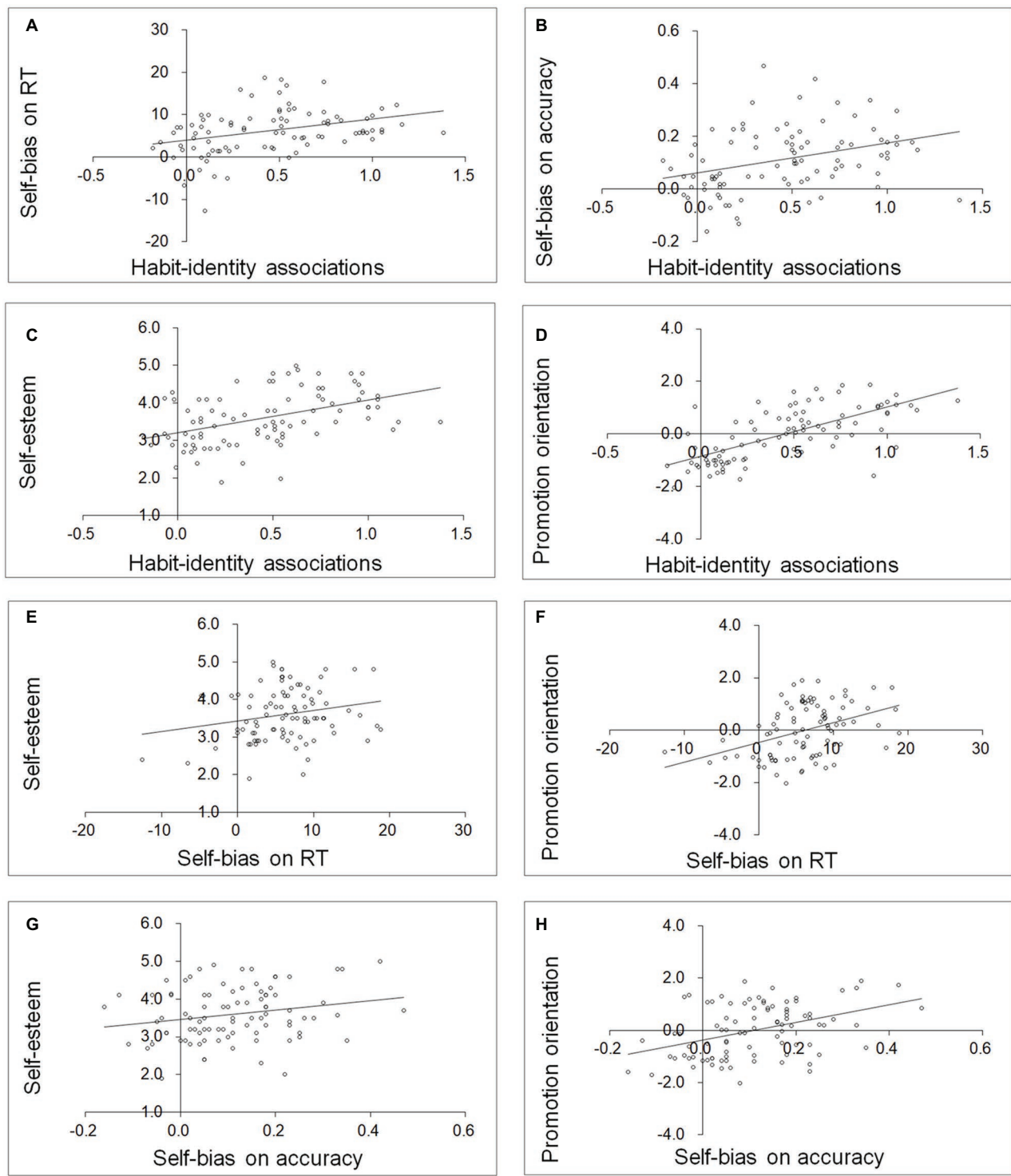


FIGURE 1 | (A–H) Scatterplots of key correlations in Study 1.

Feelings of identity derived from habits were found associated with cognitive, affective, and motivational facets of the self. The pattern of correlations suggests that individuals for whom habits are strongly related to feelings of identity

show stronger cognitive self-integration, higher self-esteem, and a stronger striving toward an ideal self. Note that the obtained correlations were between three very different types of data, that is, within-participant habit-identity correlations,

latency/accuracy data, and self-assessments, respectively, which speaks against inflated correlations due to consistency and social desirability biases.

STUDY 2

The assumption in Study 1 was that habits are implied in feelings of identity if these relate to important goals or values. Study 2 aimed to test that assumption. We contend that habit-identity associations are stronger if participants affirm the values that are perceived to be related to the respective habits. The habit-identity association task, which was used in Study 1, was thus presented under two conditions¹. In a value affirmation condition, participants were asked for each of the 80 behaviors to indicate *why* they would do the activity, in addition to the habit and true self ratings. They could choose between 10 values, which represented the motivational continuum of Schwartz's (1992) value circumplex. Participants in the control condition indicated for each activity at *which time* of the day they would likely engage in the activity and could choose between 10 specified times. The expectation was that the within-participant correlations between the habit and true self ratings would be stronger in the value affirmation versus control condition. The rationale was that value affirmation would enhance the salience of goals participants adhered to, which would thus lead to higher importance ratings.

Method

Participants and Design

The study was conducted online *via* Prolific Academic, which is a UK-based platform for online studies. A power analysis was conducted prior to this study. As there are no previous studies that could serve as a benchmark, we aimed at being able to detect a small effect size in a two-sided *t* test between two independent samples (Cohen's $d \approx 0.25$), setting an α of 0.05, and accepting a power of 0.80. The sample size needed for this setup was approximately 500. A total of 500 participants were recruited, 482 of which completed the study. All participants were students. There were 307 males and 173 females, while two participants did not indicate a gender. Their mean age was 22 years ($SD = 3.07$). Informed consent was obtained from all participants according to procedures approved by the departmental ethics committee (IRB). Participants were randomly allocated to a value affirmation versus control condition. The task took 15–20 min to complete. Participants were paid £2.25 for their contribution.

Materials

The habit-identity association task contained the same 80 behaviors that were used in Study 1. As an explanation of habit ratings, participants were told: "How much of a habit is this activity for you? A habit is something you do frequently and automatically." The ratings were then introduced as "When you have the

opportunity, how frequently and automatically do you do this?" Responses were given on a 5-point scale ranging from "never" (1) to "always" (5). As an explanation of true self ratings, participants were told: "How much does the activity reflect *who you really are* as a person? That is, to what extent does the activity represent what you would consider as your 'true self.'" The identity ratings were then introduced as: "How much does this activity reflect your true self?" Responses were given on a 5-point scale ranging from "not at all" (1) to "very much" (5). In between each habit and identity rating, participants in the value affirmation condition were asked to choose from a pull-down menu *why* they would do the activity ("If you would do this, why?"). They were presented with 10 value areas (Schwartz, 1992), which were briefly explained: "Influence (social status and prestige, control over people and resources)"; "Achievement (personal success, competence, meeting high standards)"; "Pleasure (enjoyment, sensual gratification, indulgence)"; "Excitement (adventure, novelty, seeking challenges, exploring)"; "Independence (seeking freedom, independence, uniqueness, creativity)"; "Welfare (understanding, tolerance, welfare of people and nature)"; "Helpful (helping people you meet or are in frequent contact with)"; "Tradition (respect, commitment, acceptance of customs from culture or religion)"; "Conformity (abiding by the rules, meeting others' expectations, respecting norms)"; "Security (safety, harmony, stability for yourself, others, and the community at large)."² The value labels and their descriptions were presented on an instruction page, while the pull-down menu contained the 10 value labels. In the control condition, participants were also presented with a pull-down menu but were asked to indicate *when* they would do the activity ("If you would do this, at what time would this typically occur?"). They could select one of the following 10 times: 7 AM, 9 AM, 11 AM, 1 PM, 3 PM, 5 PM, 7 PM, 9 PM, 11 PM, and 1 AM.

The validity of the value affirmation manipulation was tested in an online study among 93 participants. There were 38 males and 55 females, while two participants did not indicate a gender. Their mean age was 27 years ($SD = 8.28$). Informed consent was obtained from all participants according to procedures approved by the authors' departmental ethics committee (IRB). Participants were presented with a random selection of 25 from the 80 behaviors and were randomly assigned to the value affirmation or control condition described above. For each behavior, they were asked how important this activity would be for them on a 6-point scale ranging from "not at all" (1) to "very much" (6). The 25 ratings were averaged. Participants in the value affirmation condition indeed gave higher importance ratings than participants in the control condition, M -value affirmation = 4.02, control = 3.73, $t(91) = 2.25$, $p < 0.03$, Cohen's $d = 0.47$. This supported the validity of the value affirmation manipulation.

²We used "frequently and automatically" in Study 2 instead of "frequently" in Study 1, because, on reflection, the former is more aligned with contemporary conceptions of habit (e.g., Verplanken and Orbell, 2003; Gardner, 2015).

³Some of the labels for the value areas were slightly adapted from the original labels Schwartz (1992) presented, as some of the latter were found too abstract for the purpose of this study (e.g., we used "helpful" instead of "benevolence", and "influence" instead of "power").

¹None of the other assessments in Study 1 were included in this study.

Results and Discussion

The within-participant habit-identity correlations in this sample ranged from -0.21 to 0.99 . The median correlation was 0.69 . The median correlation was 0.71 in the value affirmation condition and 0.65 in the control condition. A t test was conducted after a Fisher- Z transformation of the correlations. The difference between the two conditions was statistically significant, $t(480) = 2.34$, $p < 0.02$, Cohen's $d = 0.21$. The results were nearly identical when untransformed scores were used, $t(480) = 2.58$, $p < 0.01$, Cohen's $d = 0.23$. This result provides proof of concept and suggests that habit-identity associations are stronger if habits are linked to value-based motivations.

GENERAL DISCUSSION

As we argued in the introduction, habits are not *necessarily* associated with identity. Individuals differ in which habits they develop, and thus in which habits, if any, make up part of their self-identity. Incidentally, we do not wish to argue that non-habitual behaviors cannot be part of someone's self-identity. Our assumption was that some habits may be more prone to relate to feelings of identity than others, namely those habits that are instantiations of chronic goals or values. In the present studies, habits were selected that were inferred from basic value domains (Schwartz, 1992). As values are inherently motivational forces, those habits are more likely to be associated with value-related goals and have a higher likelihood to be central to the self and feelings of self-identity (e.g., Verplanken and Holland, 2002). The variation in the habit-identity association measure used in both studies demonstrated that there are individual differences in the degree to which people associate habits with self-identity. In Study 1, this variation correlated with cognitive self-integration (self-prioritization), self-esteem, and a promotion-style self-regulatory focus. Study 2 demonstrated that habit-identity associations are stronger when these are explicitly considered as instantiations of values, which corroborate the assumption that value-related habits are implied in feelings of self-identity.

The correlations found in Study 1 are consistent with integrated self frameworks as suggested by Kuhl et al. (2015) and Sui and Gu (2017), which stress the interactions between cognitive, affective, and motivational aspects of the self for control of behavior. The correlations with habit-identity associations suggest that perceiving oneself to do things that fulfill important goals may be part of such a network and may thus add to feelings of self-worth and represent strivings toward an "ideal self." The latter may also be a source of positive emotional experiences, as positive emotions and higher self-esteem are consequences of successful promotion-oriented self-regulation (e.g., Higgins, 1998). Consistent with this, breaking down intrinsic associations between self and positivity leads to reduced performance in self-recognition (Ma and Han, 2010), and negative mood induces a decreased self-prioritization effect in perception (Sui et al., 2016). Kuhl et al. (2015) considered self-positivity and inner security as one of the functional

characteristics of the integrated self. The positive relations found in Study 1 may thus point to what Rogers (1961) described as characterizing "a fully functioning person," that is, someone who aims at fulfilling their full potential. While the individual components that were included in this study are interesting in their own right, the apparent relationships between these different pieces of data suggest such a more holistic integrative structure. Self-perception of habits and associated feelings of identity may thus play a role in this system, at least to the extent to which an integrated self has been developed. It should be noted though that a strong integrated self is not *necessarily* positive or wholesome but may also characterize individuals who are highly delusional or be associated with narcissism and self-aggrandizing. But in those individuals too, self-perception of habits may function to support such beliefs.

An important question is what exactly the underlying mechanisms are of an integrated self. In other words, what are the dynamics that govern the relationships between behavioral, cognitive, affective, and motivational facets of an integrated self? The correlational data of Study 1, while demonstrating relations between these entities, leave unanswered questions of causality. For instance, do stronger habit-identity associations contribute to stronger cognitive self-integration and positive self-feelings, or do individuals with a strong integrated self and high self-esteem become more attentive to what they are doing to fulfill their ideal self? A promising approach to model these relationships is provided by control-process models, which describe how individuals self-regulate in terms of behavior, cognition, affect, and motivation (e.g., Carver and Scheier, 1998; Vohs and Baumeister, 2017). While elaboration on these models is beyond the scope of this article, they describe processes that unfold when individuals experience discrepancies between a current state and a goal. Moral values that make up part of one's true self may constitute such goals. If and when the self is activated, habits may fulfill different roles in a control-process model, for instance, as a way to lower the perceived discrepancy between a current state and a goal and thus generate positive affect. Habits may also function as a standard against which goal fulfillment is evaluated, which may lead to positive or negative feelings, depending on the outcome of such an operation. Another possible role of habits is a mechanism for the mind to prioritize the action from a range of options, which would lead to goal fulfillment (e.g., Verplanken et al., 1994).

In both studies, we correlated participants' habit ratings with the degree to which they perceived these behaviors to be part of their true self. While the true self is experienced as highly personal and is fundamental to who a person thinks they are (e.g., De Freitas et al., 2018), the content of the moral beliefs, which underlie the true self, are strongly anchored in the culture the person belongs to (e.g., De Freitas et al., 2017). This makes the true self an inherently social construct. A specific habit (e.g., helping an elderly person) may thus constitute a course of action by which a culturally determined moral value (benevolence) is expressed. Habits that are strongly associated with moral values may thus function as benchmarks to evaluate not only oneself but also to make inferences, and indeed, judgments,

about other people's personality, mental state, or behavior (e.g., Newman et al., 2014, 2015; De Freitas et al., 2018).

A limitation of the present studies is that, for the obvious reason of avoiding an overload for participants, the habit and identity assessments for the 80 behaviors had to be confined to one-item measures, while for psychometric reasons, this is not ideal. A related, and arguably more fundamental, limitation is that the one-item measures of behavioral frequency leave room for the argument that we measured frequent, repetitive, or familiar behaviors, which may or may not be habitual according to the contemporary definitions of habit. This has been salvaged somewhat in Study 2 by assessing how "frequent and automatic" the behaviors were executed (but see Gardner and Tang, 2014). While we acknowledge this limitation, it has been demonstrated in numerous studies that used the Self-Report Habit Index (Verplanken and Orbell, 2003), which contains items assessing the experience of repetition as well as automaticity, that these two components are strongly correlated.

As a corollary, the present study contributes to a discussion with respect to the Self-Report Habit Index (SRHI; Verplanken and Orbell, 2003). One of the 12 items of this scale refers to self-identity ("Behavior X is something that is typically me"). It has been debated whether this item should be part of a self-assessment of habit (e.g., Gardner et al., 2012; Rebar et al., 2018). Apart from the fact that this item consistently shows high item-total correlations with the scale, the present findings support the validity of the item as part of the SRHI.

Insight into the relationship between habit and identity may have important implications for behavior change interventions, in particular the longevity of a change if an intervention is successful. Two conditions may have to be fulfilled for behavior change to be maintained over time. The first is to turn new behavior into a habit, that is, behavior that is executed frequently and automatically (e.g., Rothman et al., 2009; Walker et al., 2015; Gardner and Lally, 2018). But second, long-term behavior maintenance may be enhanced if a habit becomes part of an individual's self-identity. For instance, West (2006) posits that self-identity can be a major driver of behavior change and, importantly, the maintenance of newly acquired behavior (e.g., Tombor et al., 2015). The present studies may thus point to an exciting new direction in designing more effective behavior change

interventions, namely not only changing behavior per se but also turning new behavior into habits that are embedded in a self-identity context, and thus capitalize on an integrated self framework.

Conclusion

Some habits serve a self-identifying purpose, in particular when these are considered in the context of self-related goals or central values. The self may function as a subjective center of gravity, involving cognitive, affective, motivational, and behavioral facets (e.g., Sui, 2016). The strength of this "gravitational force" differs between individuals. For some, the self seems a relatively loosely assembled structure, whereas for others, it has a much stronger coherence. The present studies suggest that for the latter type of individuals habits may play a role in this structure and thus make up part of one's self-identity.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the British Psychological Society, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Departmental Ethics Committee, Department of Psychology, University of Bath (reference numbers #17-266 and #18-235 for Studies 1 and 2, respectively).

AUTHOR CONTRIBUTIONS

BV and JS contributed equally to the research and manuscript.

ACKNOWLEDGMENTS

The authors thank Anna Gladwin, Viknesh Jeevachandran, and Imogen Ormston for their contributions to programming and data collection and Eve Legrand and Greg Maio for their insightful comments.

REFERENCES

- Aarts, H., and Dijksterhuis, A. (2000). Habits as knowledge structures: automaticity in goal-directed behavior. *J. Pers. Soc. Psychol.* 78, 53–63. doi: 10.1037/0022-3514.78.1.53
- Ajzen, I. (1991). The theory of planned behavior. *Organ. Behav. Hum. Decis. Process.* 50, 179–211. doi: 10.1016/0749-5978(91)90020-T
- Albini, V., Butler, L. T., Traill, W. B., and Kennedy, O. B. (2018). Understanding fruit and vegetable consumption in children and adolescents: the contributions of affect, self-concept, and habit strength. *Appetite* 120, 398–408. doi: 10.1016/j.appet.2017.09.018
- Aquino, K., and Reed, A. II (2002). The self-importance of moral identity. *J. Pers. Soc. Psychol.* 83, 1423–1440. doi: 10.1037/0022-3514.83.6.1423
- Bardi, A., and Schwartz, S. H. (2003). Values and behavior: strength and structural relations. *Personal. Soc. Psychol. Bull.* 29, 1207–1220. doi: 10.1177/0146167203254602
- Bem, D. J. (1972). Self-perception theory. *Adv. Exp. Soc. Psychol.* 6, 1–62. doi: 10.1016/S0065-2601(08)60024-6
- Blasi, A. (1994). "Moral identity: its role in moral functioning" in *Fundamental research in moral development*. ed. B. Puka (New York: Garland Publishing), 168–179.
- Burkley, E., Curtis, J., Burkley, M., and Hatvany, T. (2015). Goal fusion: the integration of goals within the self-concept. *Self Identity* 14, 348–368. doi: 10.1080/15298868.2014.1000959
- Carver, C. S., and Scheier, M. (1998). *On the self-regulation of behavior*. (New York: Cambridge University Press).
- Charng, H.-W., Palliavin, J. A., and Callero, P. L. (1988). Role identity and reasoned action in the prediction of repeated behavior. *Soc. Psychol. Q.* 51, 303–317. doi: 10.2307/2786758
- De Freitas, J., Cikera, M., Grossmann, I., and Schlegel, R. (2017). Origins of the belief in good true selves. *Trends Cogn. Sci.* 21, 634–636. doi: 10.1016/j.tics.2017.05.009

- De Freitas, J., Cikera, M., Grossmann, I., and Schlegel, R. (2018). Moral goodness is the essence of personal identity. *Trends Cogn. Sci.* 22, 739–740. doi: 10.1016/j.tics.2018.05.006
- Deci, E. L., and Ryan, R. M. (1991). “A motivational approach to self: integration in personality” in *Nebraska symposium on motivation: Perspectives on motivation*. Vol. 38, ed. R. Dienstbier (Lincoln: University of Nebraska Press), 237–288.
- Fishbein, M., and Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. (Reading, MA: Addison-Wesley).
- Galla, B. M., and Duckworth, A. L. (2015). More than resisting temptation: beneficial habits mediate the relationship between self-control and positive life outcomes. *J. Pers. Soc. Psychol.* 109, 508–525. doi: 10.1037/pspp0000026
- Gardner, B. (2015). A review and analysis of the use of ‘habit’ in understanding, predicting and influencing health-related behavior. *Health Psychol. Rev.* 9, 277–295. doi: 10.1080/17437199.2013.876238
- Gardner, B., de Bruijn, G.-J., and Lally, P. (2012). Habit, identity, and repetitive action: a prospective study of binge-drinking in UK students. *Br. J. Health Psychol.* 17, 565–581. doi: 10.1111/j.2044-8287.2011.02056.x
- Gardner, B., and Lally, P. (2013). Does intrinsic motivation strengthen physical activity habit? Modeling relationships between self-determination, past behaviour, and habit strength. *J. Behav. Med.* 36, 488–497. doi: 10.1007/s10865-012-9442-0
- Gardner, B., and Lally, P. (2018). “Modelling habit formation and its determinants” in *The psychology of habit: Theory, mechanisms, change, and contexts*. ed. B. Verplanken (Cham, Switzerland: Springer), 207–229.
- Gardner, B., and Tang, V. (2014). Reflecting on non-reflective action: an exploratory think-aloud study of self-report habit measures. *Br. J. Health Psychol.* 19, 258–273. doi: 10.1111/bjhp.12060
- Gatersleben, B., Murtagh, N., and Abrahamse, W. (2014). Values, identity and pro-environmental behaviour. *Contemp. Soc. Sci.* 9, 374–392. doi: 10.1080/21582041.2012.682086
- Greenwald, A. G. (1980). The totalitarian ego: fabrication and revision of personal history. *Am. Psychol.* 35, 603–618. doi: 10.1037/0003-066X.35.7.603
- Higgins, E. T. (1998). Promotion and prevention: regulatory focus as a motivational principle. *Adv. Exp. Soc. Psychol.* 30, 1–46. doi: 10.1016/S0065-2601(08)60381-0
- Hitlin, S. (2003). Values as the core of personal identity: drawing links between two theories of self. *Soc. Psychol. Q.* 66, 118–137. doi: 10.2307/1519843
- Hu, X., Bergström, Z. M., Bodenhausen, G. V., and Rosenfeld, J. P. (2015). Suppressing unwanted autobiographical memories reduces their automatic influences: evidence from electrophysiology and an implicit autobiographical memory test. *Psychol. Sci.* 26, 1098–1106. doi: 10.1177/0956797615575734
- Hulsey, T. L., and Hampson, P. J. (2014). Moral expertise. *New Ideas Psychol.* 34, 1–11. doi: 10.1016/j.newideapsych.2014.02.001
- Johnson, J. T., Robinson, M. D., and Mitchell, E. B. (2004). Inferences about the authentic self: when do actions say more than mental states? *J. Pers. Soc. Psychol.* 87, 615–630. doi: 10.1037/0022-3514.87.5.615
- Koole, S. L., and Kuhl, J. (2003). In search of the real self: a functional perspective on optimal self-esteem and authenticity. *Psychol. Inq.* 14, 43–48. doi: 10.1207/S15327965PLI1401_02
- Kuhl, J., Quirin, M., and Koole, S. L. (2015). Being someone: the integrated self as a neuropsychological system. *Soc. Personal. Psychol. Compass* 9, 115–132. doi: 10.1111/spc3.12162
- Lindgren, K. P., Neighbors, C., Teachman, B. A., Gasser, M. L., Kaysen, D., Norris, J., et al. (2015). Habit doesn’t make the predictions stronger: implicit alcohol associations and habitualness predict drinking uniquely. *Addict. Behav.* 45, 139–145. doi: 10.1016/j.addbeh.2015.01.003
- Lockwood, P., Jordan, C. H., and Kunda, Z. (2002). Motivation by positive or negative role models: regulatory focus determines who will best inspire us. *J. Pers. Soc. Psychol.* 83, 854–864. doi: 10.1037/0022-3514.83.4.854
- Ma, Y., and Han, S. (2010). Why we respond faster to the self than to others? An implicit positive association theory of self-advantage during implicit face recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 619–633. doi: 10.1037/a0015797
- McCarthy, M. B., Collins, A. M., Flaherty, S. J., and McCarthy, S. N. (2017). Healthy eating habit: a role for goals, identity, and self-control? *Psychol. Mark.* 34, 772–785. doi: 10.1002/mar.21021
- Murtagh, N., Gatersleben, B., and Uzzell, D. (2012). Self-identity threat and resistance to change: evidence from regular travel behaviour. *J. Environ. Psychol.* 32, 318–326. doi: 10.1016/j.jenvp.2012.05.008
- Neal, D., Wood, W., Labrecque, J. S., and Lally, P. (2012). How do habits guide behavior? Perceived and actual triggers in daily life. *J. Exp. Soc. Psychol.* 48, 492–498. doi: 10.1016/j.jesp.2011.10.011
- Newman, G. E., Bloom, P., and Knobe, J. (2014). Value judgments and the true self. *Personal. Soc. Psychol. Bull.* 40, 203–216. doi: 10.1177/0146167213508791
- Newman, G. E., De Freitas, J., and Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cogn. Sci.* 39, 96–125. doi: 10.1111/cogs.12134
- Ng, N., Weinehall, L., and Öhman, A. (2007). ‘If I don’t smoke, I’m not a man’: Indonesian teenage boys’ views about smoking. *Health Educ. Res.* 22, 794–804. doi: 10.1093/her/cyl104
- Rebar, A. L., Gardner, B., Rhodes, R. E., and Verplanken, B. (2018). “The measurement of habit” in *The psychology of habit: Theory, mechanisms, change, and contexts*. ed. B. Verplanken (Cham, Switzerland: Springer), 31–49.
- Rise, J., Sheeran, P., and Hukkelberg, S. (2010). The role of self-identity in the theory of planned behavior: a meta-analysis. *J. Appl. Soc. Psychol.* 40, 1085–1105. doi: 10.1111/j.1559-1816.2010.00611.x
- Rogers, C. R. (1961). *On becoming a person: A therapist’s view of psychology*. (Boston, MA: Houghton Mifflin).
- Rosenberg, M. (1965). *Society and adolescent self-image*. (Princeton, NJ: Princeton University Press).
- Rothman, A. J., Sheeran, P., and Wood, W. (2009). Reflective and automatic processes in the initiation and maintenance of dietary change. *Ann. Behav. Med.* 38(Suppl. 1), S4–S17. doi: 10.1007/s12160-009-9188-3
- Schlegel, R. J., Hicks, J. A., Arndt, J., and King, L. A. (2009). Thine own self: true self-concept accessibility and meaning in life. *J. Pers. Soc. Psychol.* 96, 473–490. doi: 10.1037/a0014060
- Schwartz, S. H. (1992). Universals in the content and structure of values: theory and empirical tests in 20 countries. *Adv. Exp. Soc. Psychol.* 25, 1–65. doi: 10.1016/S0065-2601(08)60281-6
- Sedikides, C., and Green, J. D. (2009). Memory as a self-protective mechanism. *Soc. Personal. Psychol. Compass* 3, 1055–1068. doi: 10.1111/j.1751-9004.2009.00220.x
- Sheldon, K. M., and Elliot, A. J. (1999). Goal striving, need satisfaction, and longitudinal well-being: the self-concordance model. *J. Pers. Soc. Psychol.* 76, 482–497. doi: 10.1037/0022-3514.76.3.482
- Strohming, N., Knobe, J., and Newman, G. (2017). The true self: a psychological concept distinct from the self. *Perspect. Psychol. Sci.* 12, 551–560. doi: 10.1177/1745691616689495
- Strohming, N., and Nichols, S. (2014). The essential moral self. *Cognition* 131, 159–171. doi: 10.1016/j.cognition.2013.12.005
- Sui, J. (2016). Self-reference acts as a golden thread in binding. *Trends Cogn. Sci.* 20, 482–483. doi: 10.1016/j.tics.2016.04.005
- Sui, J., and Gu, X. (2017). Self as object: emerging trends in self research. *Trends Neurosci.* 40, 643–653. doi: 10.1016/j.tins.2017.09.002
- Sui, J., He, X., and Humphreys, G. W. (2012). Perceptual effects of social salience: evidence from self-prioritization effects on perceptual matching. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 1105–1117. doi: 10.1037/a0029792
- Sui, J., and Humphreys, G. W. (2015). The integrative self: how self-reference integrates perception and memory. *Trends Cogn. Sci.* 19, 719–728. doi: 10.1016/j.tics.2015.08.015
- Sui, J., and Humphreys, G. W. (2017). Aging enhances cognitive biases to friends but not the self. *Psychon. Bull. Rev.* 24, 2021–2030. doi: 10.3758/s13423-017-1264-1
- Sui, J., Ohrling, E., and Humphreys, G. W. (2016). Negative music disrupts self and reward biases in perception. *Q. J. Exp. Psychol.* 69, 1438–1448. doi: 10.1080/17470218.2015.1122069
- Tombor, I., Shahab, L., Herbec, A., Neale, J., Michie, S., and West, R. (2015). Smoker identity and its potential role in young adults’ smoking behavior: a meta-ethnography. *Health Psychol.* 34, 992–1003. doi: 10.1037/hea0000191
- Verplanken, B. (ed.) (2018). *The psychology of habit: Theory, mechanisms, change, and contexts*. (Cham, Switzerland: Springer).
- Verplanken, B., and Aarts, H. A. G. (1999). Habit, attitude, and planned behaviour: is habit an empty construct or an interesting case of automaticity? *Eur. Rev. Soc. Psychol.* 10, 101–134. doi: 10.1080/14792779943000035
- Verplanken, B., Aarts, H., van Knippenberg, A., and van Knippenberg, C. (1994). Attitude versus general habit: antecedents of travel mode choice. *J. Appl. Soc. Psychol.* 24, 285–300. doi: 10.1111/j.1559-1816.1994.tb00583.x

- Verplanken, B., and Holland, R. W. (2002). Motivated decision-making: effects of activation and self-centrality of values on choices and behavior. *J. Pers. Soc. Psychol.* 82, 434–447. doi: 10.1037/0022-3514.82.3.434
- Verplanken, B., and Orbell, S. (2003). Reflections on past behavior: a self-report index of habit strength. *J. Appl. Soc. Psychol.* 33, 1313–1330. doi: 10.1111/j.1559-1816.2003.tb01951.x
- Verplanken, B., and Roy, D. (2016). Empowering interventions to promote sustainable lifestyles: testing the habit discontinuity hypothesis in a field experiment. *J. Environ. Psychol.* 45, 127–134. doi: 10.1016/j.jenvp.2015.11.008
- Vignoles, V. L. (2011). “Identity motives” in *Handbook of identity. Theory and research*. eds. S. J. Schwartz, K. Luyckx, and V. L. Vignoles (New York: Springer), 403–432.
- Vohs, K. D., and Baumeister, R. F. (eds.) (2017). *Handbook of self-regulation*. 3rd Edn. (New York: Guilford).
- Wagner, N.-F., and Northoff, G. (2014). Habits: bridging the gap between personhood and personal identity. *Front. Hum. Neurosci.* 8:330. doi: 10.3389/fnhum.2014.00330
- Walker, I., Thomas, G. O., and Verplanken, B. (2015). Old habits die hard: travel habit formation and decay during an office relocation. *Environ. Behav.* 47, 1089–1106. doi: 10.1177/0013916514549619
- Ward, S. J., and King, L. A. (2018). Moral self-regulation, moral identity, and religiosity. *J. Pers. Soc. Psychol.* 115, 495–525. doi: 10.1037/pspp0000207
- West, R. (2006). *Theory of addiction*. (Oxford: Blackwell).
- Wood, W., Quinn, J. M., and Kashy, D. A. (2002). Habits in everyday life: thought, emotion, and action. *J. Pers. Soc. Psychol.* 83, 1281–1297. doi: 10.1037/0022-3514.83.6.1281
- Wood, W., and Rünger, D. (2016). Psychology of habit. *Annu. Rev. Psychol.* 67, 11.1–11.26. doi: 10.1146/annurev-psych-122414-033417

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Verplanken and Sui. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Habits, Quick and Easy: Perceived Complexity Moderates the Associations of Contextual Stability and Rewards With Behavioral Automaticity

Kiran McCloskey and Blair T. Johnson*

Institute for Collaboration on Health, Intervention, and Policy, Department of Psychological Sciences, University of Connecticut, Storrs, CT, United States

OPEN ACCESS

Edited by:

John A. Bargh,
Yale University, United States

Reviewed by:

Mark Conner,
University of Leeds, United Kingdom
Paschal Sheeran,
The University of North Carolina
at Chapel Hill, United States

*Correspondence:

Blair T. Johnson
blair.t.johnson@uconn.edu

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 14 March 2019

Accepted: 19 June 2019

Published: 24 July 2019

Citation:

McCloskey K and Johnson BT
(2019) Habits, Quick and Easy:
Perceived Complexity Moderates
the Associations of Contextual
Stability and Rewards With Behavioral
Automaticity.
Front. Psychol. 10:1556.
doi: 10.3389/fpsyg.2019.01556

Background: Habits have been proposed to develop as a function of the extent to which a behavior is rewarded, performed frequently, and executed in a stable context. The present study examines how each of these factors are associated with behavioral automaticity across a broad variety of behaviors drawn from previous habits research. This study further assesses how perceived complexity of the behavior influences the associations of rewards, frequency, and contextual stability with automaticity.

Methods: Participants ($N = 459$) completed an online survey assessing their experiences and engagement with 25 different behaviors, including exercise, handwashing, smoking, and medication adherence, among others. Exploratory factor analysis validated a short, relatively novel scale of perceived behavioral complexity, and multilevel analyses grouped by participant were used to examine the factors that contribute to automaticity.

Results: Across behaviors, frequency, contextual stability, and perceived rewards were positively associated with automaticity. Perceived complexity was negatively associated with automaticity and moderated the influence of contextual stability and rewards, but not frequency, on automaticity. Both contextual stability and rewards were stronger predictors of automaticity when behavioral complexity was high rather than low, as predicted; in addition, when contextual stability was high, more complex behaviors showed *greater* automaticity than simpler behaviors.

Conclusion: The results of this study confirm that behavioral frequency, rewards, and contextual stability are each independently associated with automaticity across a spectrum of behaviors. This study further demonstrates that perceived complexity of a behavior moderates the extent to which contextual stability and rewards are associated with automaticity. The results affirm a need to further understand the components of habits and how they differ across varying behaviors.

Keywords: automaticity, behavior, behavioral automaticity, habit, habit strength

INTRODUCTION

As people go through their days, they execute thousands of behaviors. Some behaviors may be complex, such as going to the gym in the morning, and other behaviors may be simple, such as shutting off the lights before one leaves the house. Some behaviors may promote health; others may harm it. As behavior has important consequences for individuals' life outcomes, impacting numerous domains such as health, career, and relationships, a large body of literature aimed at predicting behavior has developed. Perspectives such as the Theory of Planned Behavior (TPB) posit that behavior is the direct result of intention, and thus strive to uncover the factors that motivate individuals to engage in particular behaviors (Fishbein and Ajzen, 1975). Other approaches aim to understand the automatic influences that drive behavior regardless of an individual's intentions. One particular approach focuses on the influence of *habits*. Habits are behaviors that are performed repeatedly and with little preceding forethought (Ouellette and Wood, 1998). As about 45% of people's behavior might qualify as habitual (Neal et al., 2006), understanding habits is an important direction for behavior research.

In psychology, habits might be understood as impulses toward a behavior that are generated automatically in response to an environmental cue from a context in which that behavior has previously been repeatedly executed (Lally and Gardner, 2013), or as the dominant responses that are mentally accessible in the presence of such an environmental cue (Wood and Neal, 2009). The concept of *habit* has been applied to predict diverse behaviors such as recycling, seafood consumption, consumer behaviors, 'cyber loafing' at work, use of information technology, exercise, and even negative thinking (Low, 2016). In a meta-analysis of 72 studies of exercise behavior, Hagger et al. (2002) showed that including past behavior explained 19% of the variance in later behavior over and above the variance accounted for by TPB variables. A second meta-analysis examined a broad spectrum of behaviors and found that past behavior explained additional variance after accounting for TPB variables: 3.4% for dietary behaviors, 10.3% for physical activity behaviors, 11.4% for abstinence behaviors, and 25.3% for health-risk behaviors (McEachan et al., 2011). In fact, when including past behavior in the model, past behavior was the *only* significant predictor of health-risk behaviors. Thus, understanding the mechanisms whereby past behavior predicts future behavior is key to understanding the determinants of many important behaviors.

Three major 'ingredients' have been proposed to be associated with habit formation: contextual stability, behavioral frequency, and rewards (Wood and Neal, 2016). Habits are environmentally linked, such that a cue in the environment automatically triggers an impulse toward a behavioral tendency (Wood, 2017). When a behavior is performed regularly in a stable context, the individual is more likely to encounter consistent cues that can form the basis for a context-behavior association. As frequency of this behavior increases, so too can the strength of the context-behavior association (Wood and Neal, 2009). Rewards – either intrinsic or extrinsic – may contribute to this process by encouraging behavioral repetition (Wood and Neal, 2009;

Johnson et al., 2019), or by strengthening the ability of behavioral repetition to contribute to habit strength (de Wit and Dickinson, 2009). Previous research has examined the roles of these components individually. For instance, Verplanken (2006) established that, while behavioral frequency contributed to habits, behavioral frequency alone cannot explain the full impact of habits. Meanwhile, Wood et al. (2005) demonstrated that changing contexts disrupted habits. Indeed, the associations of frequency and contextual stability with habit strength are so well accepted that the multiplicative interaction of behavioral frequency and contextual stability ($BF \times CS$) has been often used as a measurement of habit strength (see Ouellette and Wood, 1998). Phillips et al. (2016) have also shown that intrinsic rewards predict exercise behavior through intentions for those beginning an exercise routine, but through habit strength for those maintaining a previous routine. A further, recent study found that intrinsic motivation and pleasure strengthened the repetition-habit association for new behaviors (Judah et al., 2018). Yet, to date, no single study has simultaneously mapped the relative weights of each of these three components (frequency, contextual stability, and reward) in their associations with automaticity. Further, there has been no research assessing how each of these components contribute to automaticity across a spectrum of behaviors.

As mentioned, McEachan et al. (2011) found that different types of behavior were differentially predicted by past behavior; therefore, there is a need to understand how characteristics of behaviors influence automaticity. The *complexity* of the behavior has been proposed to impact the development of habit-related automaticity (Wood et al., 2002; Verplanken, 2006; Wood and Neal, 2009; Lally et al., 2010). Behavioral complexity can be understood as the number of physical or mental steps involved in executing the behavior, in which behaviors that are complex are more time-consuming and require a greater amount of planning; for example, simple behaviors are exemplified by handwashing or cigarette smoking and complex behaviors by performing well on an intellectual task or quitting smoking (Boynton, 2005). More complex behaviors may have reduced habit strength compared to simple behaviors due to the number of steps that must be learned before the behavior becomes automatic. Verplanken (2006) showed that when behavioral complexity was experimentally manipulated in a laboratory word-search task, habit formation was impeded, even when frequency was kept constant. In a daily diary study, Wood et al. (2002) further found that greater complexity of a task was associated with more thoughts about the task, which may indicate that simpler tasks are more automatic. Further generalization of this association to a broad spectrum of behaviors can bolster these findings, and other measures can assess the influence of complexity as perceived by the individual doing the behavior.

Behavioral complexity may also moderate the associations of frequency, contextual stability, and rewards with behavioral automaticity, but these interactions have not yet been tested. We developed several hypotheses *a priori* and listed them in our institutional review board protocol, along with rationales for each (although we did not pre-register them otherwise). Specifically, behavioral frequency might be a stronger predictor of

automaticity of simple behaviors, rather than complex behaviors, due to the number of steps that need to be learned in complex behaviors. Indeed, in the previous study by Verplanken (2006), habit strength for a novel behavior depended on complexity when behavioral frequency was kept constant. If habit strength presumably began at equal points (i.e., no habit strength) for each of these novel simple and complex behaviors, the differential development of habit strength over repeated actions would imply an interaction effect between frequency and complexity. Specifically, habit strength developed more slowly over repetition when the behavior was complex, rather than when it was simple. Yet, this previous study did not directly test an interaction between frequency and contextual stability. The present study examines such an interaction.

Conversely, contextual stability may be a weaker predictor of automaticity for simple behaviors compared to complex behaviors. Whereas the habits literature has focused primarily on behaviors that are executed automatically in a *singular* context, other behavior literature has also considered behaviors that are cued in multiple contexts. The addiction literature, for example has shown that multiple environmental cues can yield increased craving and engaging in a problem behavior for a particular individual (Fatseas et al., 2015). Implementation intention research has also assessed the use of multiple cue-behavior associations, but demonstrated that developing multiple “if [cue], then [behavior]” plans does not yield effective behavioral changes, compared to setting a single if-then plan (de Vet et al., 2011; Verhoeven et al., 2013). As implementation intentions as well are thought to yield behavior by increasing cognitive accessibility of cue and behavior (Webb and Sheeran, 2008), there is need to understand the conditions under which single or multiple cues yield inclinations toward behavior. Behavioral complexity may be a factor in the association between cues and the resulting behavior, as simple behaviors might easily be performed frequently in a broad variety of contexts such that many diverse cues can become strongly associated with the behavior. A jogging habit, for instance, may be cued only once a day when a person arrives home from work, as finding the time and planning resources to go jogging frequently at multiple times during the day would be difficult. The same individual may be cued to check their phone while making coffee, while in the bathroom, and during their lunch break. The contextual variability of this simpler behavior does not disprove its automaticity or cue-behavior associations.

Complexity may also moderate the influence of rewards on behavioral automaticity. It has been argued that rewards yield habit development through increased repetition, particularly by increasing intention to re-engage in that behavior (Rothman et al., 2009; Johnson et al., 2019). Yet, in a survey assessing individuals’ engagement with 48 different behaviors, from handwashing to seatbelt use to quitting smoking, Boynton (2005) also showed that intention is a stronger predictor of engagement in behavior when behaviors are complex, rather than when they are simple. Thus, if both patterns appear, then it follows that rewards are likely to be stronger predictors of automaticity for complex behaviors rather than simple behaviors.

In order to examine the associations between behavioral frequency, contextual stability, rewards, and behavioral complexity on automaticity, this study utilizes and assesses three relatively new scales. Low (2016) developed one to assess contextual stability, and another to measure perceived rewards. Both scales can be easily adapted to different behaviors, but neither scale has undergone rigorous validation. Boynton (2005) developed and validated a similarly generalizable self-report scale measuring perceived behavioral complexity, but no subsequent research has replicated it. Moreover, of these three novel scales, none have been yet published in the scientific literature.

Low’s (2016) contextual stability scale drew on TPB literature to create a broader measure of what constitutes a behavioral context. Specifically, Ajzen and Fishbein’s (2005) Principle of Compatibility is the principle that predictors such as attitudes and intentions best predict behavior when they match on the behavioral elements of target, action, context, and time (TACT). Given the learned, associative nature of habits, an impulse toward a behavior is likely to be greatest when an individual encounters a situation that matches on TACT to a previous situation in which that individual has been rewarded for the behavior. Indeed, Low (2016) argued that habits’ strong predictive validity with future behavior may be in part due to the greater inherent TACT compatibility between past and future behavior. That said, while habit research has tended to examine the extent to which an individual repeats a given behavior, thus keeping constant ‘target’ and ‘action,’ context has been assessed primarily as the extent to which an individual engages in a behavior in the same place (e.g., Norman and Cooper, 2011) or in the presence of a single, researcher-generated cue (Ouellette and Wood, 1998). ‘Context,’ or the environment in which an individual engages in a behavior, could be considered in broader terms, and may also include other individuals present or the tools with which one performs the behavior (Ajzen, 1988, 2002). A pianist cannot play music unless an instrument is present, for example, and the presence of an electronic keyboard, compared to the presence of a piano, may afford different behavioral impulses. Low’s measure, drawing on the Principle of Compatibility, includes the social context, tools, and manner with which the behavior is performed.

Previous published research assessing rewards in habit strength have measured reward constructs with a single item (e.g., Wiedemann et al., 2014; Judah et al., 2018), or through behavior-specific scales assessing intrinsic motivation to engage in a behavior (e.g., Phillips et al., 2016). Low’s measure of rewards assesses the emotional and physical feelings of engaging in a behavior, as well as the feelings of *not* engaging in that behavior, and examines both positive and negative feelings. As a result, Low’s scale potentially affords a more expansive and broadly applicable measure than is presently available.

Behavioral complexity has been assessed in previous habits literature, either through experimental manipulation (e.g., Verplanken, 2006) or through judgment on the part of the researcher (e.g., Wood et al., 2002; Lally et al., 2010). To our knowledge, Boynton’s (2005) scale represents the only validated self-report survey of individuals’ perceptions of behavioral complexity; her study found that this scale has good reliability and construct validity across 48 different behaviors. The present

study aims to replicate these findings with our selection of 25 behaviors, including health behaviors and behaviors more contemporarily relevant to current lifestyles (e.g., mobile phone checking). Use of a measure of *perceived* behavioral complexity also has potential value for the literature, as perception of behavioral barriers do not always correlate with objective measures of such behaviors (McGinn et al., 2007), but perception of difficulty nevertheless has the potential to influence behavior (Gilpin et al., 2004).

By measuring the influence of behavioral frequency, contextual stability, and rewards on automaticity across a spectrum of 25 different behaviors, the present study examines the ‘ingredients’ of habit development proposed by Wood and Neal (2016) to draw together the wide reaches of the habits literature – from exercise behavior to negative thinking. In addition, the present study expands on the tools available for examining habitual processes by testing the psychometric characteristics of three scales related to theorized components of habits, and furthers the discussion of habits by considering how characteristics of the behavior (complexity) contribute to automaticity.

MATERIALS AND METHODS

Participants and Procedure

Participants were recruited using MTurk; they were required to be 18 or older and to reside in the United States. After reviewing an information sheet and indicating agreement with the procedures, participants were directed to complete a survey using Qualtrics. Each participant was randomized to one of three clusters in which they rated 11 behaviors on several dimensions; seven behaviors were unique in each cluster, and four behaviors (exercise, smoking, handwashing, and medication adherence) were held constant across clusters. In total, 462 surveys were returned. Three participants submitted duplicate surveys; second surveys completed by the same participant were deleted. No other surveys were removed, making for a total of 459 surveys retained for analysis (154 in the first behavior group, 152 in the second group, and 153 in the third group). Ratings were extracted only from behaviors that participants had performed, making for a total of 3,790 behavior observations. Participants were paid \$5 for completing the survey.

Ethical Considerations

The protocol for this study was approved by the University of Connecticut Institutional Review Board on August 9th, 2018 (protocol #X18-095, available from authors on request). Potential participants were informed regarding the procedures and demands of the study prior to starting the survey, and were encouraged to contact the researchers if they had any concerns. Individuals who agreed to the demands of the study were directed to then complete the survey. Written consent was not collected; the survey was designed to be anonymous and low-risk, and obtaining signed consent would result in the collection of identifying information. A waiver of signed consent was granted by the University of Connecticut Institutional Review Board.

Measures

Behavior Level (Level-1) Variables

Behaviors

In total, this study collected ratings on 25 different behaviors (see **Appendix**). For each behavior, participants first were presented with a qualifier question; participants rated the extent to which they engaged in each behavior on a 7-point Likert scale. If participants responded that they did “not at all” engage in a particular behavior, then they were directed to provide ratings only on their perceived complexity of the behavior, and their ratings were not retained for analysis in this study. All participants were presented with questions for exercise, handwashing, smoking, and medication adherence. Exercise and handwashing were chosen to act as controls across groups. Smoking and medication adherence ratings were collected from all participants to achieve power with these behaviors as the authors reasoned that most participants would neither smoke nor take medications regularly and thus, a sizeable number of participants would not be able to provide ratings about their experiences with these behaviors.

In addition to the four behaviors presented to all participants, in cluster one, participants also provided ratings on active commuting, information technology use, sunscreen use, sitting, flossing, recycling, and playing music (either by singing or playing an instrument). In cluster two, participants also provided ratings on car use, making savings deposits, condom use, negative self-thoughts, sugary drink consumption, checking their phone, and texting and driving. In cluster three, participants also provided ratings on fruit and vegetable consumption, unhealthy snacking, alcohol consumption, internet use, seafood consumption, use of food safety practices, and playing video games. These behaviors were selected to represent many behaviors that have been assessed using habits in past research, as identified in a recent meta-analysis (Low, 2016).

Behavioral frequency

Behavioral frequency was measured with a single item. Participants who reported that they did engage in the given behavior on the qualifier question used a sliding scale to indicate how many times they engaged in that behavior in the average week, from 0 to 20 (or more) times a week.

Contextual stability

Contextual stability was assessed using the eight items Low (2016) developed to assess contextual stability of a behavior based on the factors of Ajzen and Fishbein’s (2005) Principle of Compatibility. Each item in this scale was scored on a scale from 0 to 10.

Perceived rewards

Perceived rewards were assessed as the feelings elicited by doing a behavior, using the items Low (2016) developed. This scale includes six items that assess the physical and emotional feelings individuals experience as a result of doing or not doing a particular behavior, and assesses both good and bad feelings. Each item in Low’s scale is scored from 0 to 10.

Perceived behavioral complexity

Perceived behavioral complexity was measured with the six-item scale that Boynton (2005) developed and validated. This scale

assesses the perceived steps involved in executing a particular behavior by measuring the extent to which an individual views a particular behavior as difficult, time-consuming, and requiring significant planning for the average adult. Each item was assessed on a 7-point Likert scale.

TPB components

Perceived behavioral control and intention were measured based on the guidelines Fishbein and Ajzen (2011) provided. *Perceived behavioral control* was measured using two 7-point Likert items: “I am confident I am capable of [doing behavior],” and “whether or not I [do behavior] is up to me.” *Behavioral intention* was measured with a single 7-point Likert item: “I intend to engage in this behavior.” For the purposes of this analysis, we included only TPB components that have been theorized to predict behavior directly. (The TPB variables of attitude and social norm were also measured but not analyzed for the present study.)

Automaticity

Automaticity was measured using the Self-Report Behavioral Automaticity Index (SRBAI; Gardner et al., 2012). While automaticity alone does not necessarily assess solely habits, this measure has been shown to be reliable and valid, and available is an adequate shorter version of the widely used Self Report Habit Index (SRHI; Verplanken and Orbell, 2003; Gardner et al., 2012). The measure has been applied to a wide variety of behavioral domains including safe food handling, fruit consumption, and physical activity (Low, 2016). Each item is scored on a 7-point Likert scale (from low to high).

Participant Level (Level-2) Variables

Demographics

Participants provided their gender, range of annual income, and age range. Participants also reported if they had found the survey through an online forum such as Reddit. Personality traits of conscientiousness and neuroticism were also measured, but not reported, for the present study.

Preliminary Analyses

Factor analyses were used to test scale validity. Exploratory factor analysis was applied to the three relatively new scales used in this study: behavioral complexity, contextual stability, and rewards. Confirmatory factor analyses were used to test the validity of the scales that have been previously well-supported. Exploratory factor analysis was run in SPSS version 25.0 (IBM Corp., 2017). Confirmatory factor analysis was run in R (R Core Team, 2018) using the lavaan package (Rosseel, 2012). Further, intraclass correlations (ICC) were also calculated for each Level-1 variable (using adjusted scales, if deemed appropriate; see Results) to assess the extent to which the different behaviors and participants accounted for variation for each scale. Within-group ICC values, clustered by participant, were also computed between Level-1 variables using the psych package in R (Revelle, 2018).

Main Analyses (and POMP-Scored Variables)

In order to account for the multiple behavior observations taken from each participant, multilevel models were used, in which

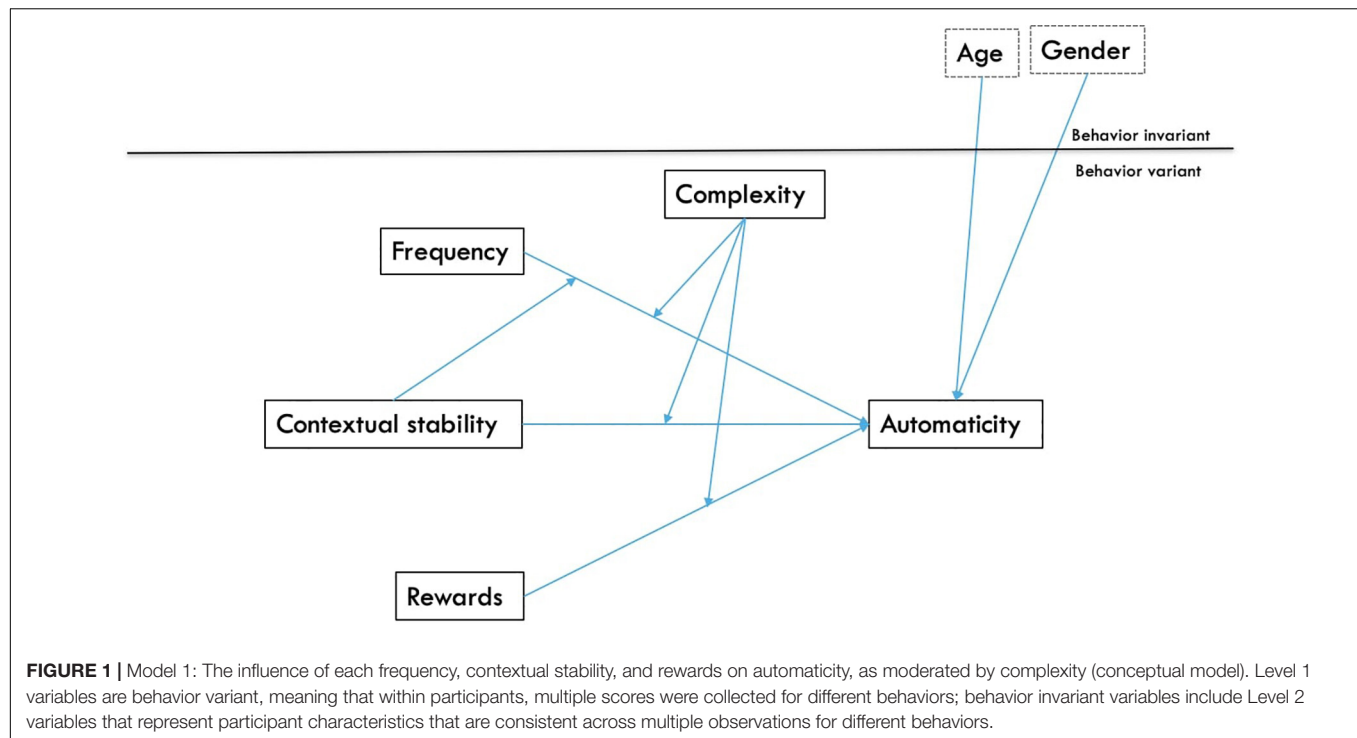
behavior ratings were nested within participants. All multilevel models were run in R using the lme4 package (Bates et al., 2015). Level-1 predictors consisted of individual ratings of behavior, including behavioral frequency, contextual stability, rewards, and complexity of the behavior. Level-2 predictors consisted of participant-level characteristics, including age and gender. Predictors were uncentered and were entered in the model in the form of percent of maximum possible (POMP) scores, such that the intercept represented the lowest score possible for each predictor (Cohen et al., 1999). Cohen et al. (1999) recommend use of POMP scores as more intuitive than presenting varying scales with unique and often meaningless units. POMP scoring has previously been used to compare across disparate scales, most frequently in meta-analysis (Cerasoli et al., 2014). In the present study, POMP scoring eases visual comparison of variables across multiple scales. Further, POMP scoring facilitates multilevel modeling and interpretation of results, as it ensures all variables are entered in the model on equivalent scales. Gender was dummy-coded. All multilevel models included random effects of behavior and participant. Significant interactions were inspected with the jtools package in R (Long, 2018). *Post hoc* mediation analyses were run using the mediation package in R (Tingley et al., 2014). Two primary models were run.

Model 1

Model 1 tested how Level-1 variables of each behavioral frequency, contextual stability, rewards, and complexity impact automaticity, as well as how complexity interacts with the other three variables to predict automaticity. An interaction between frequency and contextual stability was also included, in order to account for the association between automaticity and the popular BF \times CS measurement of habit strength. Gender and age were included as Level-2 covariates; first, main effects only were tested (reported as Model 1a), after which interactive effects were added to the model (reported as Model 1b) so as to yield accurate estimates of main and interactive effects. The model was tested with and without the interaction between frequency and contextual stability; results did not meaningfully differ, and only the model including the interaction is reported. The conceptual model appears in **Figure 1**. The general form of the model is given by:

$$\begin{aligned} \text{AUTO} = & [\gamma_{00} + \gamma_{01}\text{GENDER} + \gamma_{02}\text{AGE} + \gamma_{10}\text{FREQ} \\ & + \gamma_{20}\text{CONTEXT} + \gamma_{30}\text{REWARD} + \gamma_{40}\text{COMPLEX} \\ & + \gamma_{50}\text{COMPLEX} \times \text{FREQ} + \gamma_{60}\text{COMPLEX} \\ & \times \text{CONTEXT} + \gamma_{70}\text{COMPLEX} \times \text{REWARD} \\ & + \gamma_{80}\text{FREQ} \times \text{CONTEXT}] + \varepsilon \end{aligned}$$

Model 1 was first run as a multilevel model across behaviors, and then again individually as a regression for each of the four behaviors presented to all participants (exercise, handwashing, smoking, and medication adherence). By re-examining Model 1 for individual behaviors, extraneous confounds introduced by assessing varying behaviors in the multilevel model (such as behavioral desirability or healthiness of the behavior) were



controlled for. In particular, *objective* complexity was held constant in each individual behavior model and thus the role of *perceived* complexity was central.

Model 2

Model 2 aimed to replicate findings of Model 1 by testing the influence of rewards and complexity on habit strength, using the $BF \times CS$ interaction as a measure of habit strength. Age and gender were again included as Level-2 covariates, and a complexity \times reward interaction was entered after main effects. The conceptual model appears in **Figure 2**. The general form of the model is given by:

$$BF \times CS = [\gamma_{00} + \gamma_{01}GENDER + \gamma_{02}AGE + \gamma_{10}REWARD + \gamma_{20}COMPLEX + \gamma_{30}COMPLEX \times REWARD] + \varepsilon$$

RESULTS

Each participant provided ratings for an average of eight different behaviors, and each behavior was rated by an average of 152 participants (**Table 1**). Of all behaviors assessed in this study, handwashing was rated by the greatest number of participants (453), and texting and driving was rated by the fewest number of participants (45, representing 30% of participants presented with this behavior). **Table 2** provides descriptive statistics for both Level-1 and Level-2 variables, aggregated across behaviors. The recruited sample had similar demographic characteristics to a typical MTurk sample (Huff and Tingley, 2015). Of the 459 participants, 260 (57%) participants were male, and 197 (43%)

participants were female. A plurality (48%) of participants was between 25 and 34 years of age. Demographic information is available in the **Supplementary Materials**.

Preliminary Analyses

Missing Data

In total, 375 items were missing (0.0019% of items possible). The key dependent variable of automaticity was determined to be non-normally distributed using a Shapiro–Wilk normality test ($W = 0.90$, $p < 0.001$), and thus imputation was performed in R with the MICE package (van Buuren and Groothuis-Oudshoorn, 2011) using predictive means matching, which is particularly appropriate for non-normal data (Morris et al., 2014). Mean differences between the imputed and non-imputed datasets were assessed for each item (Diggle et al., 1995; Dong and Peng, 2013), and no significant differences were found for any items.

Differences Between Groups

There were no significant differences for behavior group for age [$F(2,456) = 2.83$, $p = 0.060$] or for gender [for being male, $F(2,456) = 3.014$, $p = 0.050$; for being female, $F(2,456) = 2.89$, $p = 0.056$; two participants selected ‘other’ as their gender]. Nonetheless, as these analyses approached significance, age and gender were retained as covariates for further analyses.

Scale Reliability and Validity

Of the scales used in this analysis, all but the scale for rewards had acceptable reliability. Contextual stability showed a reliability of $\alpha = 0.85$, 95% CI [0.85, 0.86] (ranging from $\alpha = 0.77$ to $\alpha = 0.93$ for individual behaviors); behavioral complexity had a reliability

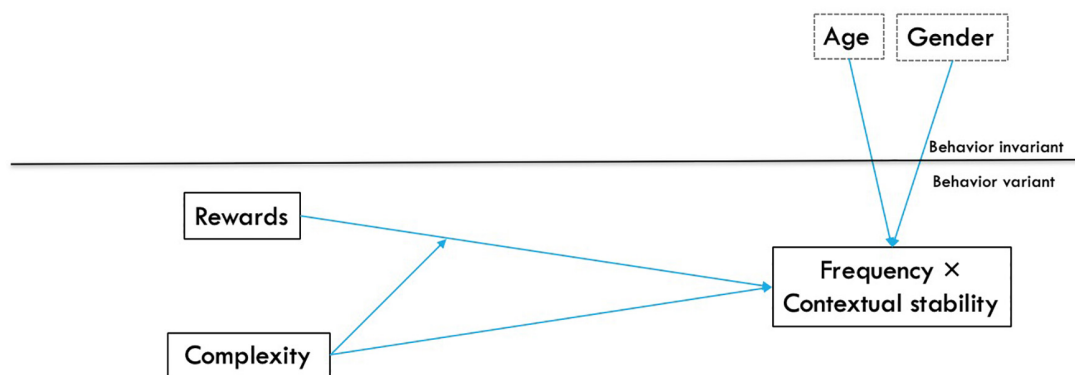


FIGURE 2 | Model 2: The influence of rewards and complexity and their interaction on habit strength, as measured by frequency and contextual stability. Level 1 variables are behavior variant, meaning that within participants, multiple scores were collected for different behaviors; behavior invariant variables include Level 2 variables that represent participant characteristics that are consistent across multiple observations for different behaviors.

TABLE 1 | Behaviors rated, ordered from most frequent to least frequently rated behaviors, along with means on key study variables.

Behavior	N Ratings	% Ratings possible	Frequency	Contextual stability	Rewards	Complexity	Habit strength	PBC	Intention
Handwashing	453	98.69	89.56	61.73	57.88	32.01	78.80	94.37	93.47
Exercise	374	81.48	26.91	62.91	64.06	65.64	40.87	89.38	85.98
Medication use	257	55.99	40.19	70.62	41.17	48.04	55.38	85.13	82.82
Fruit and vegetable consumption	153	100.00	53.73	59.00	72.75	43.79	51.70	93.42	91.13
Internet use	153	100.00	91.08	68.89	73.59	48.96	76.14	92.67	94.30
IT use	152	98.70	85.20	64.02	60.86	58.46	74.25	86.70	90.98
Sitting	152	98.70	92.66	68.50	61.71	28.03	85.48	87.03	79.89
Food safety practices	149	97.39	66.54	65.75	54.83	54.19	76.68	91.95	93.10
Phone checking	147	96.71	88.03	58.70	52.18	32.65	76.43	88.97	83.48
Smoking	143	31.15	68.25	66.42	65.94	46.77	69.26	80.42	72.13
Unhealthy snacking	139	90.85	30.76	48.17	69.21	32.33	51.82	87.62	63.72
Car use	136	89.47	50.51	63.88	53.82	59.33	58.27	87.29	86.97
Recycling	133	86.36	47.18	59.13	57.82	45.69	65.79	90.76	88.72
Playing video games	125	81.70	38.72	64.19	76.88	64.43	45.97	90.80	83.65
Flossing	122	79.22	37.91	68.73	48.11	43.51	55.15	91.74	87.70
Seafood consumption	119	78.29	17.90	48.36	73.70	42.26	34.15	89.20	77.79
Negative self-thoughts	119	77.78	38.11	30.48	13.45	44.44	69.33	67.11	33.73
Depositing savings	118	77.63	13.69	57.42	72.72	53.24	48.27	85.53	88.74
Sunscreen use	113	73.38	20.49	54.70	46.37	41.14	46.62	91.40	82.81
Sugary drink consumption	110	72.37	33.41	52.69	65.73	32.94	47.34	85.32	59.61
Active commuting	109	70.78	40.32	70.60	49.08	58.22	61.37	81.59	81.26
Alcohol consumption	97	63.40	21.55	60.48	70.31	42.39	36.49	87.26	67.75
Playing music	94	61.04	42.45	54.73	80.53	61.23	58.97	84.19	82.07
Condom use	78	51.32	18.97	54.31	58.72	44.32	57.83	85.71	80.95
Texting and driving	45	29.61	28.67	46.92	36.67	64.83	49.37	72.06	51.43

Variables are represented in the form of percent of maximum possible (POMP) scores so that higher scores represent more of the variable, using the adjusted scales where applicable (see preliminary results for more details). PBC, Perceived behavioral control. See **Appendix** for detailed definitions of each behavior.

of $\alpha = 0.84$, 95% CI [0.84, 0.85] (ranging from $\alpha = 0.55$ to $\alpha = 0.91$ for individual behaviors). One item on this scale consistently reduced the reliability of the complexity scale ("For the average adult, how automatic is this behavior?"); this item was further inspected in factor analysis and ultimately removed for multilevel analysis. Without this item, the behavioral complexity scale had a reliability of $\alpha = 0.92$ (ranging from $\alpha = 0.77$ to $\alpha = 0.96$ for

individual behaviors). The SRBAI had consistently high reliability ($\alpha = 0.96$, 95% CI [0.96, 0.96], ranging from $\alpha = 0.90$ to $\alpha = 0.97$ for individual behaviors).

The scale for rewards had a poor reliability of $\alpha = 0.51$, 95% CI [0.49, 0.54] (ranging from $\alpha = 0.03$ to $\alpha = 0.69$ for individual behaviors). Exploratory factor analyses on the underperforming rewards scale suggested two factors, but the

TABLE 2 | Descriptive statistics for within-person (Level 1) variables.

	<i>M</i>	<i>SD</i>
Habit variables		
Automaticity	60.41	29.84
Behavioral complexity	46.90	23.93
Contextual stability	60.72	21.88
Rewards	59.25	30.44
Frequency	52.02	36.85
Theory of Planned Behavior variables		
Perceived behavioral control	87.99	15.46
Intention	82.23	22.93

These descriptive statistics are drawn from the percent of maximum possible (POMP) scores, using the adjusted scales where applicable (see preliminary results for more details).

scale fit poorly onto two factors (RMSEA = 0.69, 95% *CI* [0.67, 0.71]). Given the poor reliability and validity of the rewards scale, main analyses were performed using only a single item from this scale (“When you [do behavior], how pleasurable does it feel?”). This approach is in line with previous research that has associated pleasure with habit strength (Judah et al., 2018).

Exploratory factor analysis for the behavioral complexity scale also suggested two factors, but the scale did not fit well on a two-factor model (RMSEA = 0.20, 95% *CI* [0.18, 0.22]); item analysis revealed that the second factor was driven entirely by a single item (“For the average adult, how automatic is this behavior?”). As this item also reduced the overall reliability of the scale and was determined to be particularly similar to our dependent variable of automaticity, the item was removed; when removed, the complexity scale fit well onto a single factor (RMSEA = 0.045, 95% *CI* [0.035, 0.059]). Thus, further analyses were completed using the five-item version of the complexity scale. For contextual stability, exploratory factor analysis also suggested two factors. Item analysis suggested the two factors represented a factor of stability of the physical environment, and a factor of stability of the social environment. Yet, the scale did not optimally fit onto a two-factor model (RMSEA = 0.24, 95% *CI* [0.24, 0.25]). Further, despite good reliability of the scales, the measure for contextual stability also did not map well onto a single factor (RMSEA = 0.18, 95% *CI* [0.17, 0.18]). Removing the two items that loaded on the social environment factor did not improve the fit of

this scale, and thus the full scale was retained. The SRBAI showed acceptable fit for a one-factor model (RMSEA = 0.072, 95% *CI* [0.054, 0.092]). The **Appendix** shows all scales as used for analysis.

Intraclass Correlations

First, empty multilevel linear models with random effects of *behavior* were used to compute an ICC for each Level-1 variable. As frequency and automaticity were found to be bimodally distributed around the extremes, these variables were stratified into ‘low’ and ‘high’ using a median split, and a logistic multilevel regression was run to compute ICC scores, using the formula proposed by Zeger et al. (1988). Frequency had an ICC of 0.48; automaticity had an ICC of 0.21. With a Gaussian distribution, contextual stability showed an ICC of 0.16, rewards showed an ICC of 0.22, and behavioral complexity had an ICC of 0.22. In addition, ICC values were also calculated using empty multilevel linear models with random effects of *participant*. With random effects of participant, rewards had an ICC of 0.29, contextual stability 0.36, and behavioral complexity 0.22. Using logistic models, frequency showed an ICC of 0.08 and automaticity 0.27 with random effects of participant. Within-group ICC values between Level-1 variables, clustered by participant, are reported in **Table 3**.

Main Analyses

Model 1

Model 1 (**Figure 1**) was conducted using a multilevel generalized linear model with a binomial logistic distribution, due to the non-normal distribution of automaticity. Model 1a tested main effects and found frequency, contextual stability, and rewards positively predicted behavioral automaticity, while behavioral complexity and age negatively predicted automaticity. Model 1b also included interactive effects; two significant interactions appeared (**Table 4**). At high levels of behavioral complexity, as hypothesized, rewards were more predictive of high automaticity compared to at low levels of behavioral complexity (**Figure 3**, left panel). Complexity interacted with contextual stability as predicted such that when behaviors were perceived as complex, contextual stability was a stronger predictor of high behavioral automaticity than when behaviors were perceived as simple. In addition, at low levels of contextual stability, more complex behaviors were less likely to show automaticity than simpler behaviors, while at the highest levels of contextual stability,

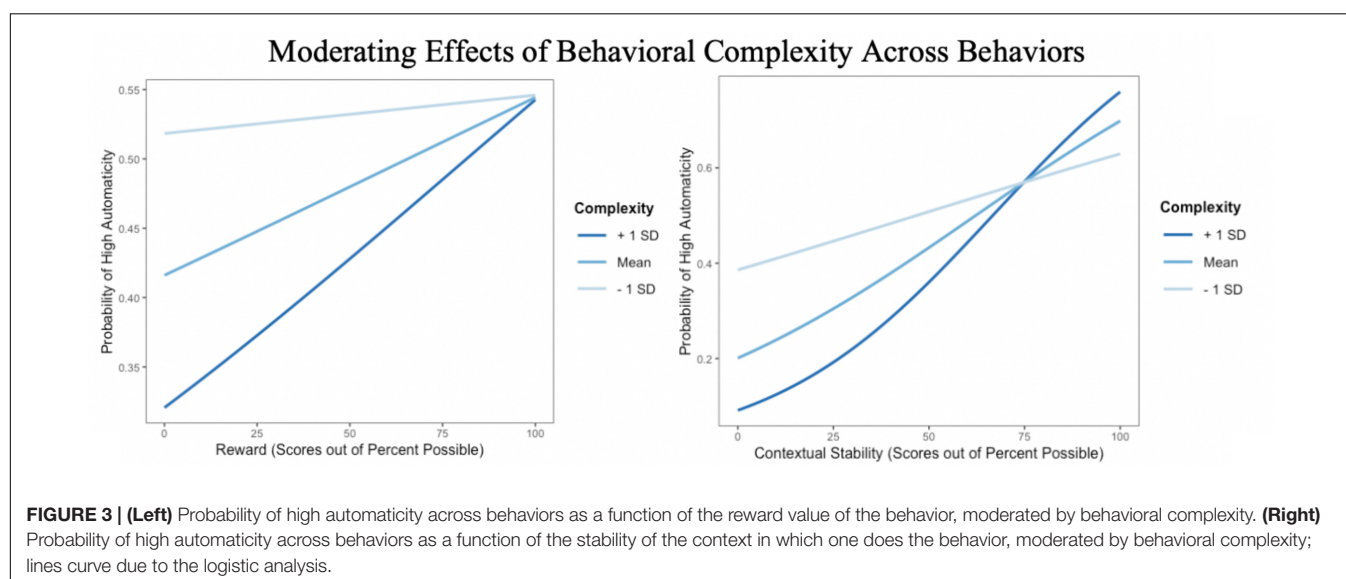
TABLE 3 | Within-group intraclass correlations (ICC) values between Level-1 variables, clustered by participant (all *ps* < 0.001).

Measure	1	2	3	4	5	6	7
(1) Automaticity	1.00						
(2) Frequency	0.65	1.00					
(3) Contextual stability	0.30	0.31	1.00				
(4) Reward	0.10	0.13	0.18	1.00			
(5) Complexity	−0.26	−0.25	0.06	0.05	1.00		
(6) Perceived behavioral control	0.12	0.18	0.26	0.26	−0.07	1.00	
(7) Intention	0.27	0.34	0.45	0.34	0.07	0.44	1.00

TABLE 4 | Results of Model 1: frequency, contextual stability, and rewards as predictors of habit strength, moderated by behavioral complexity.

Predictor variable	Unstandardized <i>B</i>	SE	Standardized β	<i>p</i>
Model 1a (Main effects only)				
Frequency	0.039	0.002	2.856	< 0.001***
Complexity	−0.006	0.003	−0.278	0.026*
Rewards	0.005	0.002	0.278	0.008**
Contextual stability	0.020	0.003	0.866	< 0.001***
Age	−0.012	0.003	−0.539	< 0.001***
Male	0.0004	0.002	0.041	0.794
Model 1b (Including interactive effects)				
Frequency × Complexity	−0.00004	0.0001	−0.214	0.380
Frequency × Contextual stability	−0.00003	0.0001	−0.16	0.666
Complexity × Rewards	0.0002	0.0001	0.813	0.019*
Complexity × Contextual stability	0.0005	0.0001	2.176	< 0.001***

Model 1a tested only the main effects; Model 1b included interactive effects alongside the previously tested main effects. Both models included random effects of behavior and participant, with behaviors nested within participant. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.



more complex behaviors were more likely to show greater automaticity than simpler behaviors (Figure 3, right panel). Frequency did not interact with behavioral complexity or contextual stability to predict high behavioral automaticity. Including interactive effects in the model significantly improved fit over the model including only main effects, $\chi^2(4, N = 459) = 31.61, p < 0.001$.

Individual behaviors

Model 1 was also run individually for the four behaviors that were rated in all three clusters: exercise, handwashing, smoking, and medication adherence (Table 5). Of these four behaviors, exercise was, on average, rated the most complex and handwashing was rated the simplest; exercise was also rated on average the most complex across the full sample of 25 behaviors, and handwashing was rated among the simplest (second only to sitting). Results for these behaviors generally showed parallel patterns to the multilevel model, with some exceptions. Behavioral frequency, contextual

stability, and rewards each predicted high automaticity for all four control behaviors, with the exception that rewards did not predict automaticity for smoking. Perceived behavioral complexity predicted high automaticity only for exercise and medication adherence. Rewards did not interact with perceived complexity to predict automaticity for any of the behaviors, but contextual stability interacted with complexity to predict high automaticity for handwashing, and a similar trend emerged for smoking. When handwashing was perceived as complex, contextual stability was positively associated with high automaticity, but when handwashing was perceived as simple, the predictive value of contextual stability on automaticity was reduced (Figure 4, left panel). When smoking was perceived as complex, contextual stability was positively associated with high automaticity, but when smoking was perceived as simple, contextual stability was negatively associated with automaticity (Figure 4, right panel). When the interaction between frequency and context was included in the model, this effect was no longer significant for smoking. Nevertheless,

TABLE 5 | Results of Model 1 by individual behaviors: frequency, contextual stability, and rewards as predictors of habit strength, moderated by behavioral complexity.

Behavior	Predictor variable	Unstandardized B	SE	Standardized β	p
Exercise					
Frequency		0.057	0.010	1.188	< 0.001***
Complexity		0.040	0.012	0.627	< 0.001***
Contextual stability		0.041	0.013	0.878	0.001**
Rewards		0.010	0.008	0.267	0.207
Male		0.002	0.003	0.100	0.561
Age		−0.014	0.008	−0.305	0.111
Frequency × Complexity		0.001	0.001	0.203	0.539
Frequency × Contextual stability		0.001	0.001	0.270	0.457
Complexity × Contextual stability		−0.0004	0.001	−0.163	0.549
Complexity × Rewards		−0.0001	0.001	−0.022	0.920
Handwashing					
Frequency		0.023	0.006	0.450	< 0.001***
Complexity		0.002	0.007	0.053	0.730
Contextual stability		0.023	0.007	0.418	0.001**
Rewards		0.011	0.004	0.334	0.009**
Male		−0.003	0.003	−0.131	0.299
Age		−0.007	0.006	−0.157	0.198
Frequency × Complexity		−0.0002	0.0003	−0.104	0.434
Frequency × Contextual stability		−0.0002	0.0004	−0.094	0.495
Complexity × Contextual stability		0.001	0.0005	0.360	0.041*
Complexity × Rewards		0.0005	0.0003	0.299	0.091
Smoking					
Frequency		0.042	0.009	1.505	< 0.001***
Complexity		0.036	0.012	0.954	0.003**
Contextual stability		0.013	0.013	0.262	0.334
Rewards		−0.002	0.010	−0.038	0.875
Male		−0.003	0.005	−0.128	0.582
Age		−0.019	0.012	−0.389	0.108
Frequency × Complexity		−0.001	0.0004	−0.524	0.201
Frequency × Contextual stability		−0.0002	0.0004	−0.148	0.608
Complexity × Contextual stability		0.001	0.0008	0.771	0.080
Complexity × Rewards		0.0004	0.0006	0.250	0.526
Medication adherence					
Frequency		0.017	0.006	0.490	0.005**
Complexity		0.015	0.008	0.342	0.055
Contextual stability		0.048	0.011	0.918	< 0.001***
Rewards		0.020	0.005	0.673	< 0.001***
Male		−0.0001	0.003	−0.003	0.985
Age		0.0001	0.0001	0.146	0.340
Frequency × Complexity		0.0002	0.0003	0.126	0.522
Frequency × Contextual stability		−0.0003	0.0004	−0.151	0.499
Complexity × Contextual stability		−0.0002	0.0006	−0.086	0.730
Complexity × Rewards		0.0004	0.0002	0.302	0.089

In all models, interactions and main effects were entered separately. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

the frequency and context interaction did not significantly predict automaticity.

Model 2

Model 2 (Figure 2) aimed to replicate findings of Model 1, using the BF × CS measurement of habit strength in place of automaticity. As Model 1 used a binomial logistic distribution, the BF × CS variable was also stratified into ‘high’

and ‘low’ using a median split in the interests of replication. In Model 2, rewards again were associated with high habit strength, and complexity was negatively associated with habit strength (Table 6). Complexity further interacted with rewards to predict habit strength, following the same patterns found in Model 1; when behaviors were perceived as complex, rewards were stronger predictors of high habit strength (Figure 5), compared to when behaviors were seen as simple. Including the

Moderating Effects of Behavioral Complexity for Individual Behaviors

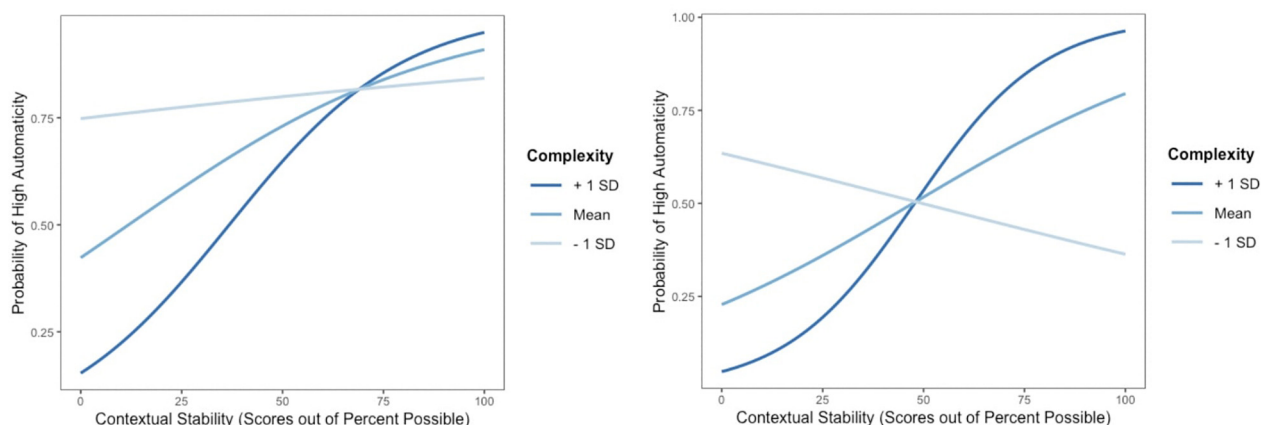


FIGURE 4 | (Left) Probability of high automaticity for *handwashing* as a function of the stability of the context in which one does the behavior, moderated by behavioral complexity; lines curve due to the logistic analysis. **(Right)** Probability of high automaticity for *smoking* as a function of the stability of the context in which one does the behavior, moderated by behavioral complexity; lines curve due to the logistic analysis.

TABLE 6 | Results of Model 2: rewards as associated with of habit strength (BF \times CS), moderated by behavioral complexity.

Predictor variable	Unstandardized B	SE	Standardized β	p
Model 2a (Main effects only)				
Rewards	0.108	0.013	0.660	< 0.001***
Complexity	-0.010	0.002	-0.491	< 0.001***
Age	-0.001	0.002	-0.028	0.774
Male	-0.0003	0.001	-0.029	0.769
Model 1b (Including interactive effects)				
Complexity \times Rewards	0.003	0.001	1.295	< 0.001***

Model 2a tested only the main effects; Model 2b included interactive effects alongside the previously tested main effects. Both models included random effects of behavior and participant, with behaviors nested within participant. *** $p < 0.001$.

interaction term significantly improved the fit of the model, $\chi^2(1, N = 459) = 23.47, p < 0.001$.

Post hoc Analyses

Preliminary analyses suggested that unhealthy behaviors were more automatic than healthy behaviors. A mediation analysis evaluated whether behavioral complexity was confounded with unhealthiness of behavior in the present study. A significant mediation effect emerged (ACME = 0.019, $p < 0.001$), with behavioral complexity accounting for 42.6% of the association between unhealthy behavior and automaticity. Unhealthiness of the behavior was no longer associated with automaticity when behavioral complexity was accounted for ($\beta = 0.122, p = 0.18$), suggesting complete mediation.

Given that rewards have been predicted to promote habit strength by promoting intention to engage in the behavior, an additional mediation analysis tested if intention explained the effect of rewards in Model 1; it did not (ACME = -0.0001, $p = 0.084$).

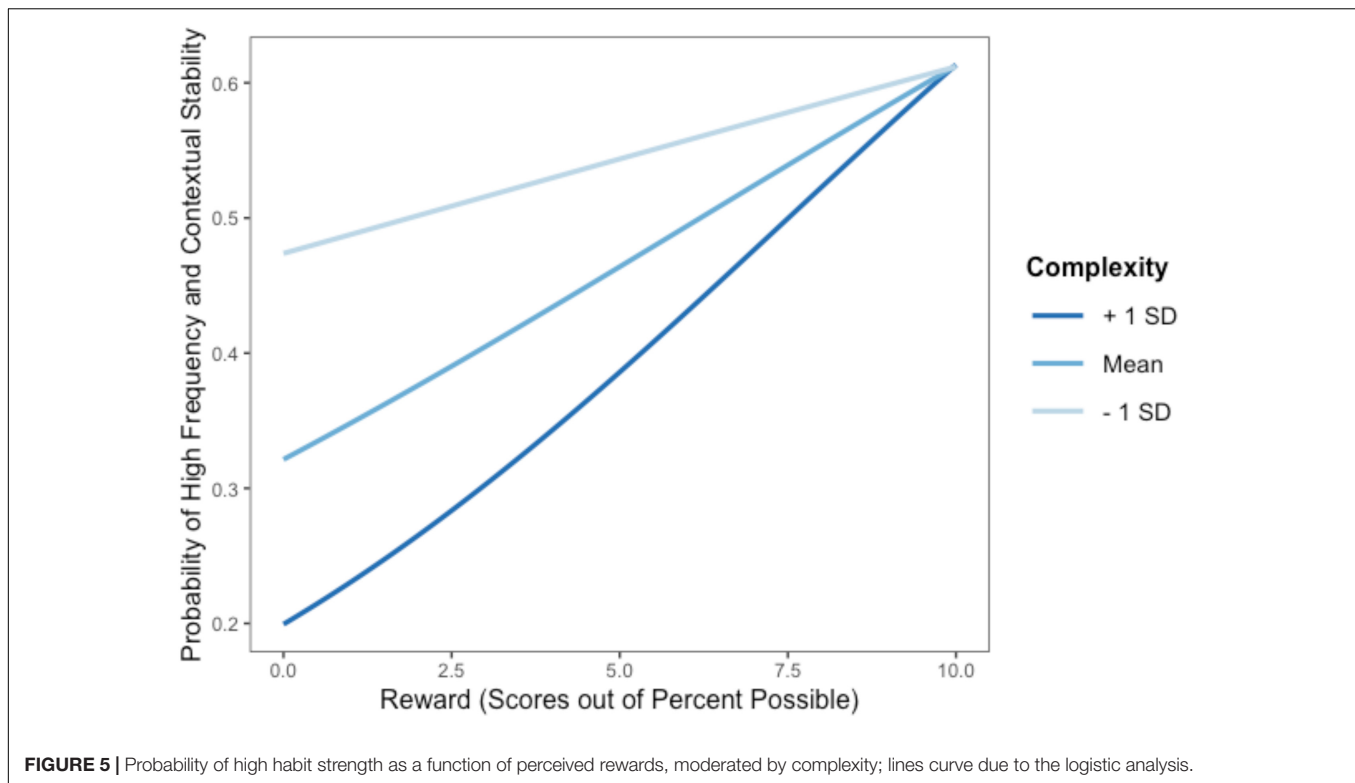
Finally, a model evaluated the predictive validity of automaticity on behavior enactment in our sample. As

behavior enactment was bimodally distributed around the extremes, a logistic analysis was again used. Results revealed that automaticity significantly predicted behavior above and beyond the effects of intention and perceived behavioral control alone, $\chi^2(1, N = 459) = 595.88, p < 0.001$.

DISCUSSION

The present study confirmed that, across 25 behaviors, behavioral frequency, contextual stability, and rewards were each associated with behavioral automaticity. It additionally established that complexity of the behavior predicts automaticity and interacts with both contextual stability and rewards, thus providing insights to the role of behavioral complexity in habitual processes (Figure 3). Together, these findings provide clarity regarding the components of habits across multiple domains of behavior.

The interactive effects of complexity on the influence of rewards and contextual stability on automaticity explains the ways in which experiences of a behavior lead to non-effortful control. Rewards are associated with positive attitudes and



intentions, and they may provide utilitarian function in promoting engagement in beneficial behaviors (e.g., even beyond the influence of intentional processes; Diamond and Loewy, 1991). Johnson et al. (2019) maintained that rewards impact habit strength by promoting intention to perform the behavior in the future, and Boynton (2005) found that executing complex behaviors (e.g., studying for an exam) is more dependent on intention than simpler behaviors (e.g., using a seatbelt). In line with this previous literature, we had expected that rewards would positively predict behavioral automaticity, and that this association would be strengthened with more complex behaviors. Both patterns appeared, when using either automaticity and the $BF \times CS$ interaction as measures of habit strength. Thus, regardless of whether one considers habit as a function of automaticity or as a function of frequency and contextual stability, perceptions of rewards and complexity are important components of habit strength.

Still, *post hoc* analyses found no significant mediation effect in which the influence of rewards on automaticity was explained by greater intention for rewarded behaviors. These findings cast doubt on an association of rewards and habit strength solely through intention, but are, nonetheless, in line with other recent research. For example, Phillips et al. (2016) found that rewards predicted exercise behavior through intention for behavior instigators, but not for behavior maintainers; possibly, in the habit formation process, intention increases initially, but diminishes as habits develop. Due to the cross-sectional nature of this study, the present research was not able to give a full picture of rewards in behavior for initiators compared to

maintainers. Judah et al. (2018) also found only inconsistent support that rewards predicted habit development through increased behavioral repetition; rather, rewards impacted habit strength by strengthening the association between doing a behavior and habit development.

The present study did not test a moderation association between rewards, behavioral frequency, and habit strength, but if complex behaviors are executed less frequently due to the number of steps and time involved in doing these behaviors, rewards may be more important for habit development for complex behaviors than simple, frequently executed behaviors by strengthening the effect of few repetitions. Additionally, Lally et al. (2010) found a logarithmic function of habit development over frequency; plausibly, rewards might drive this pattern by providing diminishing returns with each repetition. Indeed, the operant conditioning literature has established that continuous reinforcement is not as effective for long-term behavior change as variable reinforcement (Guttman, 1953), and Stawarz et al. (2015) found that although rewards effectively promoted behavior, *automaticity* development was hindered. Thus, simple behaviors that can easily be executed may not benefit as strongly as complex behaviors from the presence of rewards due to a function of diminishing returns.

Thus, while TPB approaches have argued that rewards impact behavior by promoting positive attitudes toward a behavior, which then increases intention to engage in the behavior, the present research confirms that rewards are also instrumental in non-intentional behavioral processes. In the case of positive,

healthy behaviors, this reward-based process can promote self-regulation by transferring control past the limits of intention and yielding long-term behavior change (Lally and Gardner, 2013). Yet, in the case of unhealthy or negative behaviors, rewards have the potential to circumvent self-regulation efforts (Johnson et al., 2019). The present findings support the need for a more nuanced understanding of the mechanisms through which rewards yield behavior in habits and other forms of non-effortful control.

It was hypothesized that complexity and contextual stability would interact to predict automaticity such that contextual stability would be a stronger predictor of automaticity when complexity is low. The results *did* reveal this pattern, which lends support to the argument that simple behaviors might be executed easily in multiple contexts, such that multiple cues might come to cue the same behavior. If habits are understood as the impulse toward a given behavior when an individual encounters a particular cue (Lally and Gardner, 2013), measurement of simple behavioral habits using self-report measures might not target a single habit, but rather multiple habits related to executing the same behavior. As the present study did not directly measure the specific cues that trigger habitual behaviors for each individual, this explanation cannot be further substantiated. An alternative argument might posit that while complex and simple habits have the potential to be triggered by a single environmental cue, complex behaviors require more complex cues that depend on multiple broader aspects of the overall context, while simpler habits can be initiated in response to a simple cue that can exist in multiple contexts. For instance, an individual's exercise habit might be cued when they see their sneakers by the door, but only after work and when the weather is fair, while the same individual's seatbelt habit might be cued every time they sit in a car, regardless of time of day or weather conditions. Such experiences have been reported qualitatively in previous research (Lally et al., 2011).

An unexpected interaction between contextual stability and complexity also appeared, such that when contextual stability was high, more complex behaviors were associated with greater automaticity compared to simpler behaviors. This finding appears counter-intuitive; we had no reason to expect that more complex behaviors become *more* automatic than simple behaviors when both the simple and complex behaviors are performed in stable contexts. The interaction found in this study may be an artifact of using self-report measures of automaticity across such a spectrum of behaviors; the validity of asking individuals the extent to which they enact a behavior 'without awareness' has been previously questioned (Hagger et al., 2015). It is possible – perhaps even likely – that participants scored the extent to which they executed behaviors automatically based on what they considered was automatic for that *particular* behavior, rather than across behaviors. Doing so may have yielded different criteria by which the varying behaviors were rated as automatic. For instance, we hypothesized that contextual stability would be a stronger predictor for complex behaviors rather than simple behaviors as simple behaviors could be easily executed in multiple contexts, leading to automaticity across contexts. Our participants may have been using a similar lay theory; thus,

when considering simple behaviors executed only in a particular context, they may have considered these behaviors to be less automatic *because* of their situational dependence, expecting that truly automatic simple behaviors would be executed regardless of context. Previous literature has shown, for example, that social smokers are less likely than those who smoke in multiple contexts to identify as smokers or to consider their behavior a 'personal addiction' (Moran et al., 2004), but may nevertheless reflect physiological addiction (DiFranza and Wellman, 2005).

The findings of this study largely supported the hypotheses, but other results were surprising. No effect of age was hypothesized, but age was found to be *negatively* associated with automaticity in the first model. It is possible this finding was driven by the choice of behaviors assessed in this study; alcohol consumption has been shown to peak in young adulthood (Britton et al., 2015), and several behaviors assessed in the present study are dependent on phone or internet use (such as texting and driving and IT use), which are associated with younger age (Andone et al., 2016; Neves et al., 2018).

In the first model, an interaction between behavioral frequency and complexity was predicted, such that when complexity was high, frequency would be a weaker predictor of habit strength, but no interaction was found. The present findings would suggest that the association between behavioral frequency and complexity as predictors of habit strength is purely additive. To our knowledge, the present study is the first to examine an interaction between frequency and complexity, and the present findings might support the interpretation of Verplanken's (2006) results as an additive association. While individuals in the simple task condition had higher habit strength than those in the complex task condition when frequency was held constant, perhaps the simple task condition *started* with higher habit strength due to the low levels of complexity.

Further, the BF \times CS interaction did not significantly predict automaticity after accounting for the main effects of frequency and contextual stability. This null effect is perhaps surprising given that BF \times CS is frequently used as a proxy for habit strength. Taken with the finding that contextual stability is less associated with automaticity when complexity is low rather than high, these results may suggest a need to better understand contextual stability in habits. Frequency and contextual stability may have additive rather than interactive associations with habit-related automaticity. Yet, rewards and complexity were similarly associated with the BF \times CS interaction as with automaticity; regardless of whether one considers habits as automaticity or as patterns of behavior, these components of habit hold constant. Thus the present findings appear to be relatively robust.

While the multilevel model assessed factors associated with automaticity across behaviors while accounting for random effects of individuals, the following single-level models compared individuals on a single behavior. These single-level models examining individual behaviors (see Table 5) provide insights into the components of habit strength when behavioral characteristics are held consistent. For instance, frequency was associated with automaticity for each individual behavior assessed, but rewards were associated with automaticity only for the health promotion behaviors of exercise, handwashing,

and medication adherence, and not for the health risk behavior of smoking. Thus, the prominence of frequency as a factor of habit is maintained, and rewards are important factors for behavioral automaticity, but further behavioral moderators may need to be considered.

In addition, the single-level models provide particular insights to the role of *perceived* complexity, as examining single behaviors at a time holds the objective complexity constant. When decomposing the first model to test the influence of each behavioral frequency, contextual stability, perceived rewards, and behavioral complexity on automaticity for individual behaviors, the patterns found across the full spectrum behaviors did always not hold consistent. Some associations with automaticity for individual behaviors were surprising; for each exercise, smoking, and medication adherence, perceived complexity was *positively* associated with high automaticity. Further, participants tended to rate exercise as more complex ($M = 64.06$) than handwashing ($M = 32.01$), yet, despite the finding that rewards were a stronger predictor for complex, rather than simple, behaviors when assessing all behaviors, rewards were only a significantly associated with automaticity for handwashing and not exercise. These findings further support the need to better understand the factors that yield perceptions of behavioral complexity for different behaviors; for instance, individuals who are required to take multiple daily medications may perceive medication adherence as complex, but have stronger habits for medication adherence than someone who only takes only one pill daily for a relatively minor condition. An individual who exercises moderately by jogging a few times a week may view exercise as relatively non-complex, while a 'gym rat' who devotes a significant amount of time to daily exercise may have an elaborate exercise routine. The Dunning-Kruger effect may also have played a role in the present findings, as individuals who engage more in particular behaviors may come to understand the complexities involved with that behavior, compared to those who have only had passing experiences with a behavior (Dunning, 2011). Thus, the individual behavioral models may point to additional moderators for future research examining habits across behaviors, such as health importance or knowledge of the behaviors. Further analyses with objective measures of complexity might also be compared to the present findings to confirm the influence of *perceived* complexity as compared to objective complexity. Given the theoretical non-reasoned pathways of habitual control, differential influences of perceived and objective complexity would be particularly interesting.

This study further supported the validity of a five-item version of Boynton's (2005) behavioral complexity scale using a large sample assessing a diverse span of behaviors. Future research might draw on this short, easily administered scale to assess the extent to which perceived behavioral complexity predicts behavior outcomes. Unfortunately, the other two new scales assessed by this study were not as well supported. Low's (2016) measure of contextual stability showed good reliability but was found to load onto two factors, rather than a single factor. The presence of two factors in this scale might call to question the structure of a behavioral 'context.' Previous descriptions of context in the Principle of Compatibility have

called for consideration of broad contextual factors on equal levels of generality or specificity (Ajzen, 1988), but have not detailed key facets of such contexts. Examination of the two factors that appeared in this study reveals a factor loading on the *physical* environment as well as a factor loading on the *social* environment. Future research might assess if physical and social contexts differentially influence behavioral predictors. Regardless, the scale of contextual stability did not fit particularly well on a two-factor model. The items of this scale could be adjusted and re-assessed to examine if a better-fitting two-factor structure emerges. Following such adjustments, this scale has the potential to be a valid assessment of contextual stability that provides a broader assessment than extant measures. The rewards scale showed remarkably poor reliability and validity, which may suggest this scale does not generalize to all behaviors. Different measures of rewards should be used and evaluated in future research.

Limitations and Future Directions

The findings of this study are limited by measurement validity. Several variables were assessed with a single item, and the contextual stability scale did not load well onto the expected one-factor model. Issues of measurement validity are evident in our results by the convergence of our models (Model 1 converged at gradient 0.100, while Model 2 converged at gradient 0.0004), and by the existence of standardized effect sizes greater than 1, which were not accounted for by multicollinearity. In light of considering these issues, the current findings should be interpreted with caution, and future analyses should aim to substantiate the findings of the present study with improved measures. In particular, the use of new measures for rewards and frequency would be particularly apt, given that each of these variables were measured with a single item in the present study. In addition, this study examined factors that have been theorized to lead to habit development, but only using cross-sectional methods; thus, each factor was shown to be associated with habit strength, but not explicitly to be involved with the process of habit development. Longitudinal replications are needed to support our findings.

Also, habits were measured using the SRBAI, which represents one of the shortest, validated measures tapping automaticity in habit strength. Despite the practical strengths of this measure, the SRBAI does not directly examine habits as a function of cue-behavior association, which is an important aspect of habits (Wood and Neal, 2016). As a result, the SRBAI may potentially fail to differentiate between habits and other non-learned forms of automaticity (Gardner, 2015). Regardless, findings from the second model in the present study reveal that similar patterns emerge when using alternative measurements of habit strength. No measure yet adequately taps all three dimensions of frequency, automaticity, and cue-behavior association, but as such measures are developed, findings from the present study might be further replicated with these new measures. Further, one item of the SRBAI measures the extent to which a behavior is performed frequently; in the present study, this item overlaps with the predictor of frequency, and may account for the remarkably high association between frequency and

automaticity, or for the null association between the BF \times CS and automaticity, after accounting for the main effect of frequency. An association between frequency and automaticity is unsurprising and has been supported many times in the literature, but in order to more accurately assess the relative associations between each habit ‘ingredient’ and automaticity, alternative measures that do not directly tap frequency should be used in the future.

There are alternative ways the construct of ‘rewards’ might be considered. The rewards item used in the present study assessed rewards as a function of the extent to which an individual finds the behavior to be *pleasurable* – which can be thought of as an immediate, sensory experience (Judah et al., 2018). This approach draws on the conceptualization of rewards in animal learning models of habit (e.g., Broadbent et al., 2007). Other studies have also frequently examined rewards in habits by assessing *intrinsic motivation*, or the inclination to act because of inherent enjoyment of the behavior (e.g., Gardner and Lally, 2013; Phillips et al., 2016). Pleasure and intrinsic motivation have been shown to have similar patterns of influence on habit strength, suggesting that both may be valid ways of tapping the rewards pathway (Judah et al., 2018), but future research measuring rewards as intrinsic motivation may further substantiate our findings. Rewards might also be conceptualized as extrinsic rewards: that is, as a reinforcement external to the behavior. Previous literature has suggested that external rewards might in fact undermine habit development (Wood and Neal, 2016), but future research might assess if complexity impacts this association as well.

Given that behavioral complexity and healthiness of behaviors were confounded in the present study, a different sampling of behaviors may yield a more complete picture of habits in healthy and unhealthy behaviors. Engagement in unhealthy behavior may also be influenced by low levels of social desirability and other factors specific to undesired behaviors that were not assessed in this research. Further studies might assess the different pathways by which healthy and unhealthy habits develop, controlling for complexity in order to understand the influence of these other factors. That said, the current sample of behaviors was drawn largely from the habits literature; present findings suggest that commonly studied health promotion and health risk behaviors may have different associations with habit in part *due* to varying levels of complexity, which substantiates the need to understand behavioral complexity in habits. Participants in the present study reported also consistently high levels of intention and perceived behavioral control, even for unhealthy behaviors; as such, findings may not be generalizable to unintended habits. Future research may wish to compare the factors associated with intended as compared to non-intended habits.

This study focused primarily on the components of habit *development*; future research might assess the influence of complexity on habit *disruption*. Previous research has often focused on habit disruption through changing contexts (e.g., Wood et al., 2005; Verplanken et al., 2008). If contextual stability is a stronger predictor of habit strength for complex, rather than simple behaviors, this approach might be more effective for changing complex behaviors and less effective for simpler behaviors such as the health-risk behaviors assessed in this study. Given the influence of habits on behavior beyond that of

intention, understanding the role of complexity in disruption of unwanted habits would improve efforts at behavior change in negative or health-risk behaviors.

CONCLUSIONS

In sum, this study confirms that each of the three ‘ingredients’ of habit development proposed by Wood and Neal (2016) – behavioral frequency, contextual stability, and rewards – are independently associated with automaticity across a broad spectrum of behaviors, and that complexity of the behavior often influences these associations. Perceived behavioral complexity appears to strengthen the associations of rewards and contextual stability on habit strength, and thus behavioral complexity is an important factor in mapping habitual processes and is worthy of future investigations to better understand it.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This study was approved by the University of Connecticut Institutional Review Board (IRB) on August 9th, 2018 (Protocol #X18-095). This study was exempt from collecting written consent; the procedures were deemed to be low risk, and collecting signed consent would increase the risk level as signed consent would constitute the only identifiable information collected. Before the start of the study, participants were provided with an information sheet describing the details of the study. If participants agreed to the terms described, they were instructed to continue through to the full survey.

AUTHOR CONTRIBUTIONS

KM and BTJ conceptualized and designed the study. KM collected and organized the data, and further performed statistical analysis and developed the first draft of this manuscript under the guidance of BTJ. Both authors contributed to manuscript revision, and read and approved the submitted version.

FUNDING

This study was supported, in part, by the U.S. National Institutes of Health (NIH) Science of Behavior Change Common Fund Program through an award administered by the National Institute on Aging (5U24AG052175). KM was also supported by the Jorgensen Fellowship at the University of Connecticut during the development of this study.

ACKNOWLEDGMENTS

We thank Tania B. Huedo-Medina for her support of the analyses and for comments on early drafts of this manuscript.

REFERENCES

- Ajzen, I. (1988). *Attitudes, Personality, and Behavior*. New York, NY: McGraw-Hill Education.
- Ajzen, I. (2002). *Constructing a TPB Questionnaire: Conceptual and Methodological Considerations*. Available at: <https://pdfs.semanticscholar.org/0574/b20bd58130dd5a961f1a2db10fd1fcbac95d.pdf> (accessed May, 2018).
- Ajzen, I., and Fishbein, M. (2005). "The influence of attitudes on behavior," in *The Handbook of Attitudes*, eds D. Albarracín, B. T. Johnson, and M. P. Zanna (New York: Psychology Press), 173–221.
- Andone, I., Błaskiewicz, K., Eibes, M., Trendafilov, B., Montag, C., and Markowetz, A. (2016). "How Age and Gender Affect Smartphone Usage," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, (New York, NY: Association for Computing Machinery).
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Boynton, M. (2005). *The Role of Complexity in The Prediction and Modeling of Automatic and Deliberative Behaviors*. (Masters thesis). Storrs, CT: University of Connecticut.
- Britton, A., Ben-Shlomo, Y., Benzeval, M., Kuh, D., and Bell, S. (2015). Life course trajectories of alcohol consumption in the United Kingdom using longitudinal data from nine cohort studies. *BMC Med.* 13:47. doi: 10.1186/s12916-015-0273-z
- Broadbent, N. J., Squire, L. R., and Clark, R. E. (2007). Rats depend on habit memory for discrimination learning and retention. *Lear. Mem.* 14, 145–151. doi: 10.1101/lm.455607
- Cerasoli, C. P., Nicklin, J. M., and Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: a 40-year meta-analysis. *Psychol. Bull.* 140, 980–1008. doi: 10.1037/a0035661
- Cohen, P., Cohen, J., Aiken, L. S., and West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behav. Res.* 34, 315–346. doi: 10.1207/s15327906mbr3403_2
- de Vet, E., Oenema, A., and Brug, J. (2011). More or better: do the number and specificity of implementation intentions matter in increasing physical activity? *Psychol. Sport Exerc.* 12, 471–477. doi: 10.1016/j.psychsport.2011.02.008
- de Wit, S., and Dickinson, A. (2009). Associative theories of goal-directed behaviour: a case for animal-human translational models. *Psychol. Res.* 73, 463–476. doi: 10.1007/s00426-009-0230-6
- Diamond, W. D., and Loewy, B. Z. (1991). Effects of Probabilistic Rewards on Recycling Attitudes and Behavior 1. *J. Appl. Soc. Psychol.* 21, 1590–1607. doi: 10.1111/j.1559-1816.1991.tb00489.x
- DiFranza, J. R., and Wellman, R. J. (2005). A sensitization—homeostasis model of nicotine craving, withdrawal, and tolerance: integrating the clinical and basic science literature. *Nicotine Tob. Res.* 7, 9–26. doi: 10.1080/14622200412331328538
- Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1995). *Analysis of Longitudinal Data*. New York, NY: Oxford University Press.
- Dong, Y., and Peng, C. Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus* 2, 222–239.
- Dunning, D. (2011). The Dunning-Kruger effect: on being ignorant of one's own ignorance. *Adv. Exp. Soc. Psychol.* 44, 247–296. doi: 10.1016/b978-0-12-385522-0.00005-6
- Fatseas, M., Serre, F., Alexandre, J. M., Debrabant, R., Auriacombe, M., and Swendsen, J. (2015). Craving and substance use among patients with alcohol, tobacco, cannabis or heroin addiction: a comparison of substance- and person-specific cues. *Addiction* 110, 1035–1042. doi: 10.1111/add.12882
- Fishbein, M., and Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An introduction to Theory and Research*. Boston, MA: Addison-Wesley.
- Fishbein, M., and Ajzen, I. (2011). *Predicting and Changing Behavior: The Reasoned Action Approach*. New York, NY: Psychology Press.
- Gardner, B. (2015). A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. *Health Psychol. Rev.* 9, 277–295. doi: 10.1080/17437199.2013.876238
- Gardner, B., Abraham, C., Lally, P., and de Bruijn, G.-J. (2012). Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *Int. J. Behav. Nutr. Phys. Act.* 9, 1–12. doi: 10.1186/1479-5868-9-102
- Gardner, B., and Lally, P. (2013). Does intrinsic motivation strengthen physical activity habit? Modeling relationships between self-determination, past behaviour, and habit strength. *J. Behav. Med.* 36, 488–497. doi: 10.1007/s10865-012-9442-0
- Gilpin, E. A., Lee, L., and Pierce, J. P. (2004). Does adolescent perception of difficulty in getting cigarettes deter experimentation? *Prev. Med.* 38, 485–491. doi: 10.1016/j.ypmed.2003.12.001
- Guttman, N. (1953). Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as reinforcing agent. *J. Exp. Psychol.* 46, 213–224. doi: 10.1037/h0061893
- Hagger, M. S., Chatzisarantis, N. L., and Biddle, S. J. (2002). A meta-analytic review of the theories of reasoned action and planned behavior in physical activity: predictive validity and the contribution of additional variables. *J. Sport Exerc. psychol.* 24, 3–32. doi: 10.1123/jsep.24.1.3
- Hagger, M. S., Rebar, A. L., Mullan, B., Lipp, O. V., and Chatzisarantis, N. L. (2015). The subjective experience of habit captured by self-report indexes may lead to inaccuracies in the measurement of habitual action. *Health Psychol. Rev.* 9, 296–302. doi: 10.1080/17437199.2014.959728
- Huff, C., and Tingley, D. (2015). Who are these people? Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Res. Polit.* 2, 1–12.
- Ibm Corp. (2017). *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: IBM Corp.
- Johnson, B. T., Landrum, A., and McCloskey, K. (2019). "Attitude scholarship in the 21st century: Accomplishments, challenges, and gaps," in *The Handbook of Attitudes*, Vol. 1 (2nd Edn), eds D. Albarracín and B. T. Johnson (New York, NY: Francis & Taylor), 625–649.
- Judah, G., Gardner, B., Kenward, M. G., DeStavola, B., and Aunger, R. (2018). Exploratory study of the impact of perceived reward on habit formation. *BMC Psychol.* 6:62. doi: 10.1186/s40359-018-0270-z
- Lally, P., and Gardner, B. (2013). Promoting habit formation. *Health Psychol. Rev.* 7, S137–S158.
- Lally, P., Van Jaarsveld, C. H., Potts, H. W., and Wardle, J. (2010). How are habits formed: modelling habit formation in the real world. *Eur. J. Soc. Psychol.* 40, 998–1009. doi: 10.1002/ejsp.674
- Lally, P., Wardle, J., and Gardner, B. (2011). Experiences of habit formation: a qualitative study. *Psychol. Health Med.* 16, 484–489. doi: 10.1080/13548506.2011.555774
- Long, J. A. (2018). *jtools: Analysis and Presentation of Social Scientific Data. R package Version 1.1.1*. <https://cran.r-project.org/package=jtools>
- Low, R. E. (2016). *Place in Habits and Habits in Place*. Available at: <https://opencommons.uconn.edu/dissertations/1103> (accessed June 5, 2016).
- McEachan, R. R. C., Conner, M., Taylor, N. J., and Lawton, R. J. (2011). Prospective prediction of health-related behaviours with the theory of planned behaviour: a meta-analysis. *Health Psychol. Rev.* 5, 97–144. doi: 10.1080/08870446.2011.613995
- McGinn, A. P., Evenson, K. R., Herring, A. H., Huston, S. L., and Rodriguez, D. A. (2007). Exploring associations between physical activity and perceived

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01556/full#supplementary-material>

- and objective measures of the built environment. *J. Urban Health* 84, 162–184. doi: 10.1007/s11524-006-9136-4
- Moran, S., Wechsler, H., and Rigotti, N. A. (2004). Social smoking among US college students. *Pediatrics* 114, 1028–1034. doi: 10.1542/peds.2003-0558-1
- Morris, T. P., White, I. R., and Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Med. Res. Methodol.* 14:75. doi: 10.1186/1471-2288-14-75
- Neal, D. T., Wood, W., and Quinn, J. M. (2006). Habits—A repeat performance. *Curr. Dir. Psychol. Sci.* 15, 198–202. doi: 10.1111/j.1467-8721.2006.00435.x
- Neves, B. B., Fonseca, J. R., Amaro, F., and Pasqualotti, A. (2018). Social capital and Internet use in an age-comparative perspective with a focus on later life. *PLoS One* 13:e0192119. doi: 10.1371/journal.pone.0192119
- Norman, P., and Cooper, Y. (2011). The theory of planned behaviour and breast self-examination: assessing the impact of past behaviour, context stability and habit strength. *Psychol. Health* 26, 1156–1172. doi: 10.1080/08870446.2010.481718
- Ouellette, J. A., and Wood, W. (1998). Habit and intention in everyday life: the multiple processes by which past behavior predicts future behavior. *Psychol. Bull.* 124, 54–74. doi: 10.1037//0033-2909.124.1.54
- Phillips, L. A., Chamberland, P. É., Hekler, E. B., Abrams, J., and Eisenberg, M. H. (2016). Intrinsic rewards predict exercise via behavioral intentions for initiators but via habit strength for maintainers. *Sport Exerc. Perform. Psychol.* 5, 352–364. doi: 10.1037/spy0000071
- R Core Team (2018). *R: A Language and Environment For Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Revelle, W. (2018). *Psych: Procedures for Personality and Psychological Research. R Package Version 1.8.4*. Evanston, Illinois: Northwestern University.
- Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.3389/jpsy.2014.01521
- Rothman, A. J., Sheeran, P., and Wood, W. (2009). Reflective and automatic processes in the initiation and maintenance of dietary change. *Ann. Behav. Med.* 38(Suppl. 1), s4–s17. doi: 10.1007/s12160-009-9118-3
- Stawarz, K., Cox, A. L., and Blandford, A. (2015). “Beyond self-tracking and reminders: designing smartphone apps that support habit formation,” in *Proceedings of The 33rd annual ACM Conference on Human Factors in Computing Systems*, (Seoul).
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., and Imai, K. (2014). mediation: R package for causal mediation analysis. *J. Stat. Softw.* 59, 1–38.
- van Buuren, S., and Groothuis-Oudshoorn, K. (2011). mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* 45, 1–67.
- Verhoeven, A. A., Adriaanse, M. A., De Ridder, D. T., De Vet, E., and Fennis, B. M. (2013). Less is more: the effect of multiple implementation intentions targeting unhealthy snacking habits. *Eur. J. Soc. Psychol.* 43, 344–354. doi: 10.1002/ejsp.1963
- Verplanken, B. (2006). Beyond frequency: habit as mental construct. *Br. J. Soc. Psychol.* 45, 639–656. doi: 10.1348/014466605x49122
- Verplanken, B., and Orbell, S. (2003). Reflections on past behavior: a self-report index of habit strength. *J. App. Soc. Psychol.* 33, 1313–1330. doi: 10.1111/j.1559-1816.2003.tb01951.x
- Verplanken, B., Walker, I., Davis, A., and Jurasek, M. (2008). Context change and travel mode choice: combining the habit discontinuity and self-activation hypotheses. *J. Env. Psychol.* 28, 121–127. doi: 10.1016/j.jenvp.2007.10.005
- Webb, T. L., and Sheeran, P. (2008). Mechanisms of implementation intention effects: the role of goal intentions, self-efficacy, and accessibility of plan components. *Br. J. Soc. Psychol.* 47, 373–395. doi: 10.1348/014466607x267010
- Wiedemann, A. U., Gardner, B., Knoll, N., and Burkert, S. (2014). Intrinsic rewards, fruit and vegetable consumption, and habit strength: a three-wave study testing the associative-cybernetic model. *Appl. Psychol. Health Well Being* 6, 119–134. doi: 10.1111/aphw.12020
- Wood, W. (2017). Habit in personality and social psychology. *Pers. Soc. Psychol. Rev.* 21, 389–403. doi: 10.1177/1088868317720362
- Wood, W., and Neal, D. T. (2009). The habitual consumer. *J. Consum. Psychol.* 19, 579–592. doi: 10.1016/j.jcps.2009.08.003
- Wood, W., and Neal, D. T. (2016). Healthy through habit: interventions for initiating & maintaining health behavior change. *Behav. Sci. Policy* 2, 71–83. doi: 10.1353/bsp.2016.0008
- Wood, W., Quinn, J. M., and Kashy, D. A. (2002). Habits in everyday life: thought, emotion, and action. *J. Pers. Soc. Psychol.* 83, 1281–1297. doi: 10.1037//0022-3514.83.6.1281
- Wood, W., Tam, L., and Witt, M. G. (2005). Changing circumstances, disrupting habits. *J. Pers. Soc. Psychol.* 88, 918–933. doi: 10.1037/0022-3514.88.6.918
- Zeger, S. L., Liang, K. Y., and Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44, 1049–1060.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 McCloskey and Johnson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Measures

Contextual Stability

When you [do behavior], how consistently do you do it. . . . 0(Never the Same). . . 10(Always the same)

- (1) At the same time(s) of day?
- (2) On the same day(s) of the week?
- (3) For the same amount(s) of time per sitting?
- (4) In the same particular way(s)?
- (5) In the same location(s)?
- (6) With the same people (who are also doing it)?
- (7) Around the same people (who are NOT doing it)?
- (8) Using the same object(s) or tool(s)?

Rewards

When you [do behavior] how does it feel? 0(Not at all). 10(Extremely)

- (1) Pleasurable?

Behavioral Complexity

1(Not at All). . . . 7(A Great Deal)

- (1) In general, how challenging is [behavior] for the average person?
- (2) How much attention does the average person need to give to [behavior] to do it well?
- (3) How complex is [behavior]?
- (4) How much planning is involved before [behavior]?
- (5) On average, how much time does it take to [behavior]?

Behavior Definitions

Exercise

By exercise, I mean engaging in physical behavior for 30 min or more that elevates your heart rate.

Handwashing

By handwashing, I mean washing your hands in any context.

Smoking

By smoking, I mean using cigarettes to smoke tobacco.

Taking Medication

By taking medication, I mean taking medication that has been prescribed to you by a healthcare professional.

Fruit and Vegetable Consumption

By eating fruits and vegetables I mean any time you eat fruits or vegetables, not necessarily at the same time.

Unhealthy Snacking

By unhealthy snacking I mean eating foods high in fat or sugar not at meal times.

Alcohol Consumption

By drinking alcohol, I mean drinking at least one unit of alcohol (approximately one measure of spirits, half a glass of wine, or half a pint of beer).

Internet Use

By internet use, I mean accessing the internet through computers, phones, or tablets for either leisure or work use.

Seafood Consumption

By eating seafood, I mean eating fish or shellfish.

Food Safety Practices

By food safety practices, I mean handling and treating food in such a way as to reduce the risk of getting sick from food, such as washing hands and surfaces when handling food, keeping food at the “correct” temperature, and avoiding unsafe foods.

Playing Video Games

By playing video games, I mean playing games on a computer or console.

Active Commuting

By active commuting, I mean traveling to and from work or school by a means that requires some physical activity on your part, such as walking, biking, or using public transport.

IT Use

By IT use, I mean using technology to save, receive, or send information.

Sunscreen Use

By sunscreen use, I mean applying a product with SPF protection to your skin.

Sitting

By sitting, I mean sitting in any context, such as in a car or on a bus, or at work/home/or school.

Flossing

For the following questions, I will ask you about your feelings and behaviors regarding flossing. By flossing, I mean using dental floss to floss your teeth.

Recycling

By recycling, I mean putting recyclable materials in recycling receptacles.

Playing Music

By playing music, I mean performing music by playing a musical instrument and/or by singing.

Car Use

By car use, I mean that when you need to use transportation, you drive a car.

Depositing Savings

By depositing savings, I mean putting money in a dedicated savings account.

Condom Use

By condom use, I mean using a condom while engaging in sexual activity.

Negative Self-Thoughts

By negative self-thoughts, I mean negative thoughts you have about yourself.

Sugary Drink Consumption

By drinking sugary drinks, I mean drinking beverages that are high in sugar, such as soda, energy drinks, or juice.

Phone Checking

By checking your phone, I mean checking your phone for notifications with or without a notification alert.

Texting and Driving

By texting and driving, I mean reading and sending texts and/or instant messages while driving.



Enhanced Avoidance Habits in Relation to History of Early-Life Stress

Tara K. Patterson*, Michelle G. Craske and Barbara J. Knowlton

Department of Psychology, University of California, Los Angeles, Los Angeles, CA, United States

OPEN ACCESS

Edited by:

John A. Bargh,
Yale University, United States

Reviewed by:

Elizabeth Tricomi,
Rutgers University, The State
University of New Jersey,
United States
Jony Sheynin,
Texas A&M Health Science Center,
United States
Lars Schwabe,
Universität Hamburg, Germany

*Correspondence:

Tara K. Patterson
tkpatterson@ucla.edu

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 16 March 2019

Accepted: 30 July 2019

Published: 13 August 2019

Citation:

Patterson TK, Craske MG and
Knowlton BJ (2019) Enhanced
Avoidance Habits in Relation
to History of Early-Life Stress.
Front. Psychol. 10:1876.
doi: 10.3389/fpsyg.2019.01876

The effect of stress on the balance between goal-directed behavior and stimulus-response habits has been demonstrated in a number of studies, but the extent to which stressful events that occur during development affect the balance between these systems later in life is less clear. Here, we examined whether individuals with a history of early-life stress (ELS) show a bias toward avoidance habits on an instrumental learning task as adults. Participants ($N = 189$ in Experiment 1 and $N = 112$ in Experiment 2) were undergraduate students at the University of California, Los Angeles. In Experiment 1, we hypothesized that a history of ELS and a longer training phase would be associated with greater avoidance habits. Participants learned to make button-press responses to visual stimuli in order to avoid aversive auditory outcomes. Following a training phase involving extensive practice of the responses, participants were tested for habitual responding using outcome devaluation. After completing the instrumental learning task, participants provided retrospective reports of stressful events they experienced during their first 16 years of life. We did not observe evidence for an effect of the length of training, but we did observe an effect of ELS, with greater stress predicting greater odds of performing the avoidance habit. In Experiment 2, we sought to replicate the effect of ELS observed in Experiment 1, and we also tested whether the presence of distraction during training would increase avoidance habit performance. We replicated the effect of ELS but we did not observe evidence of an effect of distraction. Taken together, these data lend support to the hypothesis that stress occurring during development can have lasting effects on the balance between goal-directed behavior and stimulus-response habits in humans. Enhancement of avoidance habits may help explain the higher levels of negative health outcomes such as heart and liver disease that have been observed in individuals with a history of ELS. Some of the negative health behaviors that contribute to these negative health outcomes, e.g., overeating and substance use, may be performed initially to avoid feelings of distress and then transition to being performed habitually.

Keywords: stress, habit, avoidance learning, instrumental learning, outcome devaluation, goal-directed action

INTRODUCTION

The effects of stress on physical and psychological health have been of increasing interest in recent years, with one area of focus being how individuals are affected by stress that occurs during development (early-life stress, ELS). Common sources of ELS are childhood abuse and neglect. Such experiences have been shown to cast a long shadow on health throughout the lifespan, affecting

outcomes in adulthood ranging from severe obesity (Anda et al., 2006), heart disease (Dong et al., 2004), and liver disease (Dong et al., 2003) to sexually transmitted disease (Hillis et al., 2000) and depressive disorders (Chapman et al., 2004). The behavioral and neural mechanisms of the associations between ELS and adult health are largely unknown. Because many negative health outcomes are linked to repetitive behaviors such as overeating or substance use, it is possible that an increased reliance on stimulus–response habits in this population could explain some of the health effects experienced by its constituents.

Stimulus–response habits can be defined as instrumental behaviors that, in contrast to goal-directed actions, have come to be automatically elicited by stimuli in whose presence the behavior has been repeatedly performed, without regard to instrumental outcomes (Dickinson, 1985). For example, an animal that has been overtrained to press a lever to obtain a food reward will persist in lever pressing even after the food outcome has been devalued (Adams, 1982). In this scenario, the animal's behavior is thought to be guided by the stimulus–response association (i.e., the association between the lever and the pressing behavior) rather than by the value of the outcome (i.e., the food reward), because the animal persists in performing the response when the stimulus is present even though the outcome associated with performing that response is no longer desired. Habits have also frequently been studied using maze navigation tasks, especially in rodents; in this assay, habitual behavior is assessed by setting up a situation where the extent to which behavior is based on stimulus–response associations can be inferred from navigation decisions or performance (e.g., Packard and McGaugh, 1992; McDonald and White, 1994). In humans, habitual behavior has also been investigated with the probabilistic classification task, in which participants learn to classify stimuli based on trial-by-trial feedback. This task can be performed using the habit memory system, as in the case of individuals with amnesia (Knowlton et al., 1994), and can also be performed using the declarative memory system, as in the case of individuals with Parkinson's disease (Knowlton et al., 1996).

A number of studies have shown that stress increases habitual behavior in both non-human animals and humans. Experimentally induced stress has been shown to decrease sensitivity to outcome devaluation (Dias-Ferreira et al., 2009; Schwabe and Wolf, 2009, 2010), increase habitual behavior in maze navigation (Kim et al., 2001; Schwabe et al., 2008), and bias competition between the declarative memory system and the habit learning system in favor of habit learning in probabilistic classification (Schwabe and Wolf, 2012). The effects of stress on habitual behavior are likely mediated by stress hormones (for review, see Wirz et al., 2018), and a study using human infants showed that this stress-induced shift to habitual responding can occur as early as 15 months of age (Seehagen et al., 2015). Although most studies of this phenomenon have measured habitual behavior shortly after stress exposure, a study of male rats exposed to stress during the first 2 weeks of life found that they showed increased habitual behavior in maze navigation as adolescents (Grissom et al., 2012), and humans whose mothers reported that they were exposed to stress prenatally

showed increased habitual behavior in maze navigation as adults (Schwabe et al., 2012).

Two other factors that have been shown to influence habitual behavior are the amount of training and the presence of distraction. Animals that receive a limited amount of training show behavior that is goal-directed (i.e., sensitive to outcome devaluation), whereas animals that receive extended training show behavior that is habitual (i.e., insensitive to outcome devaluation), indicating that with greater training, behavior transitions from being controlled by action–outcome associations to being controlled by stimulus–response associations (Adams, 1982; Dickinson, 1985). One study has successfully demonstrated this effect in humans, showing that participants who received limited training were sensitive to outcome devaluation whereas participants who received extended training were not (Tricomi et al., 2009). A second factor that appears to influence habits is distraction. For example, in the probabilistic classification task, the presence of distraction by a secondary task appears to bias competition between the declarative memory system and the habit learning system in favor of habit learning (Foerde et al., 2006, 2007).

Stimulus–response habits can be appetitive (e.g., pressing a lever to receive a food reward) or avoidant (e.g., pressing a lever to avoid a shock). Most research on stimulus–response habits has been conducted using appetitive habits, as the methods for evaluation of habit formation through devaluation of appetitive outcomes via conditioned taste aversion or selective satiation procedures have been well established (for review, see Knowlton and Patterson, 2018). However, in a pair of studies conducted by Gillan et al. (2014, 2015), a shock avoidance task incorporating a novel procedure for devaluation of aversive outcomes was used to investigate avoidance habits. In this task, participants learned to avoid electric shocks delivered to the left and right wrist by making responses to warning stimuli with the left and right foot, respectively. Next, one of the two outcomes was devalued by disconnecting one of the electrodes and leaving the other electrode connected. Participants' responding to the valued and devalued stimuli was then tested in extinction. Selective responding to the still-valued stimulus indicates that participants have flexibly adjusted their behavior (i.e., that they are behaving in a goal-directed manner), whereas persistence in responding to the devalued stimulus despite the built-in cost to performance that results from continuing to hold in mind a rule that no longer applies and executing unnecessary behaviors on the basis of this rule is interpreted as habitual behavior. Using this procedure, Gillan et al. (2014, 2015) demonstrated enhanced avoidance habits in individuals with obsessive-compulsive disorder. Like compulsions, some negative health behaviors such as overeating and substance use can be understood as avoidance habits, because they may be performed initially in order to avoid feelings of distress, and then eventually transition to being performed habitually. We were therefore interested in whether adults with a history of ELS might also show enhanced avoidance habits. If so, this tendency could represent a behavioral vulnerability that increases the likelihood of the poor health outcomes observed in this group.

We used a noise avoidance task similar to the shock avoidance task used by Gillan et al. (2014, 2015), wherein participants could avoid hearing aversive noises delivered to the left and right ears by making the correct keyboard responses to associated warning stimuli. After learning the responses, participants underwent an instructed devaluation procedure in which one of the two earphones previously delivering aversive noises was removed, and then a test for habit formation was conducted in extinction. Avoidance habit formation was measured by whether the participant persisted in making the keyboard response associated with avoiding noise to the ear from which the earphone had been removed. In addition to testing for an effect of ELS, we also manipulated the level of training participants received (Experiment 1) and the level of distraction present during training (Experiment 2). The primary hypothesis of this study was that individuals who reported a history of ELS would show enhanced avoidance habits. The secondary hypotheses were (a) that individuals who received a greater level of training prior to devaluation would show enhanced avoidance habits relative to those who received less training, and (b) that learning the stimulus–response associations in the presence of distraction would lead to enhanced avoidance habits relative to associations learned without distraction.

EXPERIMENT 1

Materials and Methods

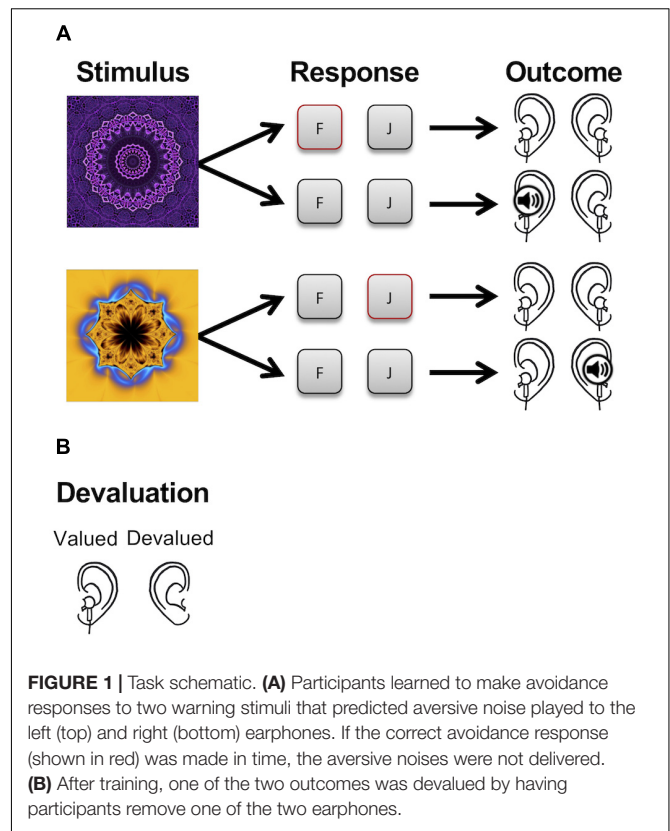
Participants

Study participants were recruited from the undergraduate student population in the Psychology Department at the University of California, Los Angeles. Participants were compensated with credit toward partial fulfillment of course requirements. Study procedures were approved by the Institutional Review Board of the University of California, Los Angeles, and all participants provided written record of informed consent.

A total of 198 participants were recruited for the study. Five participants did not complete the experiment, one participant failed to follow the instructions, two participants provided incomplete questionnaire data, and one data file was overwritten due to experimenter error, yielding a sample size of 189 (148 women, 41 men, $M_{\text{age}} = 20.31$ years, $SD_{\text{age}} = 1.81$ years, age range: 18–28 years).

Design and Procedure

The avoidance learning task was adapted from procedures described in Gillan et al. (2014, 2015). A schematic of the task is shown in **Figure 1**. Participants were instructed that their task was to avoid hearing aversive noises. Participants were shown two abstract visual warning stimuli that predicted aversive noise to the left and right earphones, respectively, and were told that they could avoid hearing the aversive noises by making the correct keyboard responses when they saw the warning stimuli. Performing the correct response with the left hand avoided noise to the left earphone, and performing the correct response with the right hand avoided noise to the right earphone. A third stimulus



was designated as the “safe” stimulus and never predicted aversive noise. Assignment of the three images to the three experimental trial types (warning stimulus 1, warning stimulus 2, and safe stimulus) was randomized across participants. On each trial, one of the three stimuli was selected randomly and presented on screen for 500 ms. Correct responses to the warning stimuli prevented aversive noise from being delivered to the earphones, but did not terminate the stimulus. If the participant pressed the incorrect key or failed to respond within 500 ms, the aversive noise (an audio file resembling a female scream) was delivered to the corresponding earphone. A female scream was selected as the aversive outcome based on the ease of implementation in comparison to an electric shock and based on prior research that used a female scream as an effective unconditional stimulus (e.g., Lau et al., 2008; Britton et al., 2011). Responses to the safe stimulus had no effect. There was a delay of 500 ms between termination of the warning stimulus and delivery of the aversive noise, and the intertrial interval was 2 s. Audio files were 1 s long and played at a volume of 82 dB.

Following demonstration of the stimulus–outcome contingencies, participants performed six practice trials (two per stimulus). Participants were allowed to repeat the practice phase if desired. The main experiment consisted of two phases, a training phase and a post-devaluation habit test. The amount of training was varied between subjects; participants in the short training condition completed 120 trials (40 per stimulus), and participants in the long training condition completed 600 trials (200 per stimulus). Assignment to condition was randomized

across participants. After training was complete, one of the two outcomes was devalued by having participants remove one of the earphones. Which earphone was removed (left versus right) was counterbalanced across participants. Participants were told that they would be evaluated based on the responses they made to avoid noise to the earphone that had not been removed, and that it was not necessary to make the response associated with avoiding noise to the earphone that had been removed. The habit test was conducted in extinction (i.e., no noises were delivered to either earphone), but participants were not informed of this. The habit test consisted of 30 trials (10 per stimulus in random order). The dependent variable of interest was whether the participant persisted in performing the response associated with avoiding aversive noise to the removed earphone, as performance of this behavior was no longer of value and thus would be evidence of habit formation. Therefore, during the post-devaluation habit test, responding to the valued stimulus was defined as performing the response associated with avoiding aversive noise to the non-removed earphone when presented with the stimulus that had predicted aversive noise to the non-removed earphone, and responding to the devalued stimulus was defined as performing the response associated with avoiding aversive noise to the removed earphone when presented with the stimulus that had predicted aversive noise to the removed earphone.

Participants completed the experiment in a private testing room on a desktop computer. Stimulus presentation and response collection were implemented in E-Prime Standard (Version 2.0). Button press responses were made using the computer keyboard. Following completion of the computer task, participants completed a packet of questionnaires. The 25-item Childhood Trauma Questionnaire – Short Form (CTQ-SF; Bernstein et al., 2003) was used to assess stress exposure during the first 16 years of life. The items on the questionnaire ask about experiences of physical abuse (e.g., being hit hard enough to leave bruises), physical neglect (e.g., not having enough to eat), emotional abuse (e.g., being called names), emotional neglect (e.g., not feeling loved), and sexual abuse (e.g., being touched in a sexual way). Each item is rated on a 5-point scale with response options ranging from “never true” to “very often true.” The mean score reported by Bernstein et al. (2003) for this measure based on a normative community sample ($N = 579$) was 39.6. The 40-item State-Trait Anxiety Inventory (STAI; Spielberger, 1983) was used to assess anxiety at the present moment (state anxiety) and in general (trait anxiety). The 20-item Beck Depression Inventory-II (BDI-II; Beck et al., 1996) was used to assess depressive symptoms during the past 2 weeks (suicidality question omitted). Finally, the 10-item Perceived Stress Scale (PSS; Cohen et al., 1983) was used to assess how unpredictable, uncontrollable, and overloading participants’ lives had been during the past month. The entire lab visit took approximately 1 h.

Data Analysis

Statistical analyses were performed using IBM SPSS Statistics (Version 25). Data from the acquisition phase (response accuracy to the two warning stimuli and false alarm rate to the safe stimulus) and level of responding to the valued stimulus during the habit test were analyzed using two (level

of training: 120 trials, 600 trials) \times two (level of ELS: low-ELS, high-ELS) between-subjects ANOVA with participants categorized as low-ELS or high-ELS based on a median split of the CTQ-SF scores. Responding to the devalued stimulus during the post-devaluation habit test was analyzed using binary logistic regression with participants’ responses binned into zero responses to the devalued stimulus (no habitual behavior) or one or more responses to the devalued stimulus (habitual behavior). Responding was binarized in this manner based on a bimodal distribution of the response data among participants who responded to the devalued stimulus, with one subgroup of participants making few responses to the devalued stimulus and a second subgroup of participants responding on the majority of devalued stimulus trials. We therefore collapsed across the two subgroups, classifying all participants who responded to the devalued stimulus as exhibiting habitual behavior. The following predictors were included in the regression model: CTQ-SF, level of training, devalued side, STAI state anxiety, STAI trait anxiety, BDI-II, PSS, CTQ-SF \times level of training, CTQ-SF \times devalued side, CTQ-SF \times STAI state anxiety, CTQ-SF \times STAI trait anxiety, CTQ-SF \times BDI-II, CTQ-SF \times PSS, age, and gender. Continuous predictors used to create interaction terms were mean-centered to reduce multicollinearity, and dichotomous predictor variables were dummy coded. A significance level of 0.05 was used for all statistical tests.

A supplemental data analysis in which the number of responses to the devalued stimulus was entered as the outcome variable is provided in **Supplementary Table 1**.

Results

Sample characteristics are reported in **Table 1** (prevalence of ELS by degree and type of stress reported) and **Table 2** (scores on questionnaire variables for low-ELS and high-ELS participants). The mean CTQ-SF score was 36.02 ($SD = 11.98$), and the median CTQ-SF score was 33.00. The low-ELS group had a mean CTQ-SF score of 27.96 ($SD = 2.17$) and the high-ELS group had a mean CTQ-SF score of 43.99 ($SD = 12.38$). The high-ELS group differed significantly from the low-ELS group on measures of state anxiety, $t(187) = 5.37$, $p < 0.001$, $d = 0.78$; trait anxiety, $t(187) = 6.57$, $p < 0.001$, $d = 0.96$; depression, $t(187) = 6.96$, $p < 0.001$, $d = 1.01$; and perceived stress, $t(187) = 3.72$, $p < 0.001$, $d = 0.54$.

We first tested for effects of the level of training (120 training trials versus 600 training trials) and level of ELS (low-ELS versus high-ELS) on response accuracy during training, false alarm rate during training, and level of responding to the valued stimulus during the post-devaluation habit test. The data from the training phase are shown in **Figure 2**. During training, response accuracy to the two warning stimuli was 81.29% ($SD = 12.10\%$), and the false alarm rate to the safe stimulus was 11.68% ($SD = 19.14\%$). Training accuracy did not differ significantly across levels of training, $F(1,185) = 2.57$, $p = 0.111$, $\eta_p^2 = 0.014$, or levels of ELS, $F(1,185) = 0.96$, $p = 0.330$, $\eta_p^2 = 0.005$, and the interaction between training and ELS was not significant, $F(1,185) = 0.78$, $p = 0.379$, $\eta_p^2 = 0.004$. False alarm rate did not differ significantly across levels of training, $F(1,185) = 0.07$, $p = 0.786$, $\eta_p^2 < 0.001$, or levels

TABLE 1 | Prevalence of ELS in sample.

	Experiment 1					Experiment 2				
	5	6–10	11–15	16–20	21–25	5	6–10	11–15	16–20	21–25
CTQ-SF subscale										
Physical abuse	62.43	31.22	4.76	0.53	1.06	59.82	33.93	3.57	2.68	0.00
Physical neglect	46.56	43.92	8.47	1.06	0.00	53.57	46.43	0.00	0.00	0.00
Emotional abuse	31.22	47.62	14.81	2.65	3.70	25.89	57.14	10.71	6.25	0.00
Emotional neglect	26.98	47.09	17.46	6.88	1.59	29.46	47.32	17.86	4.46	0.89
Sexual abuse	84.66	11.11	1.59	1.59	1.06	84.82	9.82	2.68	0.89	1.79

Percentage of participants in each experiment reporting Early-Life Stress (ELS) broken down by degree and type of ELS reported. For each subscale, five is the lowest score, corresponding to a response of “never” for all items. CTQ-SF, Childhood Trauma Questionnaire – Short Form (Bernstein et al., 2003).

TABLE 2 | Characteristics of low-ELS and high-ELS groups.

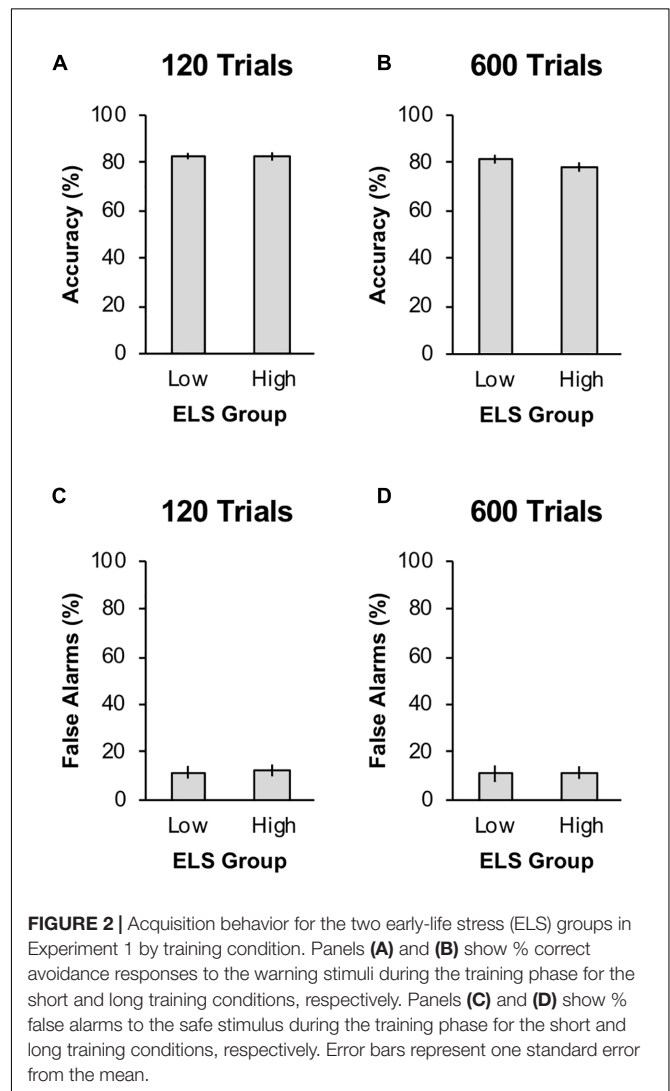
	Experiment 1		Experiment 2	
	Low-ELS	High-ELS	Low-ELS	High-ELS
CTQ-SF				
Physical abuse	27.96 (2.17)	43.99 (12.38)	27.80 (1.72)	41.15 (8.70)
STAI				
State anxiety	36.01 (10.76)	44.77 (11.62)	34.51 (12.68)	42.36 (12.64)
Trait anxiety	38.34 (9.75)	48.68 (11.78)	39.96 (11.13)	47.95 (11.57)
BDI-II				
BDI-II	7.56 (6.11)	16.42 (10.73)	7.09 (7.09)	12.57 (9.33)
PSS				
PSS	17.07 (6.32)	20.53 (6.51)	16.16 (6.44)	20.08 (7.31)

Mean (SD) scores on questionnaire measures for participants grouped by reported level of childhood stress exposure. ELS, Early-Life Stress; CTQ-SF, Childhood Trauma Questionnaire – Short Form (Bernstein et al., 2003); STAI, State-Trait Anxiety Inventory (Spielberger, 1983); BDI-II, Beck Depression Inventory-II (Beck et al., 1996); PSS, Perceived Stress Scale (Cohen et al., 1983).

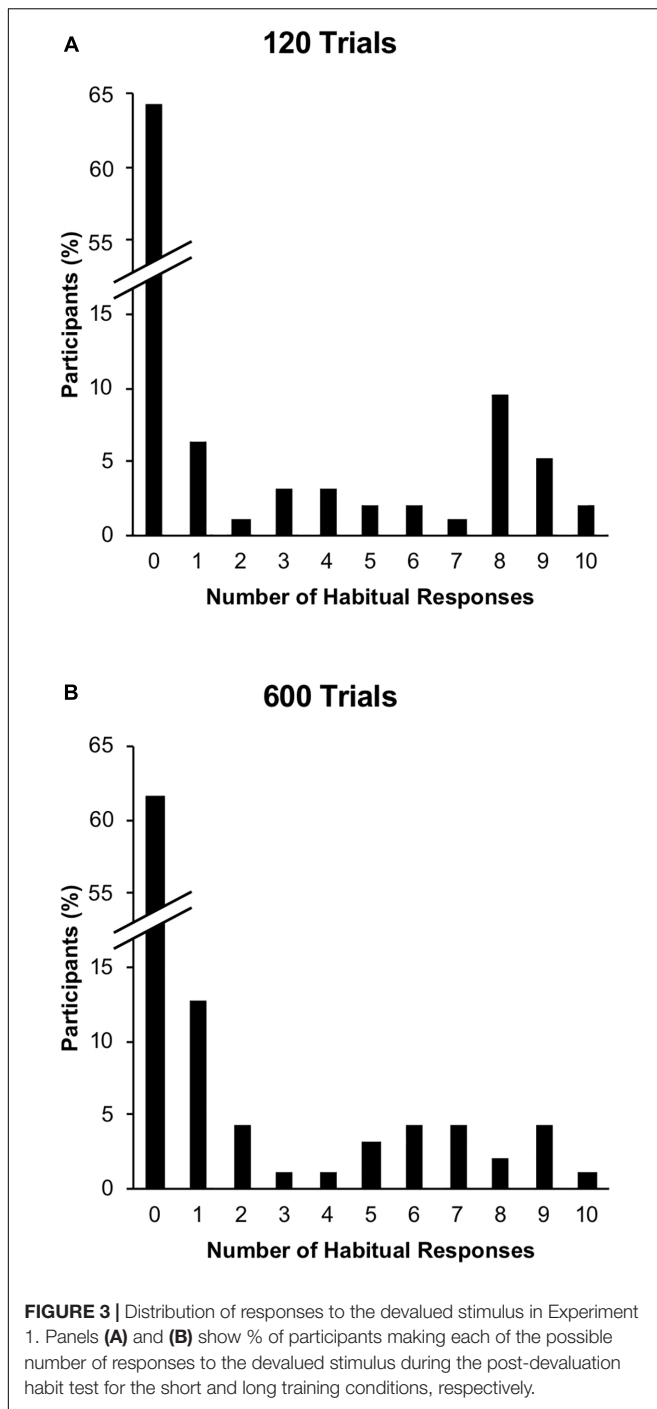
of ELS, $F(1,185) = 0.05$, $p = 0.832$, $\eta_p^2 < 0.001$, and the interaction between training and ELS was not significant, $F(1,185) < 0.01$, $p = 0.960$, $\eta_p^2 < 0.001$.

During the post-devaluation habit test, 100% of participants responded to the valued stimulus (i.e., performed the valued response in the presence of the valued stimulus), with an average response rate of 90.69% ($SD = 12.93\%$). Responding to the valued stimulus did not differ significantly across levels of training, $F(1,185) = 2.77$, $p = 0.098$, $\eta_p^2 = 0.015$, or levels of ELS, $F(1,185) = 2.43$, $p = 0.121$, $\eta_p^2 = 0.013$, and the interaction between training and ELS was not significant, $F(1,185) = 1.13$, $p = 0.289$, $\eta_p^2 = 0.006$.

The distribution of responses to the devalued stimulus (i.e., performance of the devalued response in the presence of the devalued stimulus) during the post-devaluation habit test is shown in **Figure 3**. The average response rate to the devalued stimulus was 18.57% ($SD = 30.99\%$). Participants occasionally made the valued response to the devalued stimulus (average response rate = 10.11%, $SD = 17.41\%$); these responses were not treated as habitual as they did not reflect the stimulus–response association learned during training. We tested for the effects of ELS and length of training on habitual behavior by conducting a binary logistic regression analysis on responding to the devalued stimulus during the post-devaluation habit test. Participants’ responses were binned into zero responses



to the devalued stimulus (no habitual behavior) or one or more responses to the devalued stimulus (habitual behavior). This analysis was conducted in order to test for the effect of ELS by using CTQ-SF as a continuous predictor variable while controlling for the effects of age, gender, and the other



questionnaire variables which differed across the low-ELS and high-ELS groups. We included devalued side (i.e., whether the right or left earphone was removed during the post-devaluation habit test) as a predictor because there is evidence suggesting that individuals are more likely to engage in habitual behaviors when they are using their dominant hand (Neal et al., 2011). Although we did not measure participants' handedness, it is reasonable to assume that a large majority of participants were right hand dominant and therefore might show greater habitual responding

if assigned to the condition in which the right side was devalued. We also included interaction terms to test for moderation of the effect of ELS. The results of this analysis are shown in **Table 3**. Consistent with our hypothesis of greater habitual behavior in individuals with a history of stress during development, ELS was found to be a significantly positive predictor of habitual responding, $B = 0.080$, $p = 0.020$. The odds ratio for this predictor was 1.083, meaning that for every one point increase in CTQ-SF score, the expected odds of performing a habitual response are increased by 8.3%. Contrary to our hypothesis of greater habitual responding in participants who received more training trials, level of training was not a significant predictor of habitual responding, $B = 0.015$, $p = 0.962$, and devalued side was not a significant predictor of habitual responding, $B = -0.073$, $p = 0.822$. None of the other questionnaire variables (state anxiety, trait anxiety, depression, and perceived stress) were significant predictors, smallest $p = 0.438$, and the effects of age and gender were also not significant, smallest $p = 0.367$. We did not observe evidence for moderation of the effects of ELS as is shown by the lack of significance in the interaction predictors, smallest $p = 0.092$.

EXPERIMENT 2

In Experiment 1, we found support for the hypothesis that ELS is associated with enhanced avoidance habits. Given the number of predictors included in the model, however, there is a risk that the observed effect was the result of Type I error. Therefore, in Experiment 2, we sought to replicate the effect of ELS observed in Experiment 1, and we also added a condition in which participants performed the avoidance learning task under distraction to test the hypothesis that stimulus–response associations learned under distraction would result in greater habitual responding.

Materials and Methods

Participants

As in Experiment 1, study participants were recruited from the undergraduate student population in the Psychology Department at the University of California, Los Angeles. Participants were compensated with credit toward partial fulfillment of course requirements. Study procedures were approved by the Institutional Review Board of the University of California, Los Angeles, and all participants provided written record of informed consent.

A total of 119 participants were recruited for the study. One participant failed to follow the instructions, one participant provided incomplete questionnaire data, and five participants were excluded for left-hand dominance (see the section “Design and Procedure” below), yielding a sample size of 112 (90 women, 22 men, $M_{age} = 20.54$ years, $SD_{age} = 1.59$ years, age range: 18–26 years).

Design and Procedure

Participants performed the avoidance learning task described above in Experiment 1. We manipulated the level of distraction

TABLE 3 | Summary of binary logistic regression analysis predicting responding to the devalued stimulus during the post-devaluation habit test in Experiment 1.

Variable	<i>B</i>	<i>SE</i>	Wald	<i>p</i>	Exp(<i>B</i>)
CTQ-SF	0.080	0.034	5.39	0.020	1.083
Level of training (120 trials vs. 600 trials)	0.015	0.320	0.00	0.962	1.016
Devalued side (right vs. left)	−0.073	0.326	0.05	0.822	0.929
STAI state	−0.005	0.018	0.09	0.766	0.995
STAI trait	0.013	0.025	0.27	0.601	1.013
BDI-II	−0.017	0.028	0.37	0.542	0.983
PSS	0.032	0.041	0.60	0.438	1.032
CTQ-SF × length of training	−0.045	0.031	2.07	0.150	0.956
CTQ-SF × devalued side	−0.048	0.033	2.14	0.143	0.953
CTQ-SF × STAI state	0.001	0.002	0.09	0.762	1.001
CTQ-SF × STAI trait	−0.005	0.003	2.84	0.092	0.995
CTQ-SF × BDI-II	−0.001	0.002	0.12	0.730	0.999
CTQ-SF × PSS	0.006	0.004	1.84	0.175	1.006
Age	−0.085	0.094	0.81	0.367	0.919
Gender (male vs. female)	0.166	0.405	0.17	0.681	1.181

Significant results ($p < 0.05$) shown in bold. For dichotomous predictors, the first term in parenthetical is the reference. CTQ-SF, Childhood Trauma Questionnaire – Short Form (Bernstein et al., 2003); STAI, State-Trait Anxiety Inventory (Spielberger, 1983); BDI-II, Beck Depression Inventory-II (Beck et al., 1996); PSS, Perceived Stress Scale (Cohen et al., 1983).

within subjects during the training phase of the experiment by having participants perform a counting task during alternate blocks of 30 trials. During counting blocks, participants were randomly shown an image of a dog or a cat for 500 ms after each noise avoidance trial. They were instructed to count the cats and ignore the dogs. At the end of each counting block, participants were asked to report how many cats they had counted in the previous block. Before beginning the main experiment, participants completed practice trials on both the avoidance task and the counting task, and were allowed to repeat the practice trials if desired. To minimize task difficulty, we increased the response window for the noise avoidance task from 500 to 750 ms. Six stimulus images were used for the noise avoidance task, such that the same three stimuli were shown during all counting blocks and the other three stimuli were shown during non-counting blocks. Participants completed a total of 360 training trials (180 trials per level of distraction).

The devaluation procedure was the same as in Experiment 1, except that in Experiment 2 we instructed all participants to remove the right earphone for the habit test. Although the effect of devalued side and the interaction between devalued side and ELS in Experiment 1 were not significant, there was slightly greater habitual responding in participants instructed to remove the right earphone, and the effect of ELS on habitual responding was slightly stronger among participants instructed to remove the right earphone. Therefore, in order to maximize our ability to detect an effect in Experiment 2, we screened participants for right hand dominance and then tested for habitual behavior in the right hand by having participants remove the right earphone during the devaluation procedure. The post-devaluation habit test consisted of 60 trials, 30 containing stimuli that had been learned in the no-distraction condition and 30 containing stimuli that had been learned in the distraction condition. The 60 stimuli were presented in random order. Participants were not required to perform the counting task

during the habit test. As in Experiment 1, the dependent variable of interest was whether the participant persisted in performing the response associated with avoiding aversive noise to the removed earphone.

Participants completed the experiment in a private testing room on a desktop computer. Stimulus presentation and response collection were implemented in E-Prime Standard (Version 2.0). Button press responses were made using the computer keyboard. Following completion of the computer task, participants completed the packet of questionnaires described above for Experiment 1. We additionally administered the Edinburgh Handedness Inventory (Oldfield, 1971) to screen for right hand dominance, using a cut point of 0. Seventeen participants who did not complete the handedness questionnaire were all included in the sample. The entire lab visit took approximately 1 h.

Data Analysis

Statistical analyses were performed using IBM SPSS Statistics (Version 25). Data from the acquisition phase (response accuracy to the two warning stimuli and false alarm rate to the safe stimulus) and level of responding to the valued stimulus during the habit test were analyzed using two (level of distraction: no-distraction, distraction) × two (level of ELS: low-ELS, high-ELS) mixed-model ANOVA with participants categorized as low-ELS or high-ELS based on a median split of the CTQ-SF scores. Responding to the devalued stimulus during the post-devaluation habit test was analyzed using a binary logistic regression generalized linear mixed model with level of distraction as a repeated measure. As in Experiment 1, participants' responses were binned into zero responses to the devalued stimulus (no habitual behavior) or one or more responses to the devalued stimulus (habitual behavior). The following predictors were included in the generalized linear mixed model: CTQ-SF, level of distraction, STAI state anxiety,

STAI trait anxiety, BDI-II, PSS, CTQ-SF \times level of distraction, CTQ-SF \times STAI state anxiety, CTQ-SF \times STAI trait anxiety, CTQ-SF \times BDI-II, CTQ-SF \times PSS, age, and gender. Continuous predictors used to create interaction terms were mean-centered to reduce multicollinearity, and dichotomous predictor variables were dummy coded. A significance level of 0.05 was used for all statistical tests.

A supplemental data analysis in which the number of responses to the devalued stimulus was entered as the outcome variable is provided in **Supplementary Table 2**.

Results

Sample characteristics are reported in **Table 1** (prevalence of ELS by degree and type of stress reported) and **Table 2** (scores on questionnaire variables for low-ELS and high-ELS participants). The mean CTQ-SF score was 34.83 ($SD = 9.26$) and the median CTQ-SF score was 32.00. The low-ELS group had a mean CTQ-SF score of 27.80 ($SD = 1.72$) and the high-ELS group had a mean CTQ-SF score of 41.15 ($SD = 8.70$). The high-ELS group differed significantly from the low-ELS group on measures of state anxiety, $t(110) = 3.28$, $p = 0.001$, $d = 0.62$; trait anxiety, $t(110) = 3.71$, $p < 0.001$, $d = 0.70$; depression, $t(110) = 3.47$, $p = 0.001$, $d = 0.66$; and perceived stress, $t(110) = 3.01$, $p = 0.003$, $d = 0.57$.

We first tested for effects of the level of distraction (no-distraction versus distraction) and level of ELS (low-ELS versus high-ELS) on response accuracy during training, false alarm rate during training, and level of responding to the valued stimuli during the post-devaluation habit test. The data from the training phase are shown in **Figure 4**. During training, response accuracy to the four warning stimuli was 91.74% ($SD = 7.59\%$), and the false alarm rate to the two safe stimuli was 9.45% ($SD = 19.13\%$). There was a significant effect of distraction on training accuracy, $F(1,110) = 16.05$, $p < 0.001$, $\eta_p^2 = 0.127$, such that accuracy was higher in single-task condition blocks ($M = 92.72\%$, $SD = 7.60\%$) than in dual-task condition blocks ($M = 90.76\%$, $SD = 8.41\%$). Training accuracy did not differ significantly across levels of ELS, $F(1,110) = 0.41$, $p = 0.525$, $\eta_p^2 = 0.004$, and the interaction between distraction and ELS was not significant, $F(1,110) = 0.05$, $p = 0.830$, $\eta_p^2 < 0.001$. False alarm rate did not differ significantly across levels of distraction, $F(1,110) = 0.89$, $p = 0.349$, $\eta_p^2 = 0.008$, or levels of ELS, $F(1,110) = 2.76$, $p = 0.099$, $\eta_p^2 = 0.024$, and the interaction between distraction and ELS was not significant, $F(1,110) = 0.22$, $p = 0.637$, $\eta_p^2 = 0.002$.

During the post-devaluation habit test, 100% of participants responded to the valued stimuli (i.e., performed the valued response in the presence of the valued stimuli), with an average response rate of 92.99% ($SD = 10.75\%$). Responding to valued stimuli did not differ significantly across levels of distraction, $F(1,110) = 0.07$, $p = 0.791$, $\eta_p^2 = 0.001$, or levels of ELS, $F(1,110) = 0.08$, $p = 0.773$, $\eta_p^2 = 0.001$, and the interaction between distraction and ELS was not significant, $F(1,110) = 1.56$, $p = 0.214$, $\eta_p^2 = 0.014$.

The distribution of responses to the devalued stimuli (i.e., performance of the devalued response in the presence of the devalued stimuli) during the post-devaluation habit test is shown

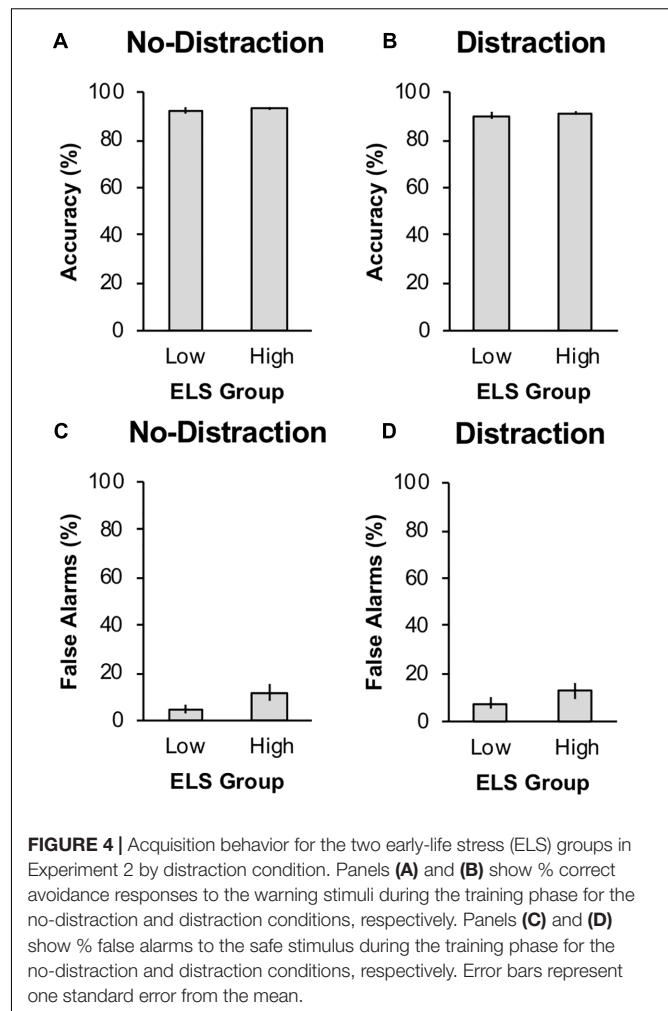
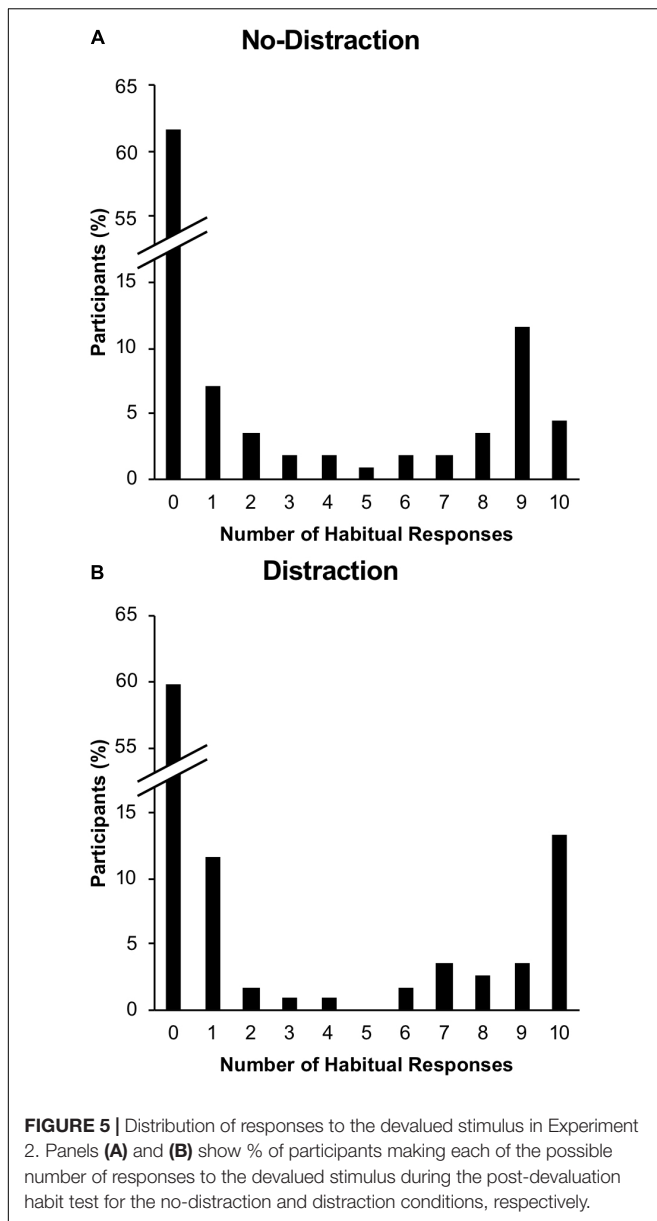


FIGURE 4 | Acquisition behavior for the two early-life stress (ELS) groups in Experiment 2 by distraction condition. Panels (A) and (B) show % correct avoidance responses to the warning stimuli during the training phase for the no-distraction and distraction conditions, respectively. Panels (C) and (D) show % false alarms to the safe stimulus during the training phase for the no-distraction and distraction conditions, respectively. Error bars represent one standard error from the mean.

in **Figure 5**. The average response rate to the devalued stimuli was 23.84% ($SD = 36.91\%$). Participants occasionally made the valued response to the devalued stimuli (average response rate = 9.87%, $SD = 18.16\%$); these responses were not treated as habitual as they did not reflect the stimulus-response associations learned during training. We tested for the effects of ELS and distraction on habitual behavior by conducting a binary logistic regression generalized linear mixed model analysis on responding to the devalued stimulus during the post-devaluation habit test. As in Experiment 1, participants' responses were binned into zero responses to the devalued stimulus (no habitual behavior) or one or more responses to the devalued stimulus (habitual behavior). The results of this analysis are shown in **Table 4**. Consistent with Experiment 1, ELS was found to be a significantly positive predictor of habitual responding, $B = 0.064$, $p = 0.022$. The odds ratio for this predictor was 1.066, meaning that for every one point increase in CTQ-SF score, the expected odds of performing a habitual response are increased by 6.6%. Contrary to our hypothesis of greater habitual responding to stimuli that were trained in the presence of distraction, level of distraction was not a significant predictor of habitual responding, $B = 0.112$, $p = 0.723$. Of the other questionnaire variables included as



predictors (state anxiety, trait anxiety, depression, and perceived stress), two were significantly positive predictors of habitual responding: state anxiety, $B = 0.047$, $p = 0.018$ and perceived stress, $B = 0.093$, $p = 0.025$. For state anxiety, the odds ratio was 1.048, meaning that for every one point increase in the STAI state anxiety score, the expected odds of performing a habitual response are increased by 4.8%. For perceived stress, the odds ratio was 1.097, meaning that for every one point increase in the PSS score, the expected odds of performing a habitual response are increased by 9.7%. The other questionnaire variables were not significant predictors, smallest $p = 0.209$, and the effects of age and gender were also not significant, smallest $p = 0.352$. We did not observe evidence for moderation of the effects of ELS as is shown by the lack of significance in the interaction predictors, smallest $p = 0.143$.

DISCUSSION

In two experiments using an avoidance learning task, we observed evidence of enhanced avoidance habits in adults who reported a history of ELS. An important implication of this finding is that this behavioral tendency may contribute to the negative health outcomes commonly experienced by individuals with a history of ELS. Some of the negative health outcomes associated with self-reported developmental stress include severe obesity (Anda et al., 2006), heart disease (Dong et al., 2004), liver disease (Dong et al., 2003), and sexually transmitted disease (Hillis et al., 2000). Negative health outcomes are frequently tied to negative health behaviors, which may be performed habitually. Some of the negative health behaviors associated with self-reported developmental stress that contribute to the aforementioned negative health outcomes include smoking (Anda et al., 1999), alcohol abuse (Dube et al., 2002), and risky sexual behavior (Hillis et al., 2001). These negative health behaviors, along with the overeating that contributes to severe obesity and obesity-related health outcomes, can be conceptualized as avoidance behaviors, which over time can become avoidance habits. For example, individuals may initially engage in overeating, substance use, or risky sexual behavior in a goal-directed manner to avoid feelings of distress, but over time these behaviors may become more automatic and stimulus-bound. It should be noted, however, that such behaviors have an appetitive aspect to them as well; understanding the relationship between ELS and negative health outcomes may require a model in which behavior is driven by both appetitive and avoidant motivations, such as in Baumeister's (1991) "escape from the self" theory of alcoholism.

One question raised by this pair of experiments is whether ELS is linked specifically to avoidance habits as opposed to avoidance behavior. Although we did not observe differences in avoidance behavior between the low-ELS and high-ELS participants during training, such differences may exist. Because the stimulus-response-outcome contingencies were demonstrated to participants explicitly at the beginning of the training phase rather than learned through experience, we may have had limited sensitivity to detect differences in the initial learning of the associations. This question could be tested in future research.

A possible biological basis for enhanced habitual behavior following ELS is that stress selectively compromises the neural structures that support goal-directed behavior, which could lead to a compensatory over-reliance on habitual responding. Goal-directed behavior relies on prefrontal cortex, dorsomedial striatum, and the hippocampus, which have been shown to atrophy following stress exposure (McEwen, 2000; Joëls et al., 2007; Dias-Ferreira et al., 2009; Soares et al., 2012). Habitual behavior, on the other hand, appears to rely on the dorsolateral striatum (Yin and Knowlton, 2004; Yin et al., 2004, 2006), which is less sensitive to stress and indeed has been shown in some cases to undergo stress-induced hypertrophy (Dias-Ferreira et al., 2009; Soares et al., 2012). The extent to which these morphological changes are reversible is not known. The presence of significant stress during a sensitive period of development may crystallize these dynamics, setting the stage

TABLE 4 | Summary of binary logistic regression generalized linear mixed model analysis predicting responding to the devalued stimulus during the post-devaluation habit test in Experiment 2.

Variable	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	Exp(<i>B</i>)
CTQ-SF	0.064	0.028	2.31	0.022	1.066
Distraction (no distraction vs. distraction)	0.112	0.315	0.36	0.723	1.119
STAI state	0.047	0.020	2.39	0.018	1.048
STAI trait	−0.033	0.026	−1.26	0.209	0.968
BDI-II	−0.003	0.026	−0.13	0.899	0.997
PSS	0.093	0.041	2.25	0.025	1.097
CTQ-SF × distraction	−0.016	0.035	−0.45	0.655	0.984
CTQ-SF × STAI state	−0.003	0.002	−1.47	0.143	0.997
CTQ-SF × STAI trait	−0.004	0.003	−1.18	0.240	0.996
CTQ-SF × BDI-II	−0.001	0.003	−0.23	0.817	0.999
CTQ-SF × PSS	0.000	0.005	0.07	0.946	1.000
Age	−0.004	0.103	−0.04	0.966	0.996
Gender (male vs. female)	−0.384	0.412	−0.93	0.352	0.681

Significant results ($p < 0.05$) shown in bold. For dichotomous predictors, the first term in parenthetical is the reference. CTQ-SF, Childhood Trauma Questionnaire – Short Form (Bernstein et al., 2003); STAI, State-Trait Anxiety Inventory (Spielberger, 1983); BDI-II, Beck Depression Inventory-II (Beck et al., 1996); PSS, Perceived Stress Scale (Cohen et al., 1983).

for an overreliance on habitual responding in adulthood. Some evidence supporting this hypothesis includes the finding that male rats exposed to maternal separation during the first 2 weeks of life are more likely to use a stimulus–response navigation strategy in early adolescence (Grissom et al., 2012), and humans exposed to stress prenatally are more likely to use a stimulus–response navigation strategy in adulthood (Schwabe et al., 2012). Future research incorporating neuroimaging of habit learning in the ELS population should investigate this possibility. Recent neuroimaging studies that target the neuroendocrine basis of the stress-induced shift toward habitual behavior are helping to elucidate the mechanisms that underlie this shift (for review, see Wirz et al., 2018); it would be interesting to see how the effects of acute stress on habit compare to the effects of ELS on habit at the neural level.

An additional finding of the present study is that in Experiment 2 we also observed enhanced avoidance habits in individuals who reported higher levels of state anxiety and higher levels of perceived stress during the past month. This finding is consistent with previous literature on stress and habitual behavior (e.g., Schwabe and Wolf, 2009, 2010, 2012), but to our knowledge an effect of stress has not previously been demonstrated with avoidance habits. However, since this result was only present in Experiment 2 and not in Experiment 1, further research should be done to confirm the finding.

In addition to providing support for the hypothesis that ELS alters the tendency toward habitual responding, the results of the present study also demonstrate the utility of avoidance learning tasks in human habit research. Research on habits in humans has traditionally been carried out in appetitive situations with participants working for monetary rewards, points, or food (e.g., Tricomi et al., 2009), but tasks employing aversive stimuli have a long history of success in the non-human animal habit learning literature, particularly in maze navigation tasks where animals are motivated to escape a negative situation such as a water tank or open surface (e.g., Packard and McGaugh, 1992;

McDonald and White, 1994). However, it should be noted that these maze navigation studies differ from the present study in that they do not use outcome devaluation to test habitual behavior. Aversive stimuli like the scream sound used in the present study are not difficult to incorporate into computer-based tasks and may provide greater motivation than appetitive stimuli.

Two hypotheses that we made in this pair of experiments were not borne out by the results. In Experiment 1, we predicted that a longer period of training would result in greater habitual responding, and in Experiment 2, we predicted that distraction during training would result in greater habitual responding. Neither of these manipulations affected the level of habit formation as measured by our post-devaluation habit test. It is possible that the manipulations we employed were not effective because the manipulations were not strong enough. Our manipulation of amount of training was a fivefold increase in the number of training trials, but participants in the long training condition still received only a single training session, and it is possible that to see an effect of training, multiple sessions would be required. A previous study conducted with appetitive stimuli that showed an effect of level of training on habitual responding implemented 12 training sessions over the course of 3 days (Tricomi et al., 2009); an avoidance learning study with a similar amount of training across multiple days may reveal a relationship between level of training and habitual responding. Similarly, the distraction task we used may have failed to provide enough of a challenge to produce the distraction-induced increase in habitual responding observed in previous studies (Foerde et al., 2006, 2007).

On the other hand, our failure to find an effect of length of training on habitual responding is consistent with a recent series of experiments conducted by de Wit et al. (2018), in which length of training was manipulated across a variety of tasks and in each case extended training failed to produce greater habitual responding. Notably, the noise avoidance procedure used in our experiments was very similar to the noise avoidance procedure

used in one of the experiments conducted by de Wit et al. (2018); therefore, the null result of extended training in the present study serves as a replication of the null result of extended training reported in de Wit et al.'s (2018) noise avoidance experiment. The response rate to the devalued stimuli in our experiments was somewhat higher than the response rate to the devalued stimuli observed in the de Wit et al. (2018) noise avoidance experiment [approximately 20% in our experiments versus approximately 10% in the de Wit et al. (2018) experiment]. This difference may be due to the fact that responses in the de Wit et al. (2018) noise avoidance experiment were performed with a foot pedal whereas our participants performed responses with their index fingers on a computer keyboard.

An area of future research suggested by the present study is whether ELS affects habit learning, habit performance, or both. Previous research employing acute stress and challenges to executive control has indicated that these factors affect both the learning and performance of habits. For example, studies using the probabilistic classification task have shown that acute stress and distraction modulate which memory system is engaged during classification learning, biasing competition between the declarative memory system and the habit learning system in favor of habit learning (Foerde et al., 2006, 2007; Schwabe and Wolf, 2012). In contrast, studies that induce acute stress or executive challenge after learning but before a habit test demonstrate how these factors influence the performance of habits that have been learned previously. For example, Schwabe and Wolf (2010) showed that acute stress after learning decreases sensitivity to devaluation, and Lin et al. (2016) showed that completion of a task designed to deplete executive resources after learning an unhealthy habit increased performance of the unhealthy habit. Of course, as ELS cannot be induced between learning and testing, the paradigms that have been used to investigate performance effects of acute stress and executive challenge cannot be applied, and different paradigms possibly incorporating neuroimaging will be necessary to tease apart the effects of ELS on habit learning versus habit performance.

One limitation of the present study is that because we used a college sample, our ELS groups may be more high-functioning and resilient to stress than individuals with a history of ELS in the general population. Nevertheless, even this sample yielded evidence in support of our hypothesis that ELS affects avoidance habit formation. Future research with a more representative sample would, however, yield important information about the generalizability of our findings and typical effect sizes. A second limitation is that our sample was primarily composed of young adult females. Neither age nor sex were found to be significant predictors of habitual responding in this set of experiments; however, the age range of participants in our sample was relatively limited and the sample of male participants was relatively small. Future studies should investigate whether these variables are truly non-significant by testing a wider age range and sampling a larger number of males.

Our findings extend recent work demonstrating enhanced avoidance habits in individuals with obsessive-compulsive

disorder (Gillan et al., 2014, 2015), identifying a second population with this behavioral pattern. Additional populations that may show similar patterns include individuals with post-traumatic stress disorder, binge eating disorder, and substance use disorders. Future research should investigate these possibilities. A deeper understanding of the role of avoidance habits in maladaptive behavior has the potential to inform interventions that may mitigate their negative effects on individuals' lives.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Institutional Review Board of the University of California, Los Angeles with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Institutional Review Board of the University of California, Los Angeles.

AUTHOR CONTRIBUTIONS

TP, MC, and BK contributed to the conception and design of the study. TP oversaw data acquisition, performed the statistical analysis, and wrote the first draft of the manuscript. All authors contributed to the manuscript revision, and read and approved the submitted version of the manuscript.

FUNDING

This research was supported by the National Science Foundation Graduate Research Fellowship (DGE1144087), an institutional training fellowship from the National Institutes of Health (T32MH096682), and a grant from the National Institutes of Health (R01DA045716).

ACKNOWLEDGMENTS

This manuscript is based on a chapter from the first author's dissertation (Patterson, 2018). The authors thank Ling Lee Chong, Zhixi Liu, and Alexander Gordon for research assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01876/full#supplementary-material>

REFERENCES

- Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q. J. Exp. Psychol. Section B* 34B, 77–98. doi: 10.1080/14640748208400878
- Anda, R. F., Croft, J. B., Felitti, V. J., Nordenberg, D., Giles, W. H., Williamson, D. F., et al. (1999). Adverse childhood experiences and smoking during adolescence and adulthood. *JAMA* 282, 1652–1658. doi: 10.1001/jama.282.17.1652
- Anda, R. F., Felitti, V. J., Bremner, J. D., Walker, J. D., Whitfield, C., Perry, B. D., et al. (2006). The enduring effects of abuse and related adverse experiences in childhood: a convergence of evidence from neurobiology and epidemiology. *Eur. Arch. Psychiatry Clin. Neurosci.* 256, 174–186. doi: 10.1007/s00406-005-0624-4
- Baumeister, R. F. (1991). *Escaping the Self: Alcoholism, Spirituality, Masochism, and Other Flights from the Burden of Selfhood*. New York, NY: Basic Books.
- Beck, A. T., Steer, R. A., and Brown, G. K. (1996). *Beck Depression Inventory*, 2 Edn. San Antonio, TX: The Psychological Corporation.
- Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., et al. (2003). Development and validation of a brief screening version of the childhood trauma questionnaire. *Child Abuse Negl.* 27, 169–190. doi: 10.1016/S0145-2134(02)00541-0
- Britton, J. C., Lissek, S., Grillon, C., Norcross, M. A., and Pine, D. S. (2011). Development of anxiety: the role of threat appraisal and fear learning. *Depress. Anxiety* 28, 5–17. doi: 10.1002/da.20733
- Chapman, D. P., Whitfield, C. L., Felitti, V. J., Dube, S. R., Edwards, V. J., and Anda, R. F. (2004). Adverse childhood experiences and the risk of depressive disorders in adulthood. *J. Affect. Disord.* 82, 217–225. doi: 10.1016/j.jad.2003.12.013
- Cohen, S., Kamarck, T., and Mermelstein, R. (1983). A global measure of perceived stress. *J. Health Soc. Behav.* 24, 385–396. doi: 10.2307/2136404
- de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., et al. (2018). Shifting the balance between goals and habits: five failures in experimental habit induction. *J. Exp. Psychol. Gen.* 147, 1043–1065. doi: 10.1037/xge0000402
- Dias-Ferreira, E., Sousa, J. C., Melo, I., Morgado, P., Mesquita, A. R., Cerqueira, J. J., et al. (2009). Chronic stress causes frontostriatal reorganization and affects decision-making. *Science* 325, 621–625. doi: 10.1126/science.1171203
- Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philos. Trans. Royal Soc. Lon., Series B Biol. Sci.* 308, 67–78. doi: 10.1098/rstb.1985.0010
- Dong, M., Dube, S. R., Felitti, V. J., Giles, W. H., and Anda, R. F. (2003). Adverse childhood experiences and self-reported liver disease: new insights into the causal pathway. *Arch. Intern. Med.* 163, 1949–1956. doi: 10.1001/archinte.163.16.1949
- Dong, M., Giles, W. H., Felitti, V. J., Dube, S. R., Williams, J. E., Chapman, D. P., et al. (2004). Insights into causal pathways for ischemic heart disease: adverse childhood experiences study. *Circulation* 110, 1761–1766. doi: 10.1161/01.CIR.0000143074.54995.7F
- Dube, S. R., Anda, R. F., Felitti, V. J., Edwards, V. J., and Croft, J. B. (2002). Adverse childhood experiences and personal alcohol abuse as an adult. *Addict. Behav.* 27, 713–725. doi: 10.1016/S0306-4603(01)00204-0
- Foerde, K., Knowlton, B. J., and Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proc. Natl. Acad. Sci., U.S.A.* 103, 11778–11783. doi: 10.1073/pnas.0602659103
- Foerde, K., Poldrack, R. A., and Knowlton, B. J. (2007). Secondary-task effects on classification learning. *Mem. Cogn.* 35, 864–874. doi: 10.3758/bf03193461
- Gillan, C. M., Apergis-Schoute, A. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Fineberg, N. A., et al. (2015). Functional neuroimaging of avoidance habits in obsessive-compulsive disorder. *Am. J. Psychiatr.* 172, 284–293. doi: 10.1176/appi.ajp.2014.14040525
- Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Apergis-Schoute, A. M., et al. (2014). Enhanced avoidance habits in obsessive-compulsive disorder. *Biol. Psychiatr.* 75, 631–638. doi: 10.1016/j.biopsych.2013.02.002
- Grissom, E. M., Hawley, W. R., Bromley-Dulfano, S. S., Marino, S. E., Stathopoulos, N. G., and Dohanich, G. P. (2012). Learning strategy is influenced by trait anxiety and early rearing conditions in prepubertal male, but not prepubertal female rats. *Neurobiol. Learn. Mem.* 98, 174–181. doi: 10.1016/j.nlm.2012.06.001
- Hillis, S. D., Anda, R. F., Felitti, V. J., and Marchbanks, P. A. (2001). Adverse childhood experiences and sexual risk behaviors in women: a retrospective cohort study. *Fam. Plann. Perspect.* 33, 206–211. doi: 10.2307/2673783
- Hillis, S. D., Anda, R. F., Felitti, V. J., Nordenberg, D., and Marchbanks, P. A. (2000). Adverse childhood experiences and sexually transmitted diseases in men and women: a retrospective study. *Pediatrics* 106 e11, 1–6. doi: 10.1542/peds.106.1.e11
- Joëls, M., Karst, H., Krugers, H. J., and Lucassen, P. J. (2007). Chronic stress: implications for neuronal morphology, function and neurogenesis. *Front. Neuroendocrinol.* 28, 72–96. doi: 10.1016/j.yfrne.2007.04.001
- Kim, J. J., Lee, H. J., Han, J.-S., and Packard, M. G. (2001). Amygdala is critical for stress-induced modulation of hippocampal long-term potentiation and learning. *J. Neurosci.* 21, 5222–5228. doi: 10.1523/jneurosci.21-14-05222.2001
- Knowlton, B. J., Mangels, J. A., and Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science* 273, 1399–1402. doi: 10.1126/science.273.5280.1399
- Knowlton, B. J., and Patterson, T. K. (2018). “Habit formation and the striatum,” in *Behavioral Neuroscience of Learning and Memory. Current Topics in Behavioral Neurosciences* Vol. 37, eds R. E. Clark and S. J. Martin (Cham: Springer), 275–295. doi: 10.1007/7854_2016_451
- Knowlton, B. J., Squire, L. R., and Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learn. Mem.* 1, 106–120. doi: 10.1101/lm.1.2.106
- Lau, J. Y. F., Lissek, S., Nelson, E. E., Lee, Y., Roberson-Nay, R., Poeth, K., et al. (2008). Fear conditioning in adolescents with anxiety disorders: results from a novel experimental paradigm. *J. Am. Acad. Child Adolesc. Psychiatr.* 47, 94–102. doi: 10.1097/chi.0b013e31815a5f01
- Lin, P.-Y., Wood, W., and Monterosso, J. (2016). Healthy eating habits protect against temptations. *Appetite* 103, 432–440. doi: 10.1016/j.appet.2015.11.011
- McDonald, R. J., and White, N. M. (1994). Parallel information processing in the water maze: evidence for independent memory systems involving dorsal striatum and hippocampus. *Behav. Neural Biol.* 61, 260–270. doi: 10.1016/S0163-1047(05)80009-3
- McEwen, B. S. (2000). Effects of adverse experiences for brain structure and function. *Biol. Psychiatr.* 48, 721–731. doi: 10.1016/S0006-3223(00)00964-1
- Neal, D. T., Wood, W., Wu, M., and Kurlander, D. (2011). The pull of the past: when do habits persist despite conflict with motives? *Personal. Soc. Psychol. Bull.* 37, 1428–1437. doi: 10.1177/0146167211419863
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Packard, M. G., and McGaugh, J. L. (1992). Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: further evidence for multiple memory systems. *Behav. Neurosci.* 106, 439–446. doi: 10.1037/0735-7044.106.3.439
- Patterson, T. K. (2018). *Effects of Early-Life Stress on Actions, Habits, and the Neural Systems Supporting Instrumental Behavior*. Los Angeles, CA: University of California, Los Angeles. [dissertation].
- Schwabe, L., Bohbot, V. D., and Wolf, O. T. (2012). Prenatal stress changes learning strategies in adulthood. *Hippocampus* 22, 2136–2143. doi: 10.1002/hipo.22034
- Schwabe, L., Dalm, S., Schächinger, H., and Oitzl, M. S. (2008). Chronic stress modulates the use of spatial and stimulus-response learning strategies in mice and man. *Neurobiol. Learn. Mem.* 90, 495–503. doi: 10.1016/j.nlm.2008.07.015
- Schwabe, L., and Wolf, O. T. (2009). Stress prompts habit behavior in humans. *J. Neurosci.* 29, 7191–7198. doi: 10.1523/JNEUROSCI.0979-09.2009
- Schwabe, L., and Wolf, O. T. (2010). Socially evaluated cold pressor stress after instrumental learning favors habits over goal-directed action. *Psychoneuroendocrinology* 35, 977–986. doi: 10.1016/j.psyneuen.2009.12.010
- Schwabe, L., and Wolf, O. T. (2012). Stress modulates the engagement of multiple memory systems in classification learning. *J. Neurosci.* 32, 11042–11049. doi: 10.1523/jneurosci.1484-12.2012
- Seehagen, S., Schneider, S., Rudolph, J., Ernst, S., and Zmyj, N. (2015). Stress impairs cognitive flexibility in infants. *Proc. Natl. Acad. Sci. U.S.A.* 112, 12882–12886. doi: 10.1073/pnas.1508345112
- Soares, J. M., Sampaio, A., Ferreira, L. M., Santos, N. C., Marques, F., Palha, J. A., et al. (2012). Stress-induced changes in human decision-making are reversible. *Trans. Psychiatr.* 2 e131, 1–7. doi: 10.1038/tp.2012.59

- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Menlo Park, CA: Mind Garden.
- Tricomi, E., Balleine, B. W., and O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29, 2225–2232. doi: 10.1111/j.1460-9568.2009.06796.x
- Wirz, L., Bogdanov, M., and Schwabe, L. (2018). Habits under stress: mechanistic insights across different types of learning. *Curr. Opin. Behav. Sci.* 20, 9–16. doi: 10.1016/j.cobeha.2017.08.009
- Yin, H. H., and Knowlton, B. J. (2004). Contributions of striatal subregions to place and response learning. *Learn. Mem.* 11, 459–463. doi: 10.1101/lm.81004
- Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19, 181–189. doi: 10.1111/j.1460-9568.2004.03095.x
- Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action–outcome contingency in instrumental conditioning. *Behav. Brain Res.* 166, 189–196. doi: 10.1016/j.bbr.2005.07.012
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2019 Patterson, Craske and Knowlton. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Expected Value of Control and the Motivational Control of Habitual Action

Andreas B. Eder^{1*} and David Dignath²

¹Department of Psychology, Julius Maximilian University of Würzburg, Würzburg, Germany, ²Department of Psychology, University of Freiburg, Freiburg, Germany

OPEN ACCESS

Edited by:

John A. Bargh,
Yale University, United States

Reviewed by:

Ludise Malkova,
Georgetown University, United States
Alexander Soutschek,
University of Zurich, Switzerland
Henk Aarts,
Utrecht University, Netherlands

*Correspondence:

Andreas B. Eder
andreas.eder@uni-wuerzburg.de

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 29 March 2019

Accepted: 22 July 2019

Published: 13 August 2019

Citation:

Eder AB and Dignath D (2019)
Expected Value of Control and the
Motivational Control of Habitual Action.
Front. Psychol. 10:1812.
doi: 10.3389/fpsyg.2019.01812

A hallmark of habitual actions is that, once they are established, they become insensitive to changes in the values of action outcomes. In this article, we review empirical research that examined effects of posttraining changes in outcome values in outcome-selective Pavlovian-to-instrumental transfer (PIT) tasks. This review suggests that cue-instigated action tendencies in these tasks are not affected by weak and/or incomplete revaluation procedures (e.g., selective satiety) and substantially disrupted by a strong and complete devaluation of reinforcers. In a second part, we discuss two alternative models of a motivational control of habitual action: a default-interventionist framework and expected value of control theory. It is argued that the default-interventionist framework cannot solve the problem of an infinite regress (i.e., what controls the controller?). In contrast, expected value of control can explain control of habitual actions with local computations and feedback loops without (implicit) references to control homunculi. It is argued that insensitivity to changes in action outcomes is not an intrinsic design feature of habits but, rather, a function of the cognitive system that controls habitual action tendencies.

Keywords: habit, outcome devaluation, Pavlovian-to-instrumental transfer, default-interventionist framework, expected value of control, cognitive control

“The chains of habit are too weak to be felt until they are too strong to be broken.” (adage credited to Samuel Johnson, 1748, “The vision of Theodore”)

Human beings like to view themselves as rationally behaving agents (Nisbett and Wilson, 1977). Yet, we are also creatures of habit. Accordingly, scientists in many different fields have been attracted to the study of habits because they invoke a dichotomy between automatic and controlled behavior (Wood and Rünger, 2016). A popular view is that habits run on autopilot until something goes wrong. For an illustration, let us take the example of our fictitious friend Tom: when he comes home from work, he has the habit to grab a can of cold beer from the fridge and to enjoy his after-work beer. On one unfortunate day, his wife bought the wrong beer, and the drama unfolds: Tom takes his usual large gulp, grimaces in distaste, and the moment is spoiled. What will happen to Tom? Will he continue with drinking, even if he cannot have his favorite beer? Maybe at a reduced rate? Or does he stop beer drinking all at once?

These questions are far from trivial, because behavior analysts commonly agree that habitual action is in principle and by definition independent of the current value of the produced outcome (see the next section). Yet, it is also clear that most people can control and correct habitual actions to some degree if the outcome is dysfunctional. In fact, a persistent inability

to correct for unwanted habitual action patterns is a hallmark of a variety of pathological states (e.g., addiction)—and hence the atypical outcome of action control in healthy adults.

This article reviews research on the motivational control of habitual action. In a first section, we will discuss insensitivity to changes in action outcomes as a defining feature of habitual actions. Then, we will review behavioral and neuroscientific studies that examined a goal-independency of cue-instigated action tendencies with posttraining outcome revaluation procedures in operant learning and outcome-selective Pavlovian-to-instrumental transfer (PIT) tasks. In the second part, we will discuss two theoretical accounts: a default-interventionist framework and expected value of control (EVC) theory. While both accounts can explain a motivational control of habitual action, we will argue that EVC theory has more potential to provide a convincing account of habit control in PIT tasks.

PART I

Dual Action Psychology: Habitual and Goal-Directed Actions

According to behavior analysts, a *habit* is an acquired behavior that is triggered by an antecedent stimulus (Dickinson, 1985). Habit is distinguished from goal-directed action that is controlled by the current value of the action goal through knowledge about the instrumental relations between the action and its consequences. Often implicit to this distinction is an assignment of features of automaticity (e.g., associative, unintentional, efficient, etc.) to habitual actions and features of non-automaticity (e.g., rule-based, intentional, capacity-limited, etc.) to goal-directed actions (Dickinson and Balleine, 1993). However, close scrutiny of this distinction makes clear that this dichotomy is not justified and too simple (for thorough discussions, see Bargh, 1994; Moors and De Houwer, 2006; Keren and Schul, 2009; for counterarguments, see Evans and Stanovich, 2013). More useful seems a functional distinction based on correlations between actions and context features and correlations between actions and valued outcomes: instrumental actions are goal-directed because they are correlated more strongly with the presence or absence of desired outcomes than with the presence of particular contexts or stimuli. For example, if Tom drinks his after-work beer because he has a desire to get drunk, he would be willing to consume another alcoholic beverage if it has the same intoxicant effect. Habitual action, by contrast, is correlated more strongly with context features than with the presence or absence of a particular outcome of the action. For example, Tom would drink his after-work beer even if he is not thirsty or keen on getting drunk. For him, it is a behavioral routine that becomes activated in the appropriate context. That means, he would not have drunk the beer at another time or place, and assuming that he has developed a habit of beer drinking, even not another beverage.

At this point, a few additional qualifications are necessary. First, the correlation of habitual actions with particular contexts (or states) does not mean that they are unrelated to the value of these contexts. Habits typically arise from frequent repetitions

of previously rewarded (instrumental) actions, that means, they often have a strong reward history (Yin and Knowlton, 2006). This rewarding context does not change with the performance of a habitual action (“Tom still gets drunk after beer consumption”) but, rather, the internal representation of this state as action outcome has changed (“getting drunk is a by-product and not an intended consequence anymore”). Complicating things further, a similar point can be made in respect to a correlation between instrumental actions and context features. Goal-directed actions are situated in particular contexts that offer a variety of informative cues for action control. Organisms exploit these cues in their active pursuit of a valued outcome and, if encountered on a regular basis, the action is correlated with the presence and absence of these contextual cues. Taken together, this means that a functional distinction between habitual and goal-directed actions based on the relative strength of correlations is gradual—and not a categorical one.

Second, for the analysis of a goal-dependency of actions, it is meaningful to distinguish between proximal and distal outcomes of actions. According to the standard definition, habitual action is not controlled by the value of proximal outcomes (“Tom does not drink his after-work beer because of the good taste of the beer”); however, the context in which the habit is performed is controlled by outcomes that are more distally related to the habitual behavior (e.g., “Tom wants to enjoy his leisure time and beer drinking serves this goal”). Thus, distal consequences can be causally involved in the performance of a habitual action even if its performance is insensitive to its immediate outcome. Note that this relationship implies a roughly hierarchical structure in which the habit (“beer drinking”) is nested in a more abstract and/or temporally extended activity (“enjoyment of the evening”). In the following, we mean an insensitivity to immediate outcomes when referring to a goal-independency of habits.

Goal-Independency of Habits

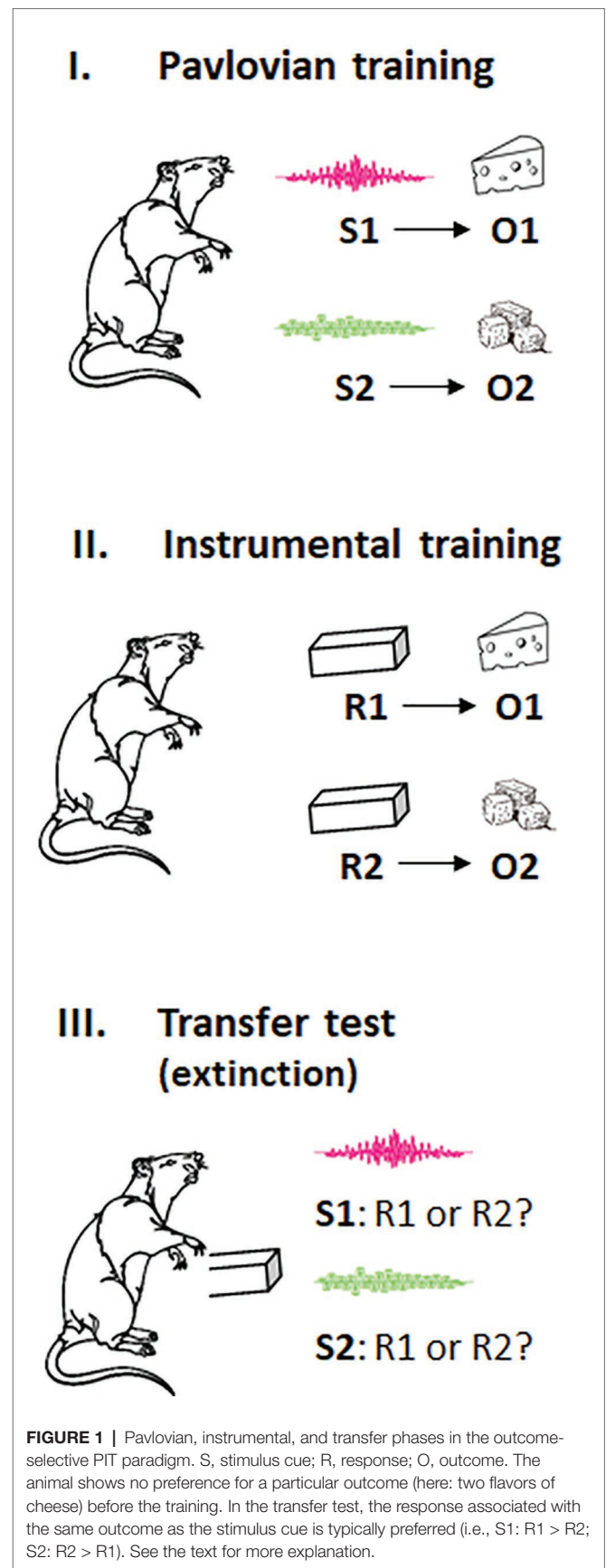
Having laid out what habitual actions are, we now discuss studies examining a goal-independency of habitual actions. Given the extensive research literature on habit acquisition and performance, this review is necessarily selective. In the following, we will focus on laboratory studies with humans and animals in which reinforcing stimuli were devalued after extensive instrumental training. For example, devaluation treatments could be the pairing of a food reinforcer with toxin, or the devaluation of a monetary reinforcer. Critically, this devaluation was done after reinforcement learning; consequently, the value of the reinforcer was changed in the absence of the associated action. Following devaluation, action performance was tested in extinction (i.e., without presentation of the reinforcer that would have allowed for new reinforcement learning). If the animal or human continued to perform the behavior which had produced the now-devalued reinforcer, it was concluded that the motivation to perform this action was not driven by the current value of the reinforcer (i.e., action outcome)—and hence habitual.

First, it should be noted that many studies with posttraining devaluations of action outcomes found that actions do *not* become habitual even after extensive training

(e.g., Adams and Dickinson, 1981). For example, a classic study trained rats to perform two distinct actions, each reinforced by a unique food reward (Colwill and Rescorla, 1985). After extensive training, one reward was devalued by pairing it with a toxin (flavor-aversion conditioning). Then, the animal was given the opportunity to engage in each of the responses in extinction. The study showed that the postlearning devaluation of the food reinforcer selectively reduced working for that food. Obviously, the rat had retrieved a memory of the devalued food outcome during the extinction test, in contradiction to early views that the reinforcer becomes not encoded in associative stimulus-response structures controlling reinforced behaviors (Thorndike, 1911; Hull, 1931). On the other hand, working for the devalued outcome was often not completely abolished in this research, which was viewed as evidence for habit formation. However, caution is warranted with this interpretation. First, other factors besides context features could have motivated the residual performance. For instance, the animal could have tested out whether the action will continue to produce no reinforcer in the extinction test (see research on the so-called “extinction burst”; Lerman and Iwata, 1995). Second, the devaluation of the reinforcer was most typically incomplete (Colwill and Rescorla, 1990). We will come back to this issue when we discuss the effectiveness of outcome devaluation treatments below.

Subsequent studies examined more specific conditions in which instrumental performance becomes insensitive to outcome values. This research suggested that overtraining, single-response training regimes, and interval-based reinforcement schedules (relative to a fixed-ratio schedule) are conducive to habit formation (e.g., Dickinson et al., 1983; Tricomi et al., 2009; Kosaki and Dickinson, 2010). However, even these protocols do not invariably lead to an insensitivity outcome values (for a recent failure, see de Wit et al., 2018) and the conditions necessary for habit formation are still not very well understood (Hogarth, 2018). Most important, the ideal “habit test” examines not only an insensitivity to correlations with (de)valued outcomes but also a sensitivity to correlations with context features. This test is found in a procedure called *outcome-selective Pavlovian-to-instrumental transfer of control* (PIT).

In outcome-selective PIT, stimuli that are predictive of specific outcomes prime instrumental responses that are associated with these outcomes. The canonical procedure is shown in **Figure 1** and consists of three separate phases: an a first *Pavlovian training phase*, participants learn predictive relations between stimuli and differential outcomes (e.g., S1-O1, S2-O2). In a subsequent *instrumental training phase*, they learn to produce these outcomes with particular actions (e.g., R1-O1, R2-O2). In a *transfer test*, both actions are then made available in extinction, and the preference for a specific action is measured in the presence of each conditioned stimulus (i.e., S1: R1 or R2?; S2: R1 or R2?). The typical result is a preference for the action whose outcome was signaled by the Pavlovian cue (i.e., S1: R1 > R2; S2: R2 > R1), suggesting that this stimulus has gained control over responding (for a review and meta-analysis, see Holmes et al., 2010;



Cartoni et al., 2016). Note that this cue-instigated action tendency cannot be explained with rote S-R learning because the action was not paired with the Pavlovian cue before the transfer test. Instead, it has been suggested that the Pavlovian cue primes the action by activating the sensory representation of the associated outcome *via* an associative S:(R-O) or S:(O-R) chain (Trapold and Overmier, 1972; Asratyan, 1974; Balleine and Ostlund, 2007; de Wit and Dickinson, 2009). According to this account, the Pavlovian cue activates a cognitive representation of the identity of the outcome (whatever its value), and this activation excites the action that is associated with the same outcome. In line with an associative S-O-R mechanism, research on “ideomotor effects” showed that presentations of action effect-related stimuli prime actions producing these effects (for reviews, see Shin et al., 2010; Hommel, 2013). An alternative account proposed that the Pavlovian cues act like discriminative stimuli in a hierarchical network that signal when a specific R-O relationship is in effect (Cartoni et al., 2013; Hogarth et al., 2014). According to this account, action choice in PIT tasks is driven by participants’ explicit beliefs about which action is more likely reinforced in the presence of a specific cue. For instance, participants in one experiment were told that the cues presented during a PIT test would indicate which action would *not* be rewarded. This instruction reversed the cue-instigated action tendency (Seabrooke et al., 2016). A follow-up study found this reversed PIT effect abolished by a cognitive load manipulation, while the standard PIT effect was spared (Seabrooke et al., 2019b). This research suggests that several processes could contribute to outcome-selective PIT effects: a resource-dependent one that is highly amenable to instructions, and a relatively resource-independent one that could be an association-based mechanism or a very simple behavioral rule. It should be noted that outcome-selective PIT effects were also observed in rodent studies, and it has been argued that the underlying mechanisms are causally involved in a broad range of “habitual” behaviors (Everitt and Robbins, 2005; Watson et al., 2012; Hogarth et al., 2013; Colagiuri and Lovibond, 2015).

Importantly, the outcome-selective PIT task can be combined with outcome devaluation treatments to examine a goal-independency of cue-instigated action tendencies. Using this research approach, animal studies found that rodents still work harder for a devalued food in the presence of a Pavlovian or discriminative cue associated with that food (Rescorla, 1994; Corbit and Balleine, 2005; Corbit et al., 2007). For example, in one study (Holland, 2004), hungry rats learned relations between stimuli and two unique food rewards (sucrose and food pellets). These food rewards were then used to reinforce two distinct actions (chain pulling and lever presses). In a subsequent extinction test, the rats had access to these responses during presentations of the Pavlovian cues. Performance in this first transfer test showed a standard outcome-selective PIT effect. After this test, one of the two food rewards was devalued by pairing it with a toxin. Then, the rats worked on a second transfer test in extinction. Although the conditioned food aversion clearly decreased working for that food at baseline,

the cue-instigated action tendency augmenting the devalued response was spared.

Results of outcome-selective PIT studies with human adults were however more mixed. While some studies confirmed the finding of animal studies that reinforcer-selective PIT does not change when the outcome is no longer desirable (Hogarth and Chase, 2011; Hogarth, 2012; Watson et al., 2014; van Steenbergen et al., 2017; De Tommaso et al., 2018), a few studies observed a change. One of these studies used a stock-market paradigm for a postlearning devaluation of outcomes (Allman et al., 2010). Human adults first learned to associate specific symbols and instrumental actions with two (fictitious) money currencies. In this phase of the experiment, both currencies had the same value, and participants knew that they can swap the earnings into real money after the study. In a first extinction test, a clear PIT effect was observed. After retraining, and immediately before a second transfer test, one of the two currencies was devalued by making the currency worthless. In the subsequent extinction test, responding for the intact currency was still elevated by matching cues; in contrast, working for the devalued money was generally disrupted and not affected by presentations of a matching cue. In short, the Pavlovian cue had lost its capacity to excite the devalued action.

Follow-up research showed that the cue-instigated action tendency is affected by a postlearning value decrease, but not by an equidistant value increase (Eder and Dignath, 2016a). The study used a stock-market paradigm similar to Allman et al. (2010). This time, however, the revaluation treatment involved three monetary outcomes: one currency was made worthless as in Allman et al. by decreasing its value by one unit ($1 \rightarrow 0$); the value of another currency was doubled ($1 \rightarrow 2$); the third currency maintained its value ($1 \rightarrow 1$) for baseline comparisons. If the cue-instigated action tendency is truly sensitive to the current value of outcomes, then it should decrease following the devaluation but increase following the upvaluation of the outcome. Results however showed that only the devaluation treatment had an effect: Outcome-selective PIT was significantly reduced after devaluation, reproducing the result of Allman et al. (2010). In contrast, PIT effects were not different from the baseline condition after the upvaluation. In short, only a decrease in the outcome value affected cue-instigated action tendencies, while an equidistant value increase had no effect.

The PIT studies reviewed above are puzzling and at odds with a large number of studies that reported no effect of postlearning changes in outcome values. In the search for an explanation, Watson and colleagues proposed that the stock-market paradigm involved highly abstract representations of values that were presumably more accessible to explicit choice strategies (Watson et al., 2018). While it is unclear why those explicit decision rules should not take a value increment into account (see Eder and Dignath, 2016a), recent studies confirmed that explicit beliefs can have a profound impact on outcome-selective PIT effects (see e.g., Seabrooke et al., 2016). In addition, the theoretical argument was made that Pavlovian cues can only activate the sensory identity of

action outcomes in PIT tasks and not their value (Balleine and Ostlund, 2007; de Wit and Dickinson, 2009). If money outcomes in the stock-market studies were represented predominantly in terms of their value, this could have made a critical difference to (animal) studies that used primary reinforcers with a more detailed sensory representation. Accordingly, it could be hypothesized that a standard PIT task with food outcomes should be not sensitive to postlearning changes in the values of outcomes.

Eder and Dignath (2016b) tested this hypothesis with liquid reinforcers. Participants were trained in separate sessions to associate specific symbols and keypresses, respectively, with red and yellow lemonades. Importantly, participants in this study were asked to consume the lemonades earned during a transfer test¹. After having worked on a first transfer test, one of the lemonades was devalued with bad-tasting Tween20. Then, a second transfer test was performed. Each transfer test was further subdivided into two test blocks. In the first experiment, participants consumed the earned lemonades immediately after each test block. In the second experiment, consumption was not immediate, and participants could take the earned lemonades with them in bottles. **Figure 2** shows the response rates in both experiments as a function of the Pavlovian cue in each test block. As can be seen, a strong and robust PIT effect was observed in both experiments before devaluation: working for a specific lemonade was elevated by presentations of cues associated with that lemonade (relative to a baseline condition with a neutral cue associated with no lemonade). However, response rates changed dramatically following the devaluation. Participants now preferred the action that produced the intact lemonade. Responding for this lemonade was still augmented by a matching cue relative to baseline. In contrast, the cue-instigated action tendency was abolished for the devalued response in Experiment 1 in which the liquids earned in a test block were consumed immediately. Interestingly, in Experiment 2 (without immediate consumption of the liquids), the cue-instigated tendency for the devalued response was abolished in the first test block only and restored in the second test block². It is plausible that the immediate consumption of the drinks increased the motivational relevance of the devalued drink. These results hence show that a strong devaluation treatment of food outcomes can also reduce cue-instigated action tendencies operating on a primary reinforcer.

¹The transfer test was carried out in nominal extinction (i.e., without feedback whether or which lemonade had been earned). This was done to prevent the feedback from influencing the response choice. Instruction explicitly stated that the actions during the transfer test procure lemonades (2.5 ml according to the fixed-ratio 9 schedule) and that the probability of a reward is not influenced by the pictures presented during this phase. Note that a reward expectancy during the extinction test is common in PIT studies (see e.g., Hogarth and Chase, 2011; Colagiuri and Lovibond, 2015). Furthermore, it increases the ecological validity of the PIT task to behavior outside of the laboratory (for a discussion of this point see Lovibond and Colagiuri, 2013).

²Collapsed across both test blocks, however, there was small PIT effect for the devalued response. Furthermore, the magnitudes of the PIT effects for the devalued response in both test blocks were not significantly different.

For an explanation, Eder and Dignath (2016b) suggested that only strong devaluation treatments suppress cue-instigated instigated actions. In fact, most studies that found no effect of the devaluation treatment used rather weak and/or incomplete devaluation treatments, such as ad libitum feeding, conditioning of a taste aversion, or health warnings (for a similar argumentation, see De Houwer et al., 2018)³. Hogarth and Chase (2011), for instance, used a specific satiety procedure to devalue a tobacco outcome. Although smoking a cigarette before a transfer test reduced participants' craving and working for cigarettes during the PIT test, cue-instigated action tendencies for that reward were not affected. Critically, working for the devalued tobacco outcome (irrespective of the cue) was still on a high level (>40%), suggesting that the devaluation was not very strong. In addition, regular smokers typically know that the state of satiety is only temporary. Therefore, it could be argued that working for cigarettes was still attractive for them during the transfer test. The devaluation treatment that is most comparable to the one used by Eder and Dignath (2016b) is conditioning of a taste aversion. Rodent studies often devalued a food reinforcer by pairing it with lithium chloride (LiCl) inducing sickness (e.g., Rescorla, 1994; Holland, 2004). Although LiCl-conditioning has a strong and lasting effect on the consumption of that food, the devaluation is often incomplete, because the animal must approach a magazine to consume the poisoned food and could reject consumption before the devaluation was complete. In fact, when Colwill and Rescorla (1990) used a standard procedure to devalue a sucrose solution with LiCl-injections before a transfer test, the devaluation treatment did not eliminate the cue-instigated action tendency. However, when the poisoned sucrose solution was injected directly into the mouth of the rat during conditioning, the stimulus lost its capacity to elevate the devalued response. Thus, animal research also found cue-instigated action tendencies abolished after a strong and immediate devaluation treatment, in line with the results of human studies reviewed above.

Our main conclusion from this short review is that the cue-instigated action tendency was suppressed when the devaluation of the associated action outcome was strong and complete. This does not mean that the action tendency scales directly with the current value of the associate outcome, as proposed for a goal-directed process. In this case, studies with a weak (but still effective) devaluation of the outcome should also have observed a reduction in cue-instigated tendencies, which was not the case (e.g., Hogarth and Chase, 2011; Watson et al., 2014; De Tommaso et al., 2018). In addition, an

³A notable exception is Experiment 1 in Seabrooke et al. (2017) that showed a PIT effect despite the use of a fairly strong devaluation treatment (coating of snacks with a distasteful paste). It should be noted, however, that (1) this study presented pictures of the food outcomes (and not Pavlovian cues) during the transfer test; (2) despite a clear reduction in subjective liking ratings, working for the devalued food (in the baseline condition) was still on a sizeable level (~25%); (3) the devalued food earned during the test was not immediately consumed (see Eder and Dignath, 2016b); (4) the same devaluation treatment affected PIT tendencies in subsequent experiments after modification of the task procedure (Seabrooke et al., 2017, 2019a).

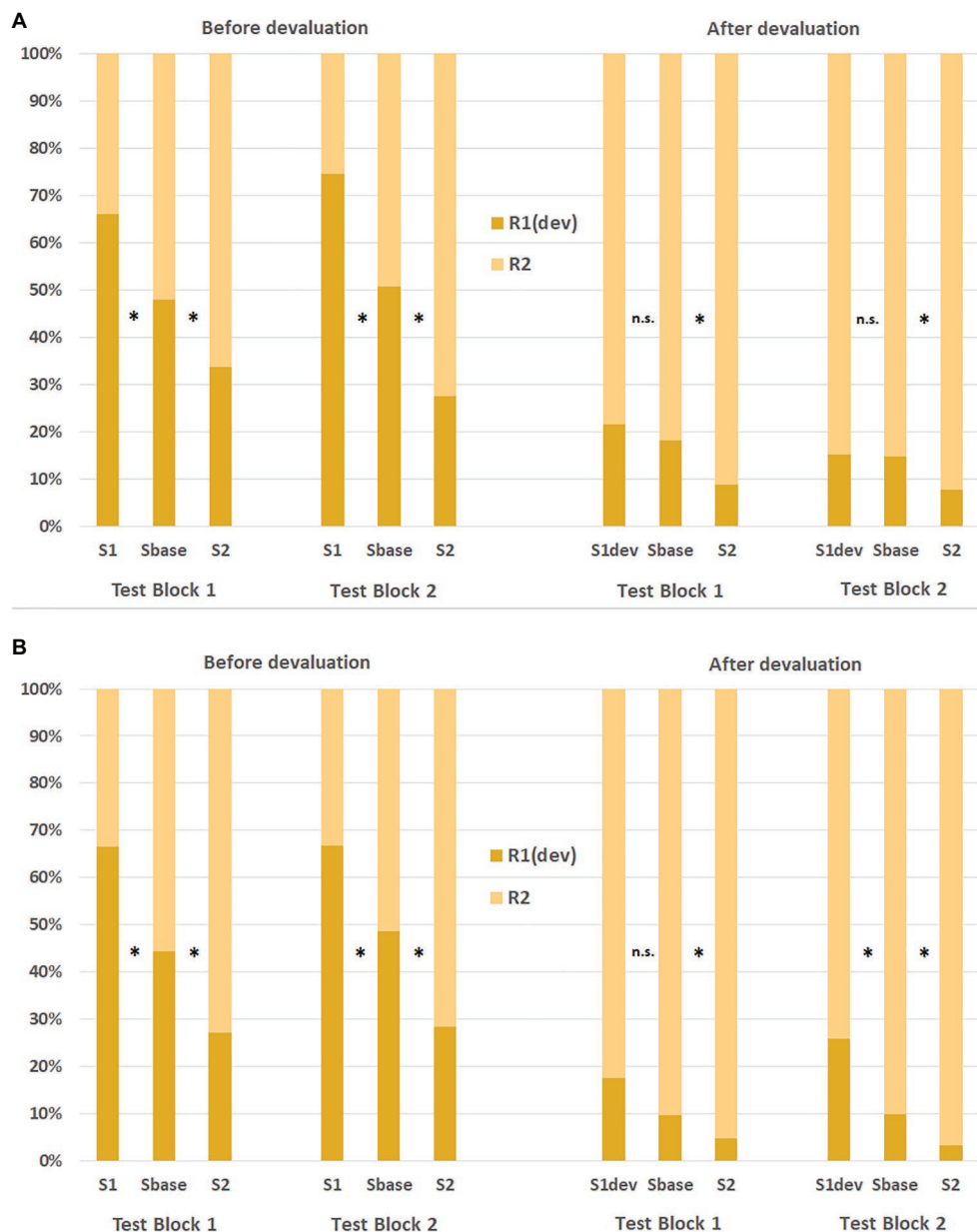


FIGURE 2 | Response ratios in Eder and Dignath (2016b) before and after the devaluation of a liquid reinforcer as a function of stimulus cue, action, and test block in Experiment 1 (upper panel A) and Experiment 2 (lower panel B). S, stimulus cue; R, response; dev, devalued outcome. *significant difference to the baseline condition at $p < 0.05$.

upvaluation of the associated outcome should have enhanced the cue-instigated action tendency, which was not observed (Eder and Dignath, 2016a). In short, the studies reviewed above do not question that the cue-instigated action tendency was “habitual” in the sense that the behavior was insensitive to the current value of the outcome; rather, they suggest that the habitual action tendency was cognitively suppressed because the devalued outcome was in conflict with other goals or intentions. According to this interpretation, an internal conflict signal is created after registration that the present state will deteriorate markedly with continued performance of the habitual action.

Detection of this conflict signal then triggers behavioral adaptations that aim to correct for the maladaptive habitual response. In the next section, we will describe two frameworks of how such a control system could be implemented on the cognitive level: a default-interventionist framework and EVC theory.

PART II

In this part, we will discuss two alternative frameworks of cognitive control: (1) a *default-interventionist framework* that

proposes a higher order cognitive control system that intervenes when the habitual action goes faulty. (2) *EVC theory* that explains the allocation of control with neural computations of the expected payoffs from engaging in cognitive control.

Default-Interventionist Framework

The default-interventionist framework postulates a cognitive control system that can intervene when the habitual “default” response becomes inappropriate, cumbersome, or defective. In its most basic form, the framework assumes two systems or control units of actions: a habitual controller and a goal-directed controller. Only the goal-directed controller is sensitive to changes in outcomes, while the habitual controller implements a stimulus-driven behavior without detailed representation of its consequences. This distinction is supported by neurophysiological research that studied dissociations in the control of voluntary and habitual actions on a neural systems level. More specifically, habitual and goal-directed controllers have been linked to two distinct (but interacting) cortico-basal ganglia networks in the brain: The associative cortico-basal ganglia loop controls goal-directed actions *via* projections from the prefrontal cortex (PFC) to the caudate nucleus and the anterior putamen. The sensorimotor loop controls habitual actions and connects the somatosensory and motor cortex with the medial and posterior putamen (for reviews, see Yin and Knowlton, 2006; Balleine et al., 2007; Graybiel and Grafton, 2015). Research found that after overtraining of a response (i.e., habit formation), neural activation is shifted from the associative loop to the sensorimotor loop (Ashby et al., 2010). Interestingly, goal-oriented behavior can be reinstated after inactivation of the infralimbic prefrontal cortex in the rodent brain (Coutureau and Killcross, 2003). This finding suggests that the circuits controlling goal-directed behavior are actively suppressed after habit formation.

The default-interventionist framework rests on the idea that there is a dynamic balance between action control systems, and that control could be shifted back from the habitual to the goal-directed control system if needed. This idea also fits with the long-standing view that prefrontal cortical areas have the capacity to override unwanted lower-order action tendencies (Koechlin et al., 2003). However, it has been argued that regaining control over habitual action tendencies is effortful and requires cognitive resources (Baddeley, 1996; Muraven and Baumeister, 2000). Furthermore, the person must be sufficiently motivated to invest resources in the executive control of the habitual action (Inzlicht and Schmeichel, 2012). Hence, a number of requirements must be met for the default-interventionist framework (for a defense and criticisms of this view, see Evans and Stanovich, 2013; Kruglanski, 2013; Hommel and Wiers, 2017; Melnikoff and Bargh, 2018).

It is likely that these conditions were met in the posttraining devaluation studies reviewed above. With a strong and complete devaluation of the outcome, participants were arguably motivated to avoid that outcome. In addition, performing the free-operant transfer task was very easy and without time pressure. However,

the explanatory problems with the default-interventionist framework are much more fundamental and concern the very architecture of this account. Specifically, it is not specified what controls the controller, leading to an infinite logical regress. This problem became apparent in early accounts that conceptualized the interventionist as a unitary system (supervisory attentional system, working memory system, goal-directed action controller, etc.). This approach was heavily criticized of introducing a “homunculus” (the executive controller) that pulls the levers to regulate lower levels if needed (Monsell and Driver, 2000). As a reaction to this criticism, the unitary control system view was replaced by more complex models that decomposed the “executive” in more specific control functions (e.g., mental set shifting, memory updating, response suppression; Miyake et al., 2000). However, as Verbruggen et al. (2014) unerringly pointed out, this approach only resulted in a multiplication of control homunculi and not in an explanation of how control is exercised. Thus, a fundamentally different approach is needed that explains cognitive control functions as an emergent phenomenon of the cognitive system.

Expected Value of Control

A model that has the potential to explain habit control in the PIT paradigm without recourse to control homunculi is found in EVC theory (Shenhav et al., 2013, 2016). This model analyzes cognitive control as a domain of reward-based decision making; that means, it is assumed that cognitive control functions serve to maximize desired outcomes through “controlled” processes when those outcomes could not otherwise be achieved by (habitual) “default” processes (Botvinick and Braver, 2015). The model aims to explain whether, where, and how much cognitive control is allocated to ongoing or planned activities. At the neural level, it is assumed that a central hub in this decision making process is the dorsal anterior cingulate cortex (dACC) that lies on the medial surfaces of the brain’s frontal lobes (see the central panel in **Figure 3**). Many studies showed that the dACC becomes active in control-demanding situations in which automatic action tendencies, such as habits, are in conflict with task-defined responses (see e.g., Procyk et al., 2000; for meta-analyses see Ridderinkhof et al., 2004; Nee et al., 2007). As a key hub in a wide network of distributed brain regions, it receives inputs from brain areas responsible for the valuation of incoming stimuli or action outcomes and sends output signals to areas responsible for the implementation of control (see **Figure 3**). In this network, it is assumed that dACC serves several functions: (1) it monitors ongoing processing to signal the need for control; (2) it evaluates the demands for control; (3) and it allocates control to downstream regions (Botvinick, 2007; Shenhav et al., 2016); for a different account of dACC functions, see Kolling et al., 2016).

According to EVC theory, two sources of value-related information are integrated in the dACC: (1) what control signal should be selected (i.e., its identity) and (2) how vigorously this control signal should be engaged (i.e., its intensity). The integration process considers the overall payoff

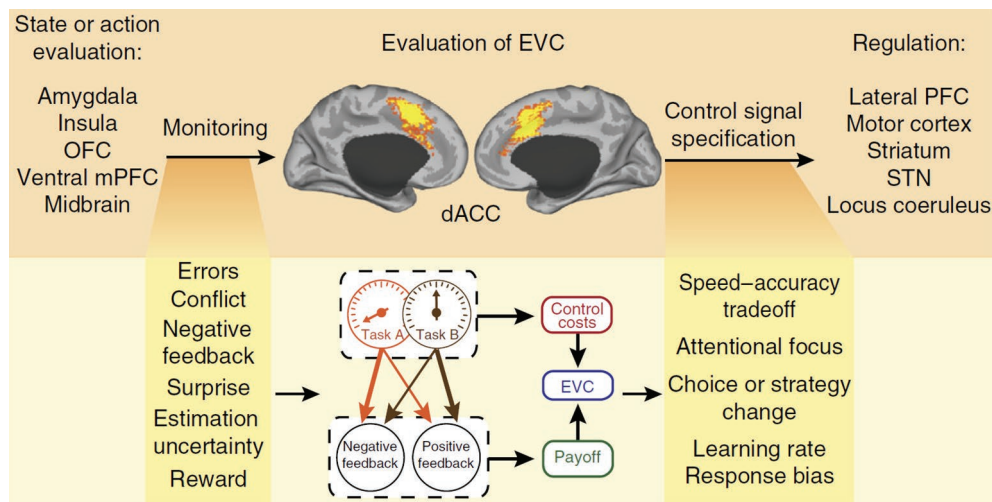


FIGURE 3 | Control allocation according to EVC theory. The dACC monitors ongoing processes for signals relevant to evaluating EVC and specifies the optimal control allocation to downstream regions for overriding a default behavior. OFC, orbitofrontal cortex; STN, subthalamic nucleus; mPFC, medial prefrontal cortex; PFC, prefrontal cortex. Figure reprinted by permission from Springer Nature: *Nature Neuroscience*, “Dorsal anterior cingulate cortex and the value of control”, © Shenhav et al. (2016).

that can be expected from engaging in a given control signal, taking into account the probabilities of positive and negative consequences that could result from performing a task. In addition, it takes into account that there is an intrinsic cost to engaging in control itself, which is a monotonic function of the intensity of the control signal (Shenhav et al., 2017). The expected value of a candidate control signal is the sum of its anticipated payoffs (weighted by their respective probabilities) minus the inherent cost of the signal (a function of its intensity). By relative comparisons, the candidate control signal with the maximum expected value is selected for a down-stream regulation of more basic processes. This selection process has been simulated as a stochastic evidence accumulation process using the drift diffusion model that avoids any recourse to a homunculus (Musslick et al., 2015). In contrast to the default-interventionist framework, EVC theory does not assume a hierarchy of action control systems but, rather, views the control of habitual actions as an emergent phenomenon of a unitary cognitive system. In addition, neural computations of the expected payoffs are continuously performed during task engagement, and control (e.g., attention) can be applied in varying degrees to the task at hand. It should be noted that the hypothesis of a neural implementation in the dACC is in principle independent of the computations proposed by the theory on the algorithmic level (Marr, 1982). In other words, it is possible that future neuroscientific research will identify other neural structures that calculate expected payoffs of engaging in control. By providing a computationally coherent and mechanistically explicit account of cognitive control functions on the algorithmic and implementational levels, EVC theory avoids the pitfall of introducing a new homunculus-like entity that magically guides cognition and behavior.

EVC theory can account for cognitive control functions and subsequent control adaptations in classic response conflict tasks (Ridderinkhof et al., 2004; Carter and van Veen, 2007; Nee et al., 2007), and the model was also used to explain behavioral flexibility that is characteristic of exploration and foraging (Shenhav et al., 2016). Most important for the present discussion, EVC theory can help to understand habit control in PIT tasks. In the remainder of this article, we provide a preliminary account of control functions in outcome-selective PIT.

In PIT tasks, the default response that must be potentially overcome is the cue-instigated action tendency that primes actions associated with shared outcomes. Before the revaluation treatment, however, there exists no motivation to override this default tendency. There is no action that would be more “correct” or valuable and that could be increased for a better payoff. To the contrary, overcoming the PIT tendency would be effortful (for indirect evidence on this assumption, see Cavanagh et al., 2013; Freeman et al., 2014; see also Yee and Braver, 2018). Therefore, the expected payoff does not justify the intrinsic cost of control. As a result, the cue-instigated action tendency is not or only minimally controlled in this phase, resulting in a PIT effect.

Expected payoffs however change dramatically after a strong revaluation of the outcome. Now, there exists a clear difference in the value of action outcomes, and response rates are adjusted to maximize the reward. At the computational level, this behavioral adjustment is implemented by prioritizing control signals that maximize the value of outcomes. As a consequence, control of action tendencies that would produce devalued outcomes is now justified, because the anticipated outcome of the intact response outweighs the effort that is necessary to override the devaluated response. Control is however not

intensified following the registration of an action tendency that would result in high-value outcomes. As a consequence, the cue-instigated action tendency is only controlled (i.e., suppressed) if it results in a devalued outcome, whereas actions resulting in desirable outcomes do not (or to a much smaller degree) demand control.

EVC theory can hence explain why studies found reduced PIT tendencies only with very strong and/or complete devaluation treatments. The outcome value arguably shrank less by a weak relative to a strong devaluation treatment. The small decrement in the expected payoff does not justify the intrinsic costs of engaging in control. Furthermore, a EVC account of the PIT task can also explain observed effects that the default interventionistic account cannot explain. For instance, computations of expected payoffs take into account a temporal discounting of future and/or past outcomes (Yi et al., 2009). Immediate outcomes are typically weighted more than temporally distant outcomes. This immediacy bias can explain why immediate (relative to delayed) consumptions had a stronger effect on cue-instigated action tendencies in the study of Eder and Dignath (2016b). Furthermore, if the negative value of the devalued drink was discounted with the time that elapsed or will elapse since the consumption of that drink (Yi et al., 2009), the expected value of engaging in control is the largest immediately after consumption of the drink. Temporal discounting of the negative outcome value can hence explain why PIT tendencies were abolished in the first test block and restored in the second test block of Eder and Dignath's experiment.

EVC theory also provides an explanation why the postlearning devaluation of the outcome had a stronger effect on the control of PIT tendencies compared to the upvaluation (Eder and Dignath, 2016a). Research on cognitive control showed that negative outcomes elicit a stronger control signal (Hajcak et al., 2005) and that conflict is aversive (Botvinick, 2007; Inzlicht et al., 2015). In line with this suggestion, studies found that conflict elicits a negative affective response (Dreisbach and Fischer, 2012) that triggers avoidance (Dignath et al., 2015; Dignath and Eder, 2015). In addition, (unexpected) positive events reduce conflict-driven behavioral adaptations, presumably because they weaken the negative conflict signal that signals need for control (e.g., van Steenbergen et al., 2009; but see also Dignath et al., 2017). It is hence plausible that a positive affective response to the (unexpected) upvaluation of a currency in the study of Eder and Dignath (2016a) has analogously decreased the intensity of the control signal that signaled need for control of the cue-instigated action tendency.

In summary, EVC theory can explain most findings of the PIT studies reviewed above. While this account is *ex post facto*, it has the benefit of providing a formal and mechanistic account of the effect of posttraining revaluation treatments on PIT tendencies. In addition, the account allows for new predictions. According to EVC theory, cognitive control of cue-instigated action tendencies should be inversely related to the intrinsic cost of control effort. Therefore, one

would expect that PIT tendencies should recover in demanding transfer tasks with high intrinsic costs of control, even when the devaluation of the associated outcome was very strong. For instance, costs of engaging in control could be manipulated by increasing the investment of resources that are necessary to reach a decision and/or to implement the action (Boureau et al., 2015). These costs could be cognitive (e.g., evaluation times), physical (e.g., energy expenditure), and/or emotional (e.g., negative affective experiences). When intrinsic costs outweigh the cost of producing a devalued outcome in a PIT task, the prediction would be that control of cue-instigated action tendencies becomes relaxed, resulting in larger outcome-selective PIT effects. Having a strong foundation in neuroscientific research, the account also makes new predictions at the neural level. Specifically, activity of dACC should increase following the strong devaluation of an outcome, indexing the monitoring and implementation of a control setting. In addition, dACC should be most active during presentations of Pavlovian cues predictive of the devalued outcome. Hence, several hypotheses can be deduced from EVC theory that could be examined in future research.

CONCLUSION

Habits have a great influence on our behavior. Some habits we strive for, and work hard to make them part of our behavioral repertoire. Other habits we want to abolish because they are problematic. Habits are consequently closely linked to cognitive control functions that regulate habitual action tendencies for the pursuit of higher-order goals. In this article, we argued on the basis of EVC theory that the allocation of control to habitual action tendencies is based on evaluations that compute the expected value of control by taking intrinsic costs of effortful control into account. Habits hence may be insensitive to changes in outcomes values because the expected benefits that follow from habit control do not justify the costs of control. The often cited insensitivity to changes in action outcomes is consequently not an intrinsic design feature of habits but, rather, a function of the cognitive system that controls habitual action tendencies.

AUTHOR CONTRIBUTIONS

AE drafted the manuscript. DD provided critical revisions. All authors approved the final version of the manuscript for submission.

FUNDING

The work described in this article was supported by grants Ed201/2-1 and Ed201/2-2 of the German Research Foundation (DFG) to AE. The funding agency had no role in writing the manuscript or the decision to submit the paper for publication.

REFERENCES

- Adams, C. D., and Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol.* 33, 109–121.
- Allman, M. J., DeLeon, I. G., Cataldo, M. F., Holland, P. C., and Johnson, A. W. (2010). Learning processes affecting human decision making: an assessment of reinforcer-selective Pavlovian-to-instrumental transfer following reinforcer devaluation. *J. Exp. Psychol. Anim. Behav. Process.* 36, 402–408. doi: 10.1037/a0017876
- Ashby, F. G., Turner, B. O., and Horvitz, J. C. (2010). Cortical and basal ganglia contributions to habit learning and automaticity. *Trends Cogn. Sci.* 14, 208–215. doi: 10.1016/j.tics.2010.02.001
- Asratyan, E. A. (1974). Conditional reflex theory and motivational behavior. *Acta Neurobiol. Exp.* 34, 15–31.
- Baddeley, A. (1996). Exploring the central executive. *Q. J. Exp. Psychol.* 49, 5–28. doi: 10.1080/713755608
- Balleine, B. W., Delgado, M. R., and Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* 27, 8161–8165. doi: 10.1523/JNEUROSCI.1554-07.2007
- Balleine, B. W., and Ostlund, S. B. (2007). “Still at the choice-point: action selection and initiation in instrumental conditioning” in *Reward and decision making in corticobasal ganglia networks*. eds. B. W. Balleine, K. Doya, J. O’Doherty, and M. Sakagami (Malden: Blackwell Publishing), 147–171.
- Bargh, J. A. (1994). “The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition” in *Handbook of social cognition*, Vol. 1: Basic processes; Vol. 2: Applications. 2nd Edn. eds. R. S. Wyer Jr., and T. K. Srull (Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.), 1–40.
- Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cogn. Affect. Behav. Neurosci.* 7, 356–366. doi: 10.3758/CABN.7.4.356
- Botvinick, M., and Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Annu. Rev. Psychol.* 66, 83–113. doi: 10.1146/annurev-psych-010814-015044
- Boureau, Y.-L., Sokol-Hessner, P., and Daw, N. D. (2015). Deciding how to decide: self-control and meta-decision making. *Trends Cogn. Sci.* 19, 700–710. doi: 10.1016/j.tics.2015.08.013
- Carter, C. S., and van Veen, V. (2007). Anterior cingulate cortex and conflict detection: an update of theory and data. *Cogn. Affect. Behav. Neurosci.* 7, 367–379. doi: 10.3758/CABN.7.4.367
- Cartoni, E., Balleine, B., and Baldassarre, G. (2016). Appetitive Pavlovian-instrumental transfer: a review. *Neurosci. Biobehav. Rev.* 71, 829–848. doi: 10.1016/j.neubiorev.2016.09.020
- Cartoni, E., Puglisi-Allegra, S., and Baldassarre, G. (2013). The three principles of action: a Pavlovian-instrumental transfer hypothesis. *Front. Behav. Neurosci.* 7:153. doi: 10.3389/fnbeh.2013.00153
- Cavanagh, J. F., Eisenberg, I., Guitart-Masip, M., Huys, Q., and Frank, M. J. (2013). Frontal theta overrides Pavlovian learning biases. *J. Neurosci.* 33, 8541–8548. doi: 10.1523/JNEUROSCI.5754-12.2013
- Colagiuri, B., and Lovibond, P. F. (2015). How food cues can enhance and inhibit motivation to obtain and consume food. *Appetite* 84, 79–87. doi: 10.1016/j.appet.2014.09.023
- Colwill, R. M., and Rescorla, R. A. (1985). Instrumental responding remains sensitive to reinforcer devaluation after extensive training. *J. Exp. Psychol. Anim. Behav. Process.* 11, 520–536.
- Colwill, R. M., and Rescorla, R. A. (1990). Effect of reinforcer devaluation on discriminative control of instrumental behavior. *J. Exp. Psychol. Anim. Behav. Process.* 16, 40–47.
- Corbit, L. H., and Balleine, B. W. (2005). Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of Pavlovian-instrumental transfer. *J. Neurosci.* 25, 962–970. doi: 10.1523/JNEUROSCI.4507-04.2005
- Corbit, L. H., Janak, P. H., and Balleine, B. W. (2007). General and outcome-specific forms of Pavlovian-instrumental transfer: the effect of shifts in motivational state and inactivation of the ventral tegmental area. *Eur. J. Neurosci.* 26, 3141–3149. doi: 10.1111/j.1460-9568.2007.05934.x
- Coutureau, E., and Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behav. Brain Res.* 146, 167–174. doi: 10.1016/j.bbr.2003.09.025
- De Houwer, J., Tanaka, A., Moors, A., and Tibboel, H. (2018). Kicking the habit: why evidence for habits in humans might be overestimated. *Motiv. Sci.* 4, 50–59. doi: 10.1037/mot0000065
- De Tommaso, M., Mastropasqua, T., and Turatto, M. (2018). Working for beverages without being thirsty: human Pavlovian-instrumental transfer despite outcome devaluation. *Learn. Motiv.* 63, 37–48. doi: 10.1016/j.lmot.2018.01.001
- de Wit, S., and Dickinson, A. (2009). Associative theories of goal-directed behaviour: a case for animal–human translational models. *Psychol. Res.* 73, 463–476. doi: 10.1007/s00426-009-0230-6
- de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., et al. (2018). Shifting the balance between goals and habits: five failures in experimental habit induction. *J. Exp. Psychol. Gen.* 147, 1043–1065. doi: 10.1037/xge0000402
- Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 308, 67–78.
- Dickinson, A., and Balleine, B. (1993). “Actions and responses: the dual psychology of behaviour” in *Spatial representation: Problems in philosophy and psychology*. eds. N. Eilan, R. A. McCarthy, and B. Brewer (Malden: Blackwell Publishing), 277–293.
- Dickinson, A., Nicholas, D. J., and Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Q. J. Exp. Psychol.* 35, 35–51. doi: 10.1080/14640748308400912
- Dignath, D., and Eder, A. B. (2015). Stimulus conflict triggers behavioral avoidance. *Cogn. Affect. Behav. Neurosci.* 15, 822–836. doi: 10.3758/s13415-015-0355-6
- Dignath, D., Janczyk, M., and Eder, A. B. (2017). Phasic valence and arousal do not influence post-conflict adjustments in the Simon task. *Acta Psychol.* 174, 31–39. doi: 10.1016/j.actpsy.2017.01.004
- Dignath, D., Kiesel, A., and Eder, A. B. (2015). Flexible conflict management: conflict avoidance and conflict adjustment in reactive cognitive control. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 975–988. doi: 10.1037/xlm0000089
- Dreisbach, G., and Fischer, R. (2012). Conflicts as aversive signals. *Brain Cogn.* 78, 94–98. doi: 10.1016/j.bandc.2011.12.003
- Eder, A. B., and Dignath, D. (2016a). Asymmetrical effects of posttraining outcome reevaluation on outcome-selective Pavlovian-to-instrumental transfer of control in human adults. *Learn. Motiv.* 54, 12–21. doi: 10.1016/j.lmot.2016.05.002
- Eder, A. B., and Dignath, D. (2016b). Cue-elicited food seeking is eliminated with aversive outcomes following outcome devaluation. *Q. J. Exp. Psychol.* 69, 574–588. doi: 10.1080/17470218.2015.1062527
- Evans, J. S. B. T., and Stanovich, K. E. (2013). Dual-process theories of higher cognition: advancing the debate. *Perspect. Psychol. Sci.* 8, 223–241. doi: 10.1177/1745691612460685
- Everitt, B. J., and Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* 8, 1481–1489. doi: 10.1038/nn1579
- Freeman, S. M., Razhas, I., and Aron, A. R. (2014). Top-down response suppression mitigates action tendencies triggered by a motivating stimulus. *Curr. Biol.* 24, 212–216. doi: 10.1016/j.cub.2013.12.019
- Graybiel, A. M., and Grafton, S. T. (2015). The striatum: where skills and habits meet. *Cold Spring Harb. Perspect. Biol.* 7:a021691. doi: 10.1101/cshperspect.a021691
- Hajcak, G., Holroyd, C. B., Moser, J. S., and Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology* 42, 161–170. doi: 10.1111/j.1469-8986.2005.00278.x
- Hogarth, L. (2012). Goal-directed and transfer-cue-elicited drug-seeking are dissociated by pharmacotherapy: evidence for independent additive controllers. *J. Exp. Psychol. Anim. Behav. Process.* 38, 266–278. doi: 10.1037/a0028914
- Hogarth, L. (2018). “A critical review of habit theory of drug dependence” in *The psychology of habit: Theory, mechanisms, change, and contexts*. ed. B. Verplanken (Cham, CH: Springer International Publishing), 325–341. doi: 10.1007/978-3-319-97529-0_18

- Hogarth, L., Balleine, B. W., Corbit, L. H., and Killcross, S. (2013). Associative learning mechanisms underpinning the transition from recreational drug use to addiction. *Ann. N. Y. Acad. Sci.* 1282, 12–24. doi: 10.1111/j.1749-6632.2012.06768.x
- Hogarth, L., and Chase, H. W. (2011). Parallel goal-directed and habitual control of human drug-seeking: implications for dependence vulnerability. *J. Exp. Psychol. Anim. Behav. Process.* 37, 261–276. doi: 10.1037/a0022913
- Hogarth, L., Retzler, C., Munafò, M. R., Tran, D. M. D., Troisi, J. R. II, Rose, A. K., et al. (2014). Extinction of cue-evoked drug-seeking relies on degrading hierarchical instrumental expectancies. *Behav. Res. Ther.* 59, 61–70. doi: 10.1016/j.brat.2014.06.001
- Holland, P. C. (2004). Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *J. Exp. Psychol. Anim. Behav. Process.* 30, 104–117. doi: 10.1037/0097-7403.30.2.104
- Holmes, N. M., Marchand, A. R., and Coutureau, E. (2010). Pavlovian to instrumental transfer: a neurobehavioural perspective. *Neurosci. Biobehav. Rev.* 34, 1277–1295. doi: 10.1016/j.neubiorev.2010.03.007
- Hommel, B. (2013). “Ideomotor action control: on the perceptual grounding of voluntary actions and agents” in *Action science: Foundations of an emerging discipline*. eds. W. Prinz, M. Beisert, and A. Herwig (Cambridge, MA US: MIT Press), 113–136.
- Hommel, B., and Wiers, R. W. (2017). Towards a unitary approach to human action control. *Trends Cogn. Sci.* 21, 940–949. doi: 10.1016/j.tics.2017.09.009
- Hull, C. L. (1931). Goal attraction and directing ideas conceived as habit phenomena. *Psychol. Rev.* 38, 487–506.
- Inzlicht, M., Bartholow, B. D., and Hirsh, J. B. (2015). Emotional foundations of cognitive control. *Trends Cogn. Sci.* 19, 126–132. doi: 10.1016/j.tics.2015.01.004
- Inzlicht, M., and Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspect. Psychol. Sci.* 7, 450–463. doi: 10.1177/1745691612454134
- Keren, G., and Schul, Y. (2009). Two is not always better than one: a critical evaluation of two-system theories. *Perspect. Psychol. Sci.* 4, 533–550. doi: 10.1111/j.1745-6924.2009.01164.x
- Koechlin, E., Ody, C., and Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science* 302, 1181–1185. doi: 10.1126/science.1088545
- Kolling, N., Wittmann, M. K., Behrens, T. E. J., Boorman, E. D., Mars, R. B., and Rushworth, M. F. S. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nat. Neurosci.* 19, 1280–1285. doi: 10.1038/nn.4382
- Kosaki, Y., and Dickinson, A. (2010). Choice and contingency in the development of behavioral autonomy during instrumental conditioning. *J. Exp. Psychol. Anim. Behav. Process.* 36, 334–342. doi: 10.1037/a0016887
- Kruglanski, A. W. (2013). Only one? The default interventionist perspective as a unimodel—commentary on Evans & Stanovich (2013). *Perspect. Psychol. Sci.* 8, 242–247. doi: 10.1177/1745691613483477
- Lerman, D. C., and Iwata, B. A. (1995). Prevalence of the extinction burst and its attenuation during treatment. *J. Appl. Behav. Anal.* 28, 93–94. doi: 10.1901/jaba.1995.28.93
- Lovibond, P. F., and Colagiuri, B. (2013). Facilitation of voluntary goal-directed action by reward cues. *Psychol. Sci.* 24, 2030–2037. doi: 10.1177/0956797613484043
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, US: W. H. Freeman and Company.
- Melnikoff, D. E., and Bargh, J. A. (2018). The mythical number two. *Trends Cogn. Sci.* 22, 280–293. doi: 10.1016/j.tics.2018.02.001
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734
- Monsell, S., and Driver, J. (2000). “Banishing the control homunculus” in *Control of cognitive processes: Attention and performance XVIII*. eds. S. Monsell, and J. Driver (Cambridge, MA US: MIT Press), 3–32.
- Moors, A., and De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychol. Bull.* 132, 297–326. doi: 10.1037/0033-2909.132.2.297
- Muraven, M., and Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: does self-control resemble a muscle? *Psychol. Bull.* 126, 247–259. doi: 10.1037/0033-2909.126.2.247
- Musslick, S., Shenhav, A., Botvinick, M. M., and Cohen, J. D. (2015). A computational model of control allocation based on the expected value of control. *The 2nd Multidisciplinary Conference on Reinforcement Learning and Decision Making*. Presented at the Edmonton, Canada. Available at: <https://tinyurl.com/y499pc7n> (Accessed July 29, 2019).
- Nee, D. E., Wager, T. D., and John, J. (2007). Interference resolution: insights from a meta-analysis of neuroimaging tasks. *Cogn. Affect. Behav. Neurosci.* 7, 1–17. doi: 10.3758/CABN.7.1.1
- Nisbett, R. E., and Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* 84, 231–259. doi: 10.1037/0033-295X.84.3.231
- Procyk, E., Tanaka, Y. L., and Joseph, J. P. (2000). Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. *Nat. Neurosci.* 3, 502–508. doi: 10.1038/74880
- Rescorla, R. A. (1994). Transfer of instrumental control mediated by a devalued outcome. *Anim. Learn. Behav.* 22, 27–33.
- Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science* 306, 443–447. doi: 10.1126/science.1100301
- Seabrooke, T., Hogarth, L., Edmunds, C. E. R., and Mitchell, C. J. (2019a). Goal-directed control in Pavlovian-instrumental transfer. *J. Exp. Psychol. Animal Learn. Cogn.* 45, 95–101. doi: 10.1037/xan0000191
- Seabrooke, T., Hogarth, L., and Mitchell, C. J. (2016). The propositional basis of cue-controlled reward seeking. *Q. J. Exp. Psychol.* 69, 2452–2470. doi: 10.1080/17470218.2015.1115885
- Seabrooke, T., Le Pelley, M. E., Hogarth, L., and Mitchell, C. J. (2017). Evidence of a goal-directed process in human Pavlovian-instrumental transfer. *J. Exp. Psychol. Animal Learn. Cogn.* 43, 377–387. doi: 10.1037/xan0000147
- Seabrooke, T., Wills, A. J., Hogarth, L., and Mitchell, C. J. (2019b). Automaticity and cognitive control: effects of cognitive load on cue-controlled reward choice. *Q. J. Exp. Psychol.* 72, 1507–1521. doi: 10.1177/1747021818797052
- Shenhav, A., Botvinick, M. M., and Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* 79, 217–240. doi: 10.1016/j.neuron.2013.07.007
- Shenhav, A., Cohen, J. D., and Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nat. Neurosci.* 19, 1286–1291. doi: 10.1038/nn.4384
- Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., et al. (2017). Toward a rational and mechanistic account of mental effort. *Annu. Rev. Neurosci.* 40, 99–124. doi: 10.1146/annurev-neuro-072116-031526
- Shin, Y. K., Proctor, R. W., and Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychol. Bull.* 136, 943–974. doi: 10.1037/a0020541
- Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. Lewiston, NY, US: Macmillan Press.
- Trapold, M. A., and Overmier, J. B. (1972). “The second learning process in instrumental learning” in *Classical conditioning II: Current research and theory*. eds. A. A. Black, and W. F. Prokasy (New York: Appleton-Century-Crofts), 427–452.
- Tricomi, E., Balleine, B. W., and O’Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29, 2225–2232. doi: 10.1111/j.1460-9568.2009.06796.x
- van Steenbergen, H., van Band, G. P. H., and Hommel, B. (2009). Reward counteracts conflict adaptation: evidence for a role of affect in executive control. *Psychol. Sci.* 20, 1473–1477. doi: 10.1111/j.1467-9280.2009.02470.x
- van Steenbergen, H., Watson, P., Wiers, R. W., Hommel, B., and de Wit, S. (2017). Dissociable corticostriatal circuits underlie goal-directed vs. cue-elicited habitual food seeking after satiation: evidence from a multimodal MRI study. *Eur. J. Neurosci.* 46, 1815–1827. doi: 10.1111/ejn.13586
- Verbruggen, F., McLaren, I. P. L., and Chambers, C. D. (2014). Banishing the control homunculi in studies of action control and behavior change. *Perspect. Psychol. Sci.* 9, 497–524. doi: 10.1177/1745691614526414
- Watson, P., de Wit, S., Hommel, B., and Wiers, R. W. (2012). Motivational mechanisms and outcome expectancies underlying the approach bias toward addictive substances. *Front. Psychol.* 3:440. doi: 10.3389/fpsyg.2012.00440

- Watson, P., Wiers, R. W., Hommel, B., and de Wit, S. (2014). Working for food you don't desire: cues interfere with goal-directed food-seeking. *Appetite* 79, 139–148. doi: 10.1016/j.appet.2014.04.005
- Watson, P., Wiers, R. W., Hommel, B., and de Wit, S. (2018). Motivational sensitivity of outcome-response priming: experimental research and theoretical models. *Psychon. Bull. Rev.* 25, 2069–2082. doi: 10.3758/s13423-018-1449-2
- Wood, W., and Rünger, D. (2016). Psychology of habit. *Annu. Rev. Psychol.* 67, 289–314. doi: 10.1146/annurev-psych-122414-033417
- Yee, D. M., and Braver, T. S. (2018). Interactions of motivation and cognitive control. *Curr. Opin. Behav. Sci.* 19, 83–90. doi: 10.1016/j.cobeha.2017.11.009
- Yi, R., Landes, R. D., and Bickel, W. K. (2009). Novel models of intertemporal valuation: past and future outcomes. *J. Neurosci. Psychol. Econ.* 2:102. doi: 10.1037/a0017571
- Yin, H. H., and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7, 464–476. doi: 10.1038/nrn1919
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Eder and Dignath. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Habit Expression and Disruption as a Function of Attention-Deficit/Hyperactivity Disorder Symptomology

Ahmet O. Ceceli^{1*}, Giavanna Esposito² and Elizabeth Tricomi¹

¹ Department of Psychology, Rutgers University-Newark, Newark, NJ, United States, ² New Jersey Institute of Technology, Newark, NJ, United States

OPEN ACCESS

Edited by:

John A. Bargh,
Yale University, United States

Reviewed by:

Laura Bradfield,
University of Technology Sydney,
Australia
Ofir Turel,
California State University, Fullerton,
United States

*Correspondence:

Ahmet O. Ceceli
ahmetc@rutgers.edu;
ahmetc36@gmail.com

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 14 March 2019

Accepted: 14 August 2019

Published: 03 September 2019

Citation:

Ceceli AO, Esposito G and
Tricomi E (2019) Habit Expression
and Disruption as a Function
of Attention-Deficit/Hyperactivity
Disorder Symptomology.
Front. Psychol. 10:1997.
doi: 10.3389/fpsyg.2019.01997

Attention-deficit/hyperactivity disorder (ADHD) is associated with neurobehavioral reward system dysfunctions that pose debilitating impairments in adaptive decision-making. A candidate mechanism for such anomalies in ADHD may be a compromise in the control of motivated behaviors. Thus, demonstrating and restoring potential motivational control irregularities may serve significant clinical benefit. The motivational control of action guides goal-directed behaviors that are driven by outcome value, and habits that are inflexibly cue-triggered. We examined whether ADHD symptomology within the general population is linked to habitual control, and whether a motivation-based manipulation can break well-learned habits. We obtained symptom severity scores from 106 participants and administered a Go/NoGo task that capitalizes on familiar, well-learned associations (green-Go and red-NoGo) to demonstrate outcome-insensitivity when compared to newly learned Go/NoGo associations. We tested for outcome-insensitive habits by changing the Go and NoGo contingencies, such that Go signals became NoGo signals and vice versa. We found that generally, participants responded less accurately when green and red stimuli were mapped to color-response contingencies that were incongruent with daily experiences, whereas novel Go/NoGo stimuli evoked similar accuracy regardless of color-response mappings. Thus, our Go/NoGo task successfully elicited outcome-insensitive habits (i.e., persistent responses to familiar stimuli without regard for consequences); however, this effect was independent of ADHD symptomology. Nevertheless, we found an association between hyperactivity and congruent Go response latency, suggesting heightened prepotency to perform habitual Go actions as hyperactivity increases. To examine habit disruption, participants returned to the lab and underwent the familiar version of the Go/NoGo task, but were given mid-experiment performance tracking information and a monetary incentive prior to contingency change. We found that this motivational boost via dual feedback prevented the incongruency-related accuracy impairment, effectively breaking the habit, albeit independent of ADHD symptomology. Our findings present only a modest link between ADHD symptomology and motivational control, which

may be due to compensatory mechanisms in ADHD driving goal-directed control, or our task's potential insensitivity to individual differences in ADHD symptomology. Further investigations may be crucial for determining whether ADHD is related to motivational impairments.

Keywords: ADHD, reward, habit, goal-directed, motivation, control

INTRODUCTION

Individuals with attention deficit-hyperactivity disorder (ADHD) are known to exhibit cognitive impairments that span domains of attention and impulsivity (American Psychiatric Association, 2013). These hallmark symptoms are often accompanied by executive control irregularities, such as diminished inhibitory control and excessive distractibility that interfere with daily functioning (Willcutt et al., 2005). Additionally, behavioral and neurobiological reports have highlighted reward-related abnormalities in ADHD, in that individuals with ADHD display impairments in learning from, interacting with, and processing rewards (Ceceli et al., 2019). Children and adults with ADHD present heightened delay aversion, such that they choose immediate, less valuable rewards over delayed yet larger rewards (Sonuga-Barke et al., 1992; Kessler et al., 2005b; Antrop et al., 2006; Marx et al., 2013). In addition to such examples of suboptimal decision-making, individuals with ADHD also exhibit abnormal reward-related neural processing in the brain's reward circuitry, such as decreased signaling in the ventral striatum during reward anticipation, and atypical orbitofrontal cortex (OFC) activity during reward delivery (Ströhle et al., 2008; Wilbertz et al., 2012; Furukawa et al., 2014; Plichta and Scheres, 2014; von Rhein et al., 2015). The affected regions of the brain that regulate reward anticipation and processing (i.e., the striatum and prefrontal cortex), are also known as integral areas for executing motivated behaviors (Balleine and O'Doherty, 2009; O'Doherty, 2016). These neurobehavioral dysfunctions in ADHD, when taken together with the cardinal presentations of inattention and impulsivity, suggest potential disparities in the control of motivated behaviors that have yet to be elucidated.

The motivational account of behavioral control posits that our actions can be either goal-directed, as in, performed deliberately in pursuit of a desirable outcome, or habitual, as in, triggered in response to a salient cue regardless of outcome value (Dickinson and Balleine, 1994). These components of motivational control have distinct neural signatures, such that the prefrontal cortex and caudate are known to be imperative for the execution of goal-directed behaviors, while cue-based habitual control is largely associated with the putamen and motor cortex (Haber, 2003; O'Doherty et al., 2004; Tricomi et al., 2009). Interestingly, a compelling body of work documents functional and structural abnormalities in ADHD when compared to neurotypicals (NTs) in these brain regions, suggesting a compromised corticostriatal system that could be indicative of motivational control deficits. For example, ADHD is associated with reduced gray matter volume in the caudate, expansion of the posterior putamen, and aberrant connectivity in the ventromedial prefrontal cortex

(vmPFC) and anterior cingulate cortex (ACC) (Qiu et al., 2009; Frodl and Skokauskas, 2012; Costa Dias et al., 2013; Norman et al., 2016; von Rhein et al., 2017; Rosch et al., 2018). Studies in rodents have suggested that a rat model of ADHD, the spontaneously hypertensive rat, exhibits a habit-dominated motivational control system, in that these rats that possess ADHD-like symptoms also display outcome-insensitive behavioral patterns (i.e., pressing a lever that predicts a food outcome to which the rat is sated) (Natsheh and Shiflett, 2015). Neural evidence suggests that this behavioral deficit is linked to imbalances in dopamine receptor activation, supporting the idea that abnormalities in the striatal systems may also manifest as an over-reliance on habitual control in ADHD (Natsheh and Shiflett, 2018).

If ADHD is indeed associated with enhanced habitual control that favors outcome-insensitive behaviors, the next logical and translationally valuable step would be to identify strategies that can overcome this behavioral deficit. For instance, performance-contingent feedback is a frequently employed tool that has been shown to improve behavioral output (Montague and Webber, 1965; Kluger and DeNisi, 1996). The positive effects of feedback in the form of performance-tracking information, as well as primary and secondary incentives, have been well-documented in the cognitive flexibility domain – namely using task-switching paradigms. Indeed, even the promise of a future performance-contingent reward has been shown to amplify task-switching performance (Yee et al., 2016). Importantly, performance-contingent monetary feedback is associated with the engagement of top-down control of task-switching processes (Umemoto and Holroyd, 2015). Taken together, we believe that the benefits of feedback on behavioral output and control over actions may carry over to the restoration of goal-directed behaviors in ADHD. Specifically, we reason that amplifying the salience of the outcomes of one's behaviors with feedback (e.g., tying task performance to monetary incentives and performance tracking) may reactivate goal-representations in otherwise stimulus-driven associations. In support of this hypothesis, we have previously demonstrated the beneficial effects of feedback on the motivational control of action (Ceceli et al., 2019).

Tackling the expression of habits and the restoration of goal-directed behaviors in potentially compromised populations may involve overcoming the methodological limitations of the traditional habit paradigm. A meaningful assessment of habit expression and disruption may require access to rigid habits with a strong association between the triggering stimulus and the behavioral response. Therefore, instead of relying on labile, newly learned habits that have been the subject of inquiry in most investigations of motivational control (Ceceli and Tricomi, 2018), it may be more effective to study habit expression and disruption

via well-learned, existing S–R associations that do not require extensive training in the laboratory (Ceceli et al., 2019).

To this end, we developed a Go/NoGo task that capitalizes on familiar green and red traffic light stimuli that activate existing stimulus–response associations (Ceceli et al., 2019). If green-Go and red-NoGo associations are habit-driven, an incongruent Go/NoGo mapping (green-NoGo, red-Go) should produce significant decrements in accuracy. Importantly, Go/NoGo mappings that involve novel stimuli with no significant behavioral representations (i.e., blue and purple light stimuli) should evoke no mapping-related performance impairments. If ADHD is associated with heightened habitual control, symptom severity might track the mapping-related impairments elicited by the familiar Go/NoGo stimuli (e.g., higher symptom severity scores should predict heightened errors of commission – response execution when instructed to withhold). Furthermore, if performance and monetary feedback are effective in restoring goal-directed control, this dual feedback delivery should protect against the mapping-related accuracy impairment, preventing the increase in commission errors when Go and NoGo associations are incongruent with daily experiences. Similarly, such a disruption in habits may also be correlated to ADHD symptom severity, such that a more severe presentation of ADHD symptoms may be less affected by the beneficial effects of feedback. Alternatively, if feedback is a salient enough motivator, highly symptomatic individuals may also benefit from our feedback manipulation, resulting in habit disruption across the board. To reveal whether ADHD is associated with habitual control, and whether a habit-dominated motivational control system may be remediated, we administered our well-learned habit task over the course of 2 days on a large sample from the general population, from whom we collected ADHD-related symptomology information. On the first day, we examined the execution of well-learned habits in our sample, and on the second day, we introduced our motivational enhancement manipulation – a combined delivery of performance information and monetary feedback – to restore goal-directed control. Importantly, per our pre-registered analysis plan (document URL)¹, we used ADHD-related measures to detect whether symptoms of the disorder tracked well-learned habit expression and disruption.

MATERIALS AND METHODS

Participants

To determine the sample size for our study, we performed an *a priori* power analysis on data from an existing study that examined inhibitory control capacity and ADHD-related symptoms (Wodushak and Neumann, 2003). In this study, healthy adults were categorized into high vs. low ADHD symptom groups for inhibitory control comparisons. We extracted effect sizes from the correlations between inhibitory control and non-verbal inattention in both symptom severity groups, and averaged the two resulting projected sample sizes.

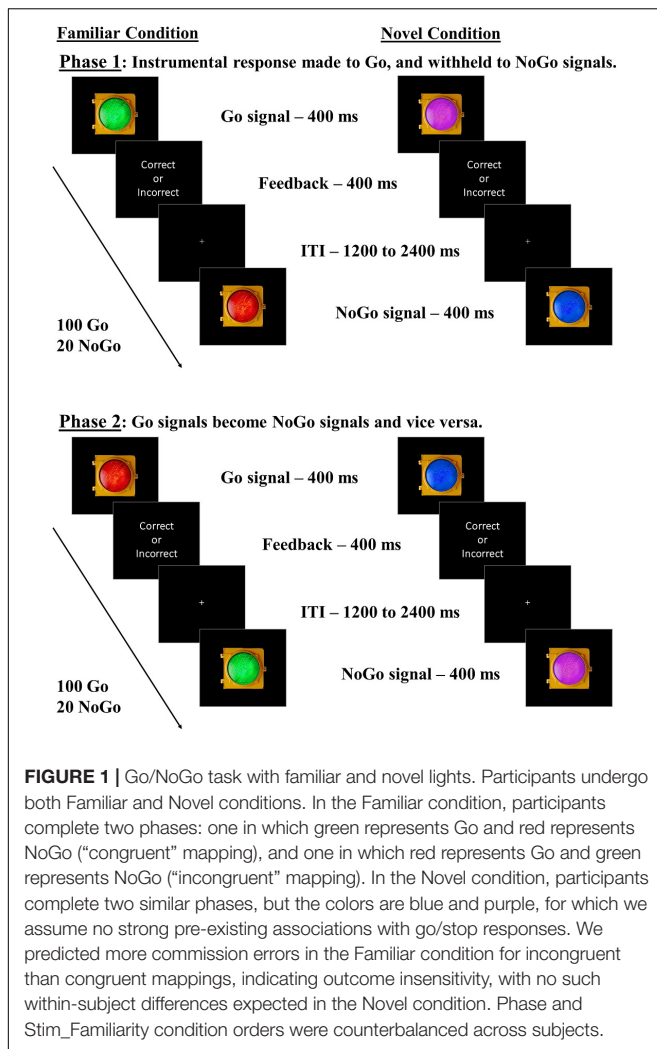
¹<https://osf.io/fjcbw>

The averaged sample size needed to reach 80% statistical power was determined to be 105. We recruited 106 participants to make up for one participant's corrupted data. Thus, 106 undergraduate students (79 female, 27 male; $M_{\text{age}} = 20.23$, $SD_{\text{age}} = 4.07$) from the Rutgers University-Newark campus participated for course credit. Informed consent was provided by all subjects per Declaration of Helsinki human subject protection guidelines. The Rutgers University Institutional Review Board approved study protocols. Individuals were excluded from participation for self-reported color-blindness. Two participants' data were excluded from analyses due to attrition ($n = 1$) and data corruption ($n = 1$). Thus, the statistical analyses were performed on the remaining 104 participants (77 female, 27 male participants; $M_{\text{age}} = 20.20$, $SD_{\text{age}} = 4.10$).

Materials and Procedures

Participants performed Go/NoGo tasks adapted from Ceceli et al. (2019) over 2 days. On day one, all participants underwent Go/NoGo tasks with familiar green and red traffic light stimuli (Familiar condition), and novel blue and purple traffic light stimuli (Novel condition) as Go and NoGo signals. Participants were instructed to respond as quickly and accurately to these stimuli as possible using the keyboard. A second phase followed in each Stim_Familiarity condition (Familiar/Novel conditions), where the color-response mappings were swapped (see **Figure 1**). In the Familiar condition, the Green-Go/Red-NoGo color-response mapping was considered “congruent” with daily experiences, while the Red-Go/Green-NoGo mapping was considered “incongruent,” in that it required the participant to override the well-established go and stop meanings of these stimuli. The Novel condition stimuli, however, are assumed to have no well-established Go or NoGo associations in daily life, in that the swapping of the color-response mappings should not require overriding associations that have been well-established. If familiar associations elicit habitual, cue-driven behavioral control, participants should experience a significant impairment in NoGo accuracy when green is mapped with NoGo. In the Novel condition, participants should perform similarly when managing either color-response mapping due to blue and purple not being strongly associated with Go/NoGo signals, reflecting goal-directed performance. We counterbalanced the order in which participants underwent the two phases within each Stim_Familiarity condition to ensure that our results were not due to a specific order of managing color-response contingencies. We also counterbalanced the order in which participants underwent the Familiar and Novel conditions. Lastly, participants completed the Adult ADHD Self-Report Scale (ASRS), a two-part survey that captures inattentive and hyperactive symptom manifestation associated with ADHD (Kessler et al., 2005a), and a demographic survey, concluding day one's procedures.

Day two was completed within 3 days of day one and examined the potential habit-disrupting effect of a motivational enhancement. We separated these sessions by at least 1 day to minimize potential training effects. On day two, all participants underwent the Familiar condition of the Go/NoGo task, completing the “congruent” color-mapping first. Next, we



induced motivational enhancement via the delivery of cumulative performance feedback and a monetary incentive. Specifically, participants’ cumulative task performance was displayed as a percentage score on the screen. Additionally, the experimenter briefly left the room, returning shortly after with a \$5 cash bonus. The participants were informed that the \$5 bonus was due to their performance on the task. The participants were then instructed to perform the “incongruent” color-mapping of the Familiar condition, and were informed that they may receive another performance-contingent cash bonus afterward. Unbeknownst to the participants, the mid-session cash bonus was not actually contingent on performance. We did not counterbalance color-mapping of Go/NoGo contingencies on day two to render the congruent color-mapping performance as baseline. Thus, we were able to test whether the presence of a mid-experiment motivational manipulation affected subsequent incongruent color-mapping performance (i.e., overriding the green-Go/red-NoGo habit). Lastly, participants completed the Creature of Habit Survey (COHS) (Ersche et al., 2017), quantifying the frequency of daily habitual tendencies, and a brief post-experiment questionnaire.

In each phase, there was a 5:1 Go/NoGo ratio, with 100 Go and 20 NoGo trials. Each Go/NoGo stimulus remained on the screen for 400 ms. Participants were required to respond to Go signals before the offset of the stimulus for a correct response. After offset, each response produced a brief “correct” or “incorrect” text slide. To ensure engagement with the task, inter-trial intervals varied randomly between 1200 and 2400 ms. Participants completed a practice session prior to each Stim_Familiarity condition, which consisted of six correct Go or NoGo responses using that condition’s stimuli. The experimenter remained present to ensure the instructions were understood during the practice sessions.

Data Analysis

We pre-registered our task procedures and analyses prior to data collection via the Open Science Framework project registration portal (document URL: see text footnote 1). Analyses that were not outlined in our pre-registration document are marked as exploratory below. Data analysis was performed using the nlme package in R (version 3.5.1).

We used NoGo accuracy as our primary measure of outcome-sensitivity, as the moderate Go to NoGo ratio was hypothesized to produce prepotent Go responses (Young et al., 2018). NoGo accuracy has been the gold standard in studying behavioral control (Schulz et al., 2007; Meule, 2017). We selected this measure as our primary outcome of interest because our hypotheses are grounded in the idea that overriding the prepotent Go response will differ based on the real-world familiarity associated with color-response mappings in the task, and be further driven by ADHD symptom severity. As a secondary measure of outcome-sensitivity, we also performed all analyses using Go accuracy to supplement our assertions of differential outcome-sensitivity across Familiar and Novel conditions, and reveal the potential role of ADHD symptom severity in contributing to outcome-sensitivity. An alternative method of reporting Go/NoGo results is centered on the signal detection approach, in which Z-scored “hits” are subtracted from Z-scored “false alarms” to derive a sensitivity bias estimate for that particular run (Stanislaw and Todorov, 1999). However, this approach may complicate extracting color-specific accuracy information that is spread out over multiple runs—for example, extracting a sensitivity bias for green would require hits from the congruent, and false alarms from the incongruent run. Nonetheless, when sensitivity biases are derived on familiarity and congruency (e.g., when measured using Green-Go hits together with Red-NoGo false alarms to yield a sensitivity bias for the familiar-congruent mapping) the results mirror the analyses reported here using traditional accuracy rates. The corresponding signal detection analyses can be found in our shared analysis scripts and data output materials in the section **Supplementary Data Sheet 1, “Signal Detection Analyses”** in **Supplementary Material**.

Participants with standardized residuals less than -3.3 and greater than 3.3 were identified as outliers (Tabachnick and Fidell, 2007). Analyses excluding outliers are reported if data removal produces substantial changes in results (i.e., changes in statistical significance of any regressor). Bootstrapped

95% confidence interval values for all model regressors are included in their corresponding data tables (1000 bootstrap iterations in each model).

ADHD Symptom Severity and Well-Learned Habits

We performed an omnibus hierarchical multiple regression test to discern the contributions of symptom severity on outcome-sensitivity within Familiar and Novel condition data collected on day 1. This hierarchical structure permitted us to extract information about the amount of variance explained by groups of regressors (i.e., controlled variables, individual difference measures, and experimental variables), while also obtaining the predictive strengths of each individual regressor. Importantly, each additional step in the hierarchy updates the parameter estimates of the regressors in the previous steps, such that we are also able to detect how controlled variables may influence other regressors of interest. We used $\Delta\text{NoGo_Accuracy}$ (i.e., change in NoGo accuracy scores across mappings) as our dependent variable (DV) to measure the within-subject mapping-related change in accuracy. A greater mapping-related impairment represents greater outcome-insensitivity (e.g., heightened difficulty overriding a color-response mapping). In a hierarchical structure, we first input the regressors Age, Gender, Stim_Familiarity_Order (order in which participants underwent Familiar and Novel conditions), Phase_Order (order in which participants underwent color-response mappings within each Stim_Familiarity condition), and Driving (each participant's experience driving, scaled in months), with Subject as a random factor into a linear mixed model. This model extracted the predictive strength of each of these controlled variables on outcome-sensitivity. In the next hierarchical step, we added the regressors ASRS_Inattentive (part A of the ASRS measure capturing symptoms of inattention), ASRS_Hyperactive (part B of the ASRS measure capturing symptoms of hyperactivity), and ASRS_Total (parts A and B aggregated to derive a composite score of ADHD symptom severity). Because our sample included six participants who had received ADHD diagnoses, we also input a Diagnosis regressor to determine whether clinical manifestation of ADHD – albeit in a small proportion of participants – affects outcome-sensitivity. We used COHS scores as a regressor to find potential correlations with tendency to behave habitually in daily life and outcome-sensitivity in our task. These regressors served to explain the main effects of each individual difference measure on outcome-sensitivity. In the third step of the hierarchical model, we input Stim_Familiarity (Familiar/Novel) as a regressor to specifically detect whether participants exhibited differential outcome-sensitivity across Familiar and Novel conditions. A significant contribution of this variable would confirm that the familiar red and green stimuli indeed elicit outcome-insensitive, habitual control, while the novel stimuli are labile, and thus controlled by goal-directed processes. We performed *post hoc t*-tests of NoGo accuracy between phases in each Stim_Familiarity condition to ascertain differential mapping-related impairment across Familiar and Novel conditions. Lastly, because of our specific focus on the influence of ADHD symptomology on habitual control, we also entered all individual difference measures' interactions with Stim_Familiarity as

regressors (e.g., ADHD_Inattentive \times Stim_Familiarity) into step four of the model. Thus, we were able to distinguish the effects of each variable on outcome-sensitivity across Familiar and Novel conditions.

In brief, we expected the controlled demographic and counterbalancing variables (Age, Gender, Driving, Stim_Familiarity_Order, and Phase_Order) to be trivial in predicting outcome-sensitivity. We did not expect the Driving regressor to play a significant role in altering outcome-sensitivity, as we expect our well-learned habit task to capture well-established associations that extend beyond experience with these color-response mappings in a traffic context. We input both main effect and interaction regressors related to individual differences in ADHD symptomology and daily habitual tendencies to reveal potential associations with outcome-sensitivity. This way, we were able to inquire whether these individual difference regressors yielded strong associations with global outcome-sensitivity (i.e., main effects predicting mapping-related impairments independent of stimulus familiarity), and further interrogate whether such an association existed with well-learned habit expression in particular (i.e., ADHD-related measure \times Stim_Familiarity interaction predicting an effect on outcome-sensitivity differentially across Familiar/Novel conditions). We also expected Stim_Familiarity to serve as a significant predictor in driving outcome-sensitivity, as the Familiar condition stimuli should selectively elicit outcome-insensitive habits, while the Novel condition stimuli should have no such effect on behavior.

ADHD Symptom Severity and Habit Disruption

We have previously shown the habit-disrupting effect of cumulative performance and monetary feedback (Ceceli et al., 2019). Here, we test via another omnibus regression whether ADHD symptom severity predicts habit disruption success. We performed a similar linear mixed model on the aggregate of Familiar data across 2 days, encompassing performance to the Familiar stimuli with and without feedback. We input our controlled variables of Age, Gender, Driving, Stim_Familiarity_Order, and Phase_Order, with Subject as a random factor into the first step. Our model similarly included ASRS_Inattentive, ASRS_Hyperactive, ASRS_Total, Diagnosis, and COHS in the second step to detect the main effects of individual differences on outcome-sensitivity. In the third step, our regression included a Feedback regressor that coded the availability of the mid-experiment dual-feedback manipulation. Because this analysis was performed only on the Familiar condition data (the Novel condition was not administered on the second day with feedback), we included no Stim_Familiarity regressor. Lastly, we included in step 4 our individual difference measures' interactions with Feedback as regressors (e.g., ASRS_Inattentive \times Feedback) to examine habit disruption per variations in ADHD-related behaviors and daily habitual tendencies.

Similar to our previous omnibus regression, we expected trivial contribution from our controlled variables, but a significant contribution from the Feedback regressor, as

the delivery of dual feedback should disrupt the well-learned habit. We expected that symptom severity may affect outcome-sensitivity globally (significant main effects of individual difference measures), but also differentially across Feedback sessions (e.g., significant contribution of ADHD_Inattentive \times Feedback). Additionally, we identified an alternative hypothesis – the possibility of habit disruption across the board (pre-registration document, Hypothesis 2b_alt). We expected no directionality in subtypes governing outcome-sensitivity (as in, inattentiveness or hyperactivity specifically driving habits), but we do note that if either subtype plays a major role in driving motivational control in the previous omnibus regression detecting the role of symptom severity on habitual control, that same subtype should predict habit disruption. We expected the frequency of habitual tendencies in daily life, as assayed by COHS, to yield a negative correlation with habit disruption (i.e., a significant COHS \times Feedback result).

Supplementary Index of Outcome-Sensitivity: Go Accuracy

We used Go accuracy as a supplemental measure of outcome-sensitivity. Thus, we repeated all mixed models that examined Δ NoGo_Accuracy using Δ Go_Accuracy as DV.

Exploratory Analyses: Go RT and Individual Difference Measures

We extended our analyses beyond the pre-registered plans and explored the potential correlations between Go reaction time (RT) and our individual difference measures of symptom severity (ASRS_Inattentiveness and ASRS_Hyperactivity) and daily habitual tendencies (COHS). These variables were entered into a correlation matrix, and Pearson's r values were corrected for multiple comparisons using the Holm–Bonferroni method. Specifically, we expected a negative correlation between RT and our individual difference measures. Most notably, we expected such an association between RT and ASRS_Hyperactivity, which would suggest quicker familiar Go actions to be associated with pronounced hyperactivity.

RESULTS

ADHD Symptom Severity and Well-Learned Habits

We performed a linear mixed model using Δ NoGo_Accuracy as the DV and Subject as a random factor to determine whether ADHD symptom severity significantly predicts outcome-sensitivity in our well-learned habit task. Our proposed model violated the assumptions of non-multicollinearity, in that three pairs of fixed factors were highly correlated with each other (for the associated Variance Inflation Factors, see section “Supplementary Material”). Thus, we report the analyses as registered in the section “Supplementary Material,” and report below an adjusted model that meets the assumptions of non-multicollinearity, normality and homoscedasticity (see Table 1). Specifically, we revised our model to remove the regressors Age, Stim_Familiarity_Order, and ASRS_Total to prevent

multicollinearity with the regressors Driving, Phase_Order, and ASRS_Inattentive/Hyperactive that are more crucial for our hypotheses.

Standard within-group residuals were within -3.3 and 3.3 ; thus, no participants were identified as outliers (Tabachnick and Fidell, 2007). In the first step of our hierarchical mixed model, contrary to our hypothesis, Gender significantly predicted outcome-sensitivity, $\beta_{\text{Gender}} = -0.15$, $p = 0.036$, in that female participants displayed significantly worse mapping-related impairments. Neither Driving experience nor the counterbalancing variable, Phase_Order, predicted outcome-sensitivity ($ps > 0.252$), model $R^2 = 0.03$. In the second step of the model, we added the individual difference measures of ADHD symptom severity, clinical ADHD diagnosis, and frequency of habitual tendencies in daily life (COHS). We found no main effects of individual difference measures on outcome-sensitivity (all $ps > 0.548$). The log likelihood estimate derived by comparing first and second steps of our model yielded no significant global (as in, non-Stim_Familiarity specific) contribution attributable to the ASRS_Inattentive, ASRS_Hyperactive, Diagnosis, and COHS regressors, $\chi^2(4) = 0.70$, $p = 0.952$, $R^2 = 0.03$, $\Delta R^2 < 0.01$. In the third step, we entered the Stim_Familiarity regressor, which significantly improved the predictive strength of the model, $\chi^2(1) = 21.53$, $p < 0.001$, $R^2 = 0.13$, $\Delta R^2 = 0.10$, $\beta_{\text{Stim_Familiarity}} = 0.31$, $t(103) = 4.66$, $p < 0.001$, meaning outcome-sensitivity was differentially affected by whether participants managed the Familiar or Novel versions of the task. *Post hoc t*-tests confirmed that mapping-related NoGo accuracy impairments were evident only when managing Go/NoGo contingencies in the Familiar condition, $t(103) = 5.33$, $p < 0.001$, while performance in the Novel condition was comparable regardless of color-mapping associations, $t(103) = -1.09$,

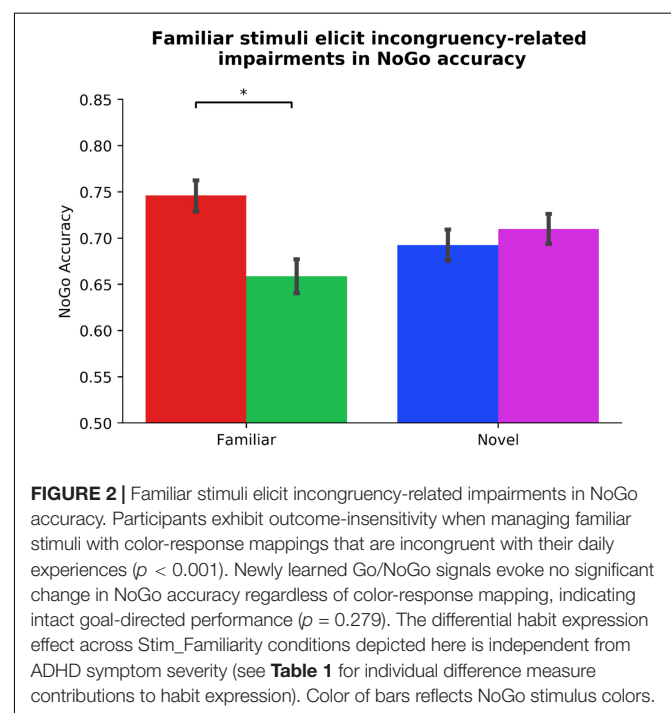


TABLE 1 | Hierarchical mixed model of ADHD symptomology and habit expression: $\Delta\text{NoGo_Accuracy}$.

Variable	VIF	β (SE)	B [95% CI]	t	sig.
Model 1					
Gender	1.01	-0.15 (0.07)	-0.06 [- 0.11, > -0.01]	-2.13	0.036
Phase_Order	1.01	-0.01 (0.07)	> -0.01 [- 0.02, 0.02]	-0.09	0.931
Driving	1.00	0.08 (0.07)	< 0.01 [> -0.01, < 0.01]	1.15	0.252
Model 2					
Gender	1.08	-0.14 (0.07)	-0.05 [- 0.11, > -0.01]	-2.00	0.049
Phase_Order	1.04	-0.01 (0.07)	> -0.01 [- 0.02, 0.02]	-0.15	0.877
Driving	1.30	0.08 (0.08)	< 0.01 [> -0.01, < 0.01]	1.08	0.283
ASRS_Inattentive	1.62	-0.01 (0.09)	> -0.01 [- 0.01, < 0.01]	-0.08	0.939
ASRS_Hyperactive	1.71	0.05 (0.09)	< 0.01 [> -0.01, 0.01]	0.54	0.591
Diagnosis	1.30	0.01 (0.08)	0.01 [- 0.11, 0.12]	0.14	0.891
COHS	1.06	-0.04 (0.07)	> -0.01 [> -0.01, < 0.01]	-0.60	0.548
Model 3					
Gender	1.08	-0.14 (0.07)	-0.05 [- 0.10, > -0.01]	-2.10	0.039
Phase_Order	1.04	-0.01 (0.07)	> -0.01 [- 0.02, 0.02]	-0.16	0.871
Driving	1.30	0.09 (0.08)	< 0.01 [> 0.01, < 0.01]	1.13	0.260
ASRS_Inattentive	1.62	-0.01 (0.08)	> -0.01 [> -0.01, < 0.01]	-0.08	0.936
ASRS_Hyperactive	1.71	0.05 (0.09)	< 0.01 [> -0.01, 0.01]	0.57	0.573
Diagnosis	1.30	0.01 (0.08)	0.01 [- 0.10, 0.11]	0.14	0.885
COHS	1.06	-0.04 (0.07)	> -0.01 [> -0.01, < 0.01]	-0.63	0.528
Stim_Familiarity	1	0.31(0.07)	0.10 [0.06, 0.15]	4.66	< 0.001
Model 4					
Gender	1.08	-0.14 (0.07)	-0.05 [- 0.10, > -0.01]	-2.11	0.039
Phase_Order	1.04	-0.01 (0.07)	> -0.01 [- 0.02, 0.02]	-0.16	0.871
Driving	1.30	0.09 (0.08)	< 0.01 [> -0.01, < 0.01]	1.14	0.260
ASRS_Inattentive	3.17	-0.01 (0.08)	< 0.01 [- 0.01, 0.01]	-0.08	0.936
ASRS_Hyperactive	3.31	0.05 (0.09)	< 0.01 [> -0.01, 0.01]	0.57	0.573
Diagnosis	2.35	0.01 (0.08)	-0.06 [- 0.20, 0.08]	0.14	0.885
COHS	2.12	-0.04 (0.07)	> -0.01 [> -0.01, < 0.01]	-0.64	0.528
Stim_Familiarity	64.79	0.31 (0.07)	0.13 [- 0.47, 0.22]	4.68	< 0.001
ASRS_Inattentive \times Stim_Familiarity	16.51	-0.02 (0.08)	> -0.01 [- 0.01, 0.01]	-0.29	0.774
ASRS_Hyperactive \times Stim_Familiarity	13.60	-0.05 (0.08)	> -0.01 [- 0.01, 0.01]	-0.56	0.575
Diagnosis \times Stim_Familiarity	2.16	0.10 (0.07)	0.15 [- 0.04, 0.34]	1.51	0.134
COHS \times Stim_Familiarity	57.37	0.12 (0.07)	< 0.01 [> -0.01, 0.01]	1.74	0.085
Model comparisons					
Model	R^2	Log likel.	χ^2	χ^2 sig.	ΔR^2
Model 1	0.03	79.53			
Model 2	0.03	79.87	0.70	0.952	< 0.01
Model 3	0.13	90.64	21.53	< 0.001	0.10
Model 4	0.15	93.73	6.19	0.186	0.03

Top layer of table depicts all regressors included in the hierarchical model. Model Comparisons layer depicts the predictive strength of each model, as compared to its previous step. VIF, Variance Inflation Factor; SE, Standard Error; CI, Confidence Interval; Log likel., Log likelihood. Significant p-values depicted in bold typeface. Analyses have been outlier corrected, with resulting deviations highlighted in the text. 95% confidence intervals were obtained by bootstrapping 1000 samples in each model.

$p = 0.279$ (see **Figure 2**). In the fourth step of the model, we input the interaction of each individual difference regressor with Stim_Familiarity to detect their potentially differential effects on outcome-sensitivity across Familiar and Novel conditions, but found no significant contribution from any ADHD-related or daily habit frequency variable (all $ps > 0.085$, $\chi^2(4) = 6.19$, $p = 0.186$, $R^2 = 0.15$, $\Delta R^2 = 0.03$). These results suggest that our sample exhibited outcome-insensitive well-learned habits across the board, but the degree of habitual control as assessed by

change in NoGo accuracy was not significantly related to ADHD symptom severity.

ADHD Symptom Severity and Habit Disruption

Similarly, we altered our pre-registered model to prevent multicollinearity, and performed a linear mixed model to examine the link between ADHD symptomology and habit disruption (see **Table 2**). The pre-registered analysis that

TABLE 2 | Hierarchical mixed model of ADHD symptomology and habit disruption: $\Delta\text{NoGo_Accuracy}$.

Variable	VIF	β	B [95% CI]	t	sig.
Model 1					
Gender	1.01	0.04 (0.07)	0.02 [−0.03, 0.07]	0.60	0.553
Phase_Order	1.01	0.10 (0.07)	0.02 [−0.01, 0.04]	1.48	0.142
Driving	1.00	−0.02 (0.07)	> −0.01 [> −0.01, < 0.01]	−0.28	0.779
Model 2					
Gender	1.08	0.04 (0.07)	0.02 [−0.04, 0.07]	0.62	0.537
Phase_Order	1.04	0.09 (0.07)	0.01 [−0.01, 0.03]	1.25	0.215
Driving	1.30	0.02 (0.08)	> −0.01 [> −0.01, < 0.01]	0.24	0.807
ASRS_Inattentive	1.62	−0.06 (0.09)	> −0.01 [−0.01, < 0.01]	−0.69	0.491
ASRS_Hyperactive	1.71	0.10 (0.09)	< 0.01 [> −0.01, 0.01]	1.12	0.263
Diagnosis	1.30	−0.05 (0.08)	−0.04 [−0.16, 0.08]	−0.61	0.542
COHS	1.06	−0.10 (0.07)	> −0.01 [> −0.01, < 0.01]	−1.41	0.162
Model 3					
Gender	1.08	0.04 (0.07)	0.02 [−0.03, 0.07]	0.64	0.521
Phase_Order	1.04	0.09 (0.07)	0.01 [−0.01, 0.03]	1.30	0.198
Driving	1.30	0.02 (0.08)	> −0.01 [> −0.01, < 0.01]	0.25	0.799
ASRS_Inattentive	1.62	−0.06 (0.08)	> −0.01 [−0.01, < 0.01]	−0.72	0.474
ASRS_Hyperactive	1.71	0.10 (0.09)	< 0.01 [> −0.01, 0.01]	1.17	0.245
Diagnosis	1.30	−0.05 (0.08)	−0.04 [−0.15, 0.07]	−0.64	0.526
COHS	1.06	−0.10 (0.07)	> −0.01 [> −0.01, < 0.01]	−1.47	0.146
Feedback	1	−0.28 (0.07)	−0.09 [−0.14, −0.05]	−4.13	< 0.001
Model 4					
Gender	1.08	0.04 (0.07)	0.02 [−0.03, 0.07]	0.64	0.525
Phase_Order	1.04	0.09 (0.07)	0.01 [−0.01, 0.07]	1.28	0.202
Driving	1.30	0.02 (0.08)	> −0.01 [−0.01, 0.03]	0.25	0.801
ASRS_Inattentive	3.15	−0.06 (0.08)	> −0.01 [−0.01, < 0.01]	−0.71	0.478
ASRS_Hyperactive	3.29	0.10 (0.09)	< 0.01 [> −0.01, 0.01]	1.16	0.250
Diagnosis	2.34	−0.05 (0.08)	−0.05 [−0.19, 0.10]	−0.63	0.530
COHS	2.10	−0.10 (0.07)	> −0.01 [> −0.01, < 0.01]	−1.45	0.150
Feedback	64.79	−0.28 (0.07)	0.01 [−0.34, 0.37]	−4.12	< 0.001
ASRS_Inattentive × Feedback	16.49	0.05 (0.08)	< 0.01 [−0.01, 0.01]	0.62	0.539
ASRS_Hyperactive × Feedback	13.58	−0.01 (0.08)	> −0.01 [−0.01, 0.01]	−0.16	0.869
Diagnosis × Feedback	2.14	0.02 (0.07)	0.02 [−0.17, 0.21]	0.24	0.811
COHS × Feedback	57.35	0.06 (0.07)	> −0.01 [> −0.01, < 0.01]	−0.86	0.391
Model comparisons					
Model	R^2	Log likel.	χ^2	χ^2 sig.	ΔR^2
Model 1	0.01	72.53			
Model 2	0.03	74.13	3.19	0.526	0.01
Model 3	0.11	82.68	17.10	< 0.001	0.08
Model 4	0.11	83.46	1.56	0.815	0.01

Top layer of table depicts all regressors included in the hierarchical model. Model Comparisons layer depicts the predictive strength of each model, as compared to its previous step. VIF, Variance Inflation Factor. SE, Standard Error. CI, Confidence Interval. Log likel., Log likelihood. Significant p-values depicted in bold typeface. Analyses have been outlier corrected, with resulting deviations highlighted in the text. 95% confidence intervals were obtained by bootstrapping 1000 samples in each model.

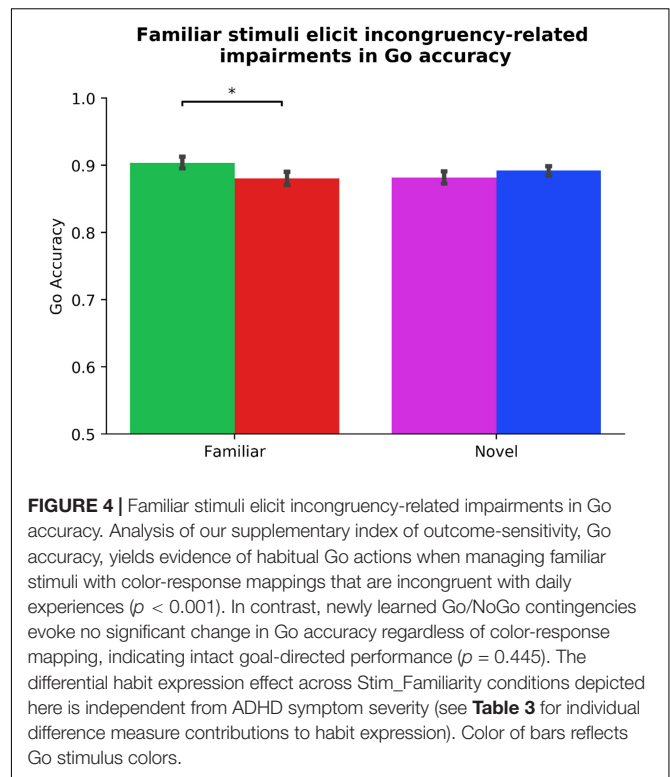
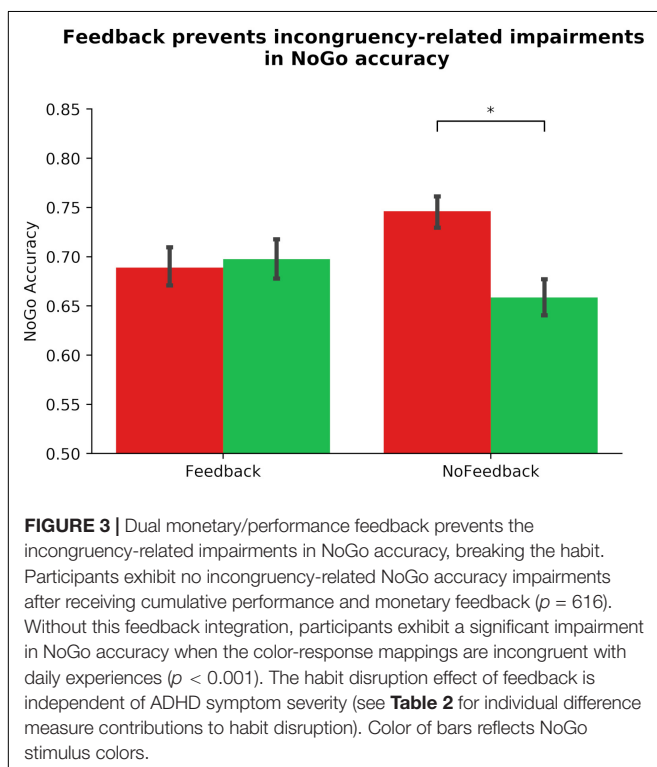
violated assumptions of multicollinearity can be found in the section “**Supplementary Material.**” In our corrected model, we input Gender, Phase_Order, and Driving experience into step one, where none significantly predicted outcome-sensitivity (all $ps > 0.142$), model $R^2 = 0.01$. In step two, we added ASRS_Inattentive, ASRS_Hyperactive, Diagnosis, and COHS into the model, and found that none of these regressors yielded main effects on outcome-sensitivity (all $ps > 0.162$), and they

did not significantly improve the predictive strength of the model, $\chi^2(4) = 3.19$, $p = 0.526$, $R^2 = 0.03$, $\Delta R^2 = 0.01$. We input Feedback as a regressor in step three, which contributed significantly to predicting outcome-sensitivity, $\beta_{\text{Feedback}} = -0.28$, $t(103) = -4.13$, $p < 0.001$, and rendered the model a significant predictor of $\Delta\text{NoGo_Accuracy}$, $\chi^2(1) = 17.10$, $p < 0.001$, $R^2 = 0.11$, $\Delta R^2 = 0.08$. We performed *post hoc* paired-samples *t*-tests to confirm the beneficial effect of dual feedback.

We found that a significant NoGo accuracy impairment was evident in absence of dual feedback, $t(103) = 5.33$, $p < 0.001$, whereas the delivery of feedback yielded no significant accuracy impairments, $t(103) = -0.50$, $p = 0.616$ (see **Figure 3**). No individual difference measures' interaction regressor in step four significantly predicted outcome-sensitivity (all $ps > 0.391$, $\chi^2(4) = 1.56$, $p = 0.815$, $R^2 = 0.11$, $\Delta R^2 = 0.01$). These results suggest that the delivery of dual feedback indeed had a protective effect on outcome-sensitivity when managing familiar stimuli, albeit independent of ADHD symptom severity.

Supplementary Analysis of ADHD Symptom Severity and Well-Learned Habits

We performed identical analyses using $\Delta\text{Go_Accuracy}$ as DV and Subject as a random factor to capture the potential association between ADHD symptomology and a supplemental assay of outcome-sensitivity (see **Table 3**). Two participants' data were identified as outliers. Due to changes in statistical significance following outlier correction, we report our outlier-removed dataset below, highlighting any change in statistical significance due to outlier correction. Neither Gender, Phase_Order, nor Driving experience predicted $\Delta\text{Go_Accuracy}$ (all $ps > 0.323$), model $R^2 = 0.01$. In step two, the Diagnosis regressor, which codes for the presence of a clinical ADHD diagnosis, made a significant contribution, $\beta_{\text{Diagnosis}} = 0.17$, $t(94) = 2.11$, $p = 0.038$ (without outlier correction: $\beta_{\text{Diagnosis}} = 0.14$, $t(96) = 1.80$, $p = 0.076$). Specifically, the presence of a diagnosis predicted more flexible Go actions. No



other step two regressor significantly predicted $\Delta\text{Go_Accuracy}$ (all $ps > 0.259$). The step two model was not significantly improved from step one, $\chi^2(4) = 5.56$, $p = 0.235$, $R^2 = 0.04$, $\Delta R^2 = 0.03$. The Stim_Familiarity regressor in step three served as a significant predictor, $\beta_{\text{Stim_Familiarity}} = 0.14$, $t(101) = 2.07$, $p = 0.010$, improving the predictive strength of the model, $\chi^2(1) = 4.44$, $p = 0.035$, $R^2 = 0.06$, $\Delta R^2 = 0.02$. Paired-samples t -tests revealed a significant Go accuracy impairment in the Familiar condition, $t(101) = 3.80$, $p < 0.001$, but not the Novel condition, $t(101) = -0.77$, $p = 0.445$ (see **Figure 4**). Lastly in step four, other than $\text{Diagnosis} \times \text{Stim_Familiarity}$, $\beta_{\text{Diagnosis} \times \text{Stim_Familiarity}} = 0.19$, $t(97) = 2.71$, $p = 0.008$, no individual difference measures significantly predicted $\Delta\text{Go_Accuracy}$ across the Familiar and Novel conditions (all other interaction $ps > 0.125$, $\chi^2(4) = 10.43$, $p = 0.034$, $R^2 = 0.10$, $\Delta R^2 = 0.05$). Because we only had six individuals with an ADHD diagnosis, we refrain from further interpretations of the contribution of the Diagnosis regressor. These results suggest that Go accuracy is differentially affected by whether familiar or novel stimuli serve as Go/NoGo signals, and a significant impairment is evident when familiar contingencies are incongruent with daily experiences. However, the habitual Go actions elicited by our familiar stimuli are independent of ADHD symptom severity.

Supplementary Analysis of ADHD Symptom Severity and Habit Disruption

We investigated habit disruption via mapping-related changes in Go accuracy using a similar mixed model (see **Table 4**). Our

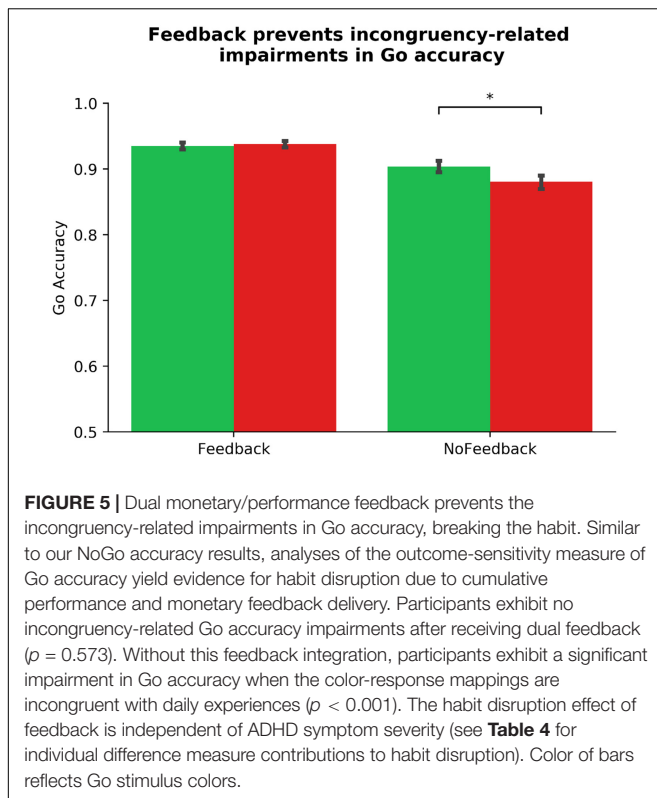
TABLE 3 | Hierarchical mixed model of ADHD symptomology and habit expression: $\Delta\text{Go_Accuracy}$.

Variable	VIF	β	B [95% CI]	t	sig.
Model 1					
Gender	1.01	< 0.01 (0.07)	< 0.01 [- 0.03, 0.03]	-0.01	0.997
Phase_Order	1.01	0.07 (0.07)	< 0.01 [- 0.01, 0.01]	0.99	0.323
Driving	1.00	0.06 (0.07)	< 0.01 [> -0.01, < 0.01]	0.88	0.383
Model 2					
Gender	1.08	0.02 (0.07)	0.01 [- 0.02, 0.04]	0.34	0.731
Phase_Order	1.04	0.08 (0.07)	< 0.01 [- 0.01, 0.01]	1.13	0.260
Driving	1.30	-0.01 (0.08)	> -0.01 [> -0.01, < 0.01]	-0.23	0.815
ASRS_Inattentive	1.59	0.02 (0.09)	> -0.01 [> -0.01, < 0.01]	0.22	0.828
ASRS_Hyperactive	1.66	-0.03 (0.09)	> -0.01 [> -0.01, < 0.01]	-0.39	0.699
Diagnosis	1.30	0.17 (0.08)	0.06 [- 0.01, 0.12]	2.11	0.038
COHS	1.06	-0.03 (0.07)	> -0.01 [> -0.01, < 0.01]	-0.41	0.681
Model 3					
Gender	1.09	0.02 (0.07)	0.01 [- 0.02, 0.04]	0.35	0.729
Phase_Order	1.04	0.08 (0.07)	< 0.01 [- 0.01, 0.01]	1.14	0.256
Driving	1.30	-0.02 (0.08)	> -0.01 [> -0.01, < 0.01]	-0.24	0.813
ASRS_Inattentive	1.59	0.02 (0.09)	> -0.01 [> -0.01, < 0.01]	0.22	0.826
ASRS_Hyperactive	1.66	-0.03 (0.09)	> -0.01 [> -0.01, < 0.01]	-0.39	0.696
Diagnosis	1.30	0.17 (0.08)	0.06 [> -0.01, 0.12]	2.13	0.036
COHS	1.06	-0.03 (0.07)	> -0.01 [> -0.01, < 0.01]	-0.42	0.678
Stim_Familiarity	1	0.14 (0.07)	0.03 [0.01, 0.06]	2.07	0.041
Model 4					
Gender	1.09	0.02 (0.07)	0.01 [- 0.02, 0.04]	0.35	0.729
Phase_Order	1.04	0.08 (0.07)	< 0.01 [- 0.01, 0.01]	1.16	0.256
Driving	1.30	-0.02 (0.08)	> -0.01 [> -0.01, < 0.01]	-0.24	0.813
ASRS_Inattentive	3.11	0.02 (0.09)	> -0.01 [> -0.01, 0.01]	0.22	0.826
ASRS_Hyperactive	3.23	-0.03 (0.09)	> -0.01 [- 0.01, < 0.01]	-0.40	0.696
Diagnosis	2.35	0.17 (0.08)	> -0.01 [- 0.08, 0.07]	2.16	0.036
COHS	2.12	-0.03 (0.07)	> -0.01 [> -0.01, < 0.01]	-0.42	0.678
Stim_Familiarity	65.83	0.14 (0.07)	< 0.01 [- 0.19, 0.21]	2.10	0.038
ASRS_Inattentive \times Stim_Familiarity	16.65	-0.08 (0.08)	> -0.01 [- 0.01, < 0.01]	-0.95	0.343
ASRS_Hyperactive \times Stim_Familiarity	13.71	0.04 (0.09)	< 0.01 [> -0.01, 0.01]	0.53	0.599
Diagnosis \times Stim_Familiarity	2.16	0.19 (0.07)	0.12 [0.01, 0.23]	2.71	0.008
COHS \times Stim_Familiarity	57.08	0.11 (0.07)	< 0.01 [> -0.01, < 0.01]	1.55	0.125
Model comparisons					
Model	R^2	Log likel.	χ^2	χ^2 sig.	ΔR^2
Model 1	0.01	218.44			
Model 2	0.04	221.22	5.56	0.235	0.03
Model 3	0.06	223.44	4.44	0.035	0.02
Model 4	0.10	228.65	10.40	0.034	0.05

Top layer of table depicts all regressors included in the hierarchical model. Model Comparisons layer depicts the predictive strength of each model, as compared to its previous step. VIF, Variance Inflation Factor; SE, Standard Error; CI, Confidence Interval. Log likel., Log likelihood. Significant p -values depicted in bold typeface. Analyses have been outlier corrected, with resulting deviations highlighted in the text. 95% confidence intervals were obtained by bootstrapping 1000 samples in each model.

multicollinearity-corrected model identified two outliers. We report outlier-removed results below, accompanied by any changes in statistical significance following outlier correction. In step one of the mixed model, no controlled regressors predicted $\Delta\text{Go_Accuracy}$ (all $ps > 0.093$), model $R^2 = 0.02$. In step two, COHS was a near significant variable, $\beta_{\text{COHS}} = -0.14$, $t(94) = -1.95$, $p = 0.054$ (without outlier-correction: $\beta_{\text{COHS}} = -0.08$, $t(96) = -1.05$, $p = 0.296$), suggesting

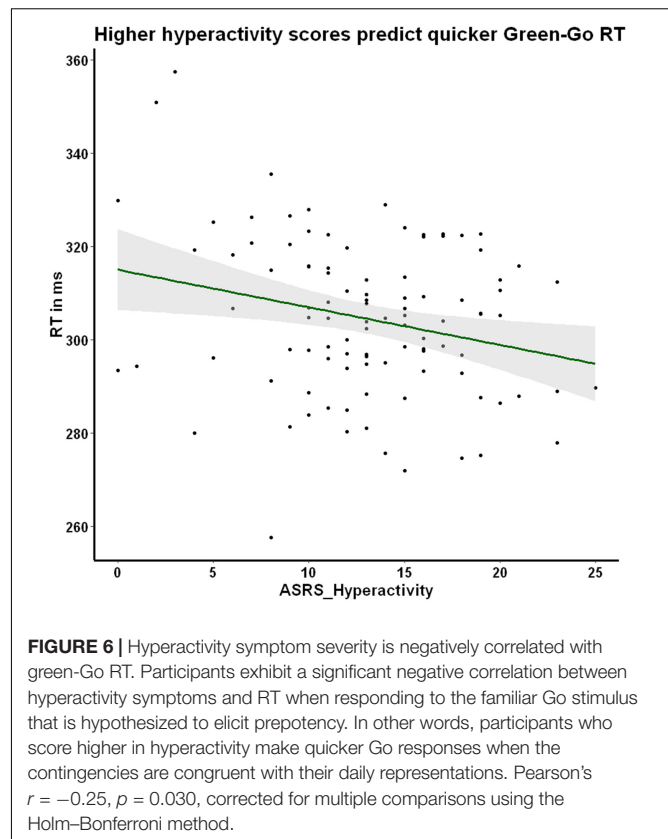
that a higher frequency of daily habits may predict more outcome-insensitive Go actions. Otherwise, no individual difference regressor served as a significant predictor of $\Delta\text{Go_Accuracy}$ (all other $ps = 0.149$), although the inclusion of step two regressors resulted in the Phase_Order variable to yield a near-significant p -value, $p = 0.066$. Step two regressors in aggregate yielded only a near-significant contribution on the DV, $\chi^2(4) = 8.56$, $p < 0.073$, $R^2 = 0.06$, $\Delta R^2 = 0.04$. In step



three, the Feedback regressor significantly predicted outcome-sensitivity as indexed by $\Delta\text{Go_Accuracy}$, $\beta_{\text{Feedback}} = -0.26$, $t(101) = -4.07$, $p < 0.001$, improving the predictive strength of the model, $\chi^2(1) = 16.01$, $p < 0.001$, $R^2 = 0.13$, $\Delta R^2 = 0.07$. This finding suggests that outcome-sensitivity as assessed by $\Delta\text{Go_Accuracy}$ is differentially impacted depending on the availability of dual feedback. Indeed, a *post hoc* paired-samples *t*-test confirms a significant impairment in Go accuracy when no feedback is delivered, $t(103) = 3.85$, $p < 0.001$, whereas with feedback, no such impairment is evident, $t(103) = -0.56$, $p = 0.573$ (see **Figure 5**). In step four, we found that $\text{COHS} \times \text{Feedback}$ significantly predicted habit disruption, $\beta_{\text{COHS} \times \text{Feedback}} = -0.16$, $t(97) = -2.46$, $p = 0.016$ (without outlier-correction: $p = 0.120$), suggesting that an increased daily habit frequency predicts a reduction in the beneficial effects of dual feedback in restoring goal-directed control. No other individual difference \times Feedback regressor predicted habit disruption (all $ps > 0.188$, $\chi^2(4) = 9.70$, $p = 0.046$, $R^2 = 0.16$, $\Delta R^2 = 0.04$). Similar to our primary measure of outcome-sensitivity using NoGo accuracy, the protective effect of dual feedback on Go accuracy was independent from ADHD symptomology. However, we do observe a significant association between habitual tendencies in daily life and a difficulty in suppressing a well-learned habit.

Exploratory Analyses: Go RT and Individual Difference Measures

We explored the potential association between prepotency to respond to the familiar Go stimulus and our individual difference



measures of ADHD symptom severity (ASRS_Inattentiveness and ASRS_Hyperactivity) and daily habit frequency (COHS). We reasoned that hyperactive individuals may exhibit a more pronounced prepotency to respond to Go stimuli, thus we were especially interested in the hyperactivity scale's association with RT. As hypothesized, we found a significant negative correlation between Go RT to the familiar green-Go color-response mapping and ASRS_Hyperactivity, $r = -0.25$, $p = 0.030$, Holm–Bonferroni corrected (**Figure 6**), suggesting that higher hyperactivity scores are associated with faster Go responses. This relationship between hyperactivity and response latency was not apparent when the Go signal was incongruent with lifelong experiences (red-Go $r = -0.05$, $p = 1$, Holm–Bonferroni corrected), or when the Novel condition stimuli served as the Go signal (purple-Go $r = -0.12$, $p = 0.630$; blue-Go $r = -0.10$, $p = 0.770$, Holm–Bonferroni corrected). The association between familiar Go RT and ASRS_Hyperactivity may suggest that individuals high in hyperactive symptoms may be exhibiting abnormally pronounced prepotency to stimuli that evoke habitual control.

DISCUSSION

The neurobehavioral evidence of atypical reward-related processes in ADHD, and the scarcity of strategies to restore potential behavioral rigidities, motivated us to examine the expression and disruption of well-learned habits as a

TABLE 4 | Hierarchical mixed model of ADHD symptomology and habit disruption: $\Delta\text{Go_Accuracy}$.

Variable	VIF	β	B [95% CI]	t	sig.
Model 1					
Gender	1.02	−0.03 (0.07)	−0.02 [− 0.02, 0.01]	−0.41	0.684
Phase_Order	1.02	0.12 (0.07)	< 0.01 [> −0.01, 0.01]	1.69	0.093
Driving	1.00	−0.03 (0.07)	> −0.01 [> −0.01, < 0.01]	−0.49	0.623
Model 2					
Gender	1.09	−0.03 (0.07)	> −0.01 [− 0.02, 0.01]	−0.41	0.679
Phase_Order	1.04	0.13 (0.07)	0.01 [< 0.01, 0.01]	1.86	0.066
Driving	1.30	−0.05 (0.08)	> −0.01 [> −0.01, < 0.01]	−0.63	0.529
ASRS_Inattentive	1.59	0.02 (0.09)	> 0.01 [> −0.01, < 0.01]	0.27	0.788
ASRS_Hyperactive	1.66	−0.14 (0.09)	> −0.01 [> −0.01, < 0.01]	−1.57	0.121
Diagnosis	1.30	−0.04 (0.08)	−0.01 [− 0.04, 0.03]	−0.53	0.598
COHS	1.06	−0.14 (0.07)	> −0.01 [> −0.01, < 0.01]	−1.95	0.054
Model 3					
Gender	1.09	−0.03 (0.07)	> −0.01 [− 0.02, 0.01]	−0.42	0.678
Phase_Order	1.04	0.13 (0.07)	0.01 [< 0.01, 0.01]	1.87	0.065
Driving	1.30	−0.05 (0.08)	> −0.01 [> −0.01, < 0.01]	−0.63	0.527
ASRS_Inattentive	1.59	0.02 (0.09)	< 0.01 [> −0.01, < 0.01]	0.27	0.787
ASRS_Hyperactive	1.66	−0.14 (0.09)	> −0.01 [> −0.01, < 0.01]	−1.57	0.119
Diagnosis	1.30	−0.04 (0.08)	−0.01 [− 0.04, 0.03]	−0.53	0.596
COHS	1.06	−0.14 (0.07)	> −0.01 [> −0.01, < 0.01]	−1.96	0.053
Feedback	1	−0.26 (0.06)	−0.03 [− 0.04, −0.01]	−4.07	< 0.001
Model 4					
Gender	1.09	−0.03 (0.07)	> −0.01 [− 0.02, 0.01]	−0.41	0.681
Phase_Order	1.04	0.13 (0.07)	0.01 [< 0.01, 0.01]	1.85	0.068
Driving	1.30	−0.05 (0.08)	> −0.01 [> −0.01, < 0.01]	−0.62	0.531
ASRS_Inattentive	3.11	0.02 (0.09)	> −0.01 [> −0.01, < 0.01]	0.27	0.789
ASRS_Hyperactive	3.23	−0.14 (0.09)	> −0.01 [> −0.01, < 0.01]	−1.56	0.123
Diagnosis	2.35	−0.04 (0.08)	−0.01 [− 0.05, 0.03]	−0.53	0.600
COHS	2.12	−0.14 (0.07)	< 0.01 [> 0.01, < 0.01]	−1.94	0.056
Feedback	65.83	−0.26 (0.06)	0.03 [− 0.09, 0.14]	−4.22	< 0.001
ASRS_Inattentive × Feedback	16.65	0.10 (0.08)	< 0.01 [> −0.01, 0.01]	1.33	0.188
ASRS_Hyperactive × Feedback	13.71	−0.07 (0.08)	> −0.01 [> −0.01, < 0.01]	−0.83	0.410
Diagnosis × Feedback	2.16	< 0.01 (0.06)	< 0.01 [− 0.06, 0.06]	0.02	0.984
COHS × Feedback	57.08	−0.16 (0.06)	> −0.01 [> −0.01, < 0.01]	−2.46	0.016
Model comparisons					
Model	R ²	Log likel.	χ^2	χ^2 sig.	ΔR^2
Model 1	0.02	336.38			
Model 2	0.06	340.66	8.56	0.730	0.04
Model 3	0.13	348.66	16.01	< 0.001	0.07
Model 4	0.16	353.52	9.70	0.046	0.04

Top layer of table depicts all regressors included in the hierarchical model. Model Comparisons layer depicts the predictive strength of each model, as compared to its previous step. VIF, Variance Inflation Factor; SE, Standard Error; CI, Confidence Interval; Log likel., Log likelihood. Significant p-values depicted in bold typeface. Analyses have been outlier corrected, with resulting deviations highlighted in the text. 95% confidence intervals were obtained by bootstrapping 1000 samples in each model.

function of ADHD symptom severity. To this end, we collected ADHD symptom severity metrics from a wide sample of participants in the general population and administered our Go/NoGo task that capitalizes on familiar green-Go/red-NoGo associations. Importantly, our incorporation of a motivational enhancement manipulation (i.e., cumulative performance and monetary feedback) permitted the study of habit expression and disruption. Our results replicate our recent documentation

of familiar Go/NoGo stimuli evoking rigid habitual control, which is also rendered more flexible (i.e., goal-directed) with motivational enhancement (Ceceli et al., 2019). However, we found only modest support for the hypothesis of ADHD symptomology tracking behavioral rigidity and habit disruption. No measure of ADHD significantly predicted outcome-insensitivity as assayed by color-response mapping-related NoGo or Go accuracy impairments. Our exploratory analyses,

however, supported our hypothesis of a significant association between pre-potency of habitual Go actions (i.e., familiar green-Go RT) and hyperactivity presentation. Furthermore, although not directly associated with ADHD, we also found a link between the frequency of habitual tendencies in daily life and habit disruption as indexed by our supplementary measure of outcome-sensitivity: mapping-related Go accuracy impairments. This significant association between daily habit frequency and difficulty breaking well-learned Go associations lends further credence to the idea that the familiar associations we capitalize on are indeed related to well-established, ecologically relevant habits.

A cardinal indicator of habitual control is the performance of an action regardless of the outcome value (Dickinson and Balleine, 1994). Accordingly, we believe that our Go/NoGo task captures outcome-sensitivity, in that the contingency change requires the agent to update which action produces the desired outcome. An impairment in the ability to override the well-learned habit may cause difficulties in flexibly updating the associations between cues and actions (i.e., the color-response mappings) that yield desirable outcomes (e.g., the value of performing a correct action).

We assert that our familiar stimuli elicit outcome-insensitive habits due to their well-established nature. The newly formed associations (e.g., purple-Go) are more labile, allowing the agent to exert goal-directed control regardless of changes to the color-response mappings. By this logic, these novel associations should eventually elicit habitual control with sufficient exposure – similar to overtraining of S–R associations in rodents (Adams, 1982). The magnitude of training necessary for this switch in motivational control using a change in Go and NoGo contingencies remains unknown. Previous research has suggested that pre-training stimuli over the course of an extra training session can yield stronger S–R execution in comparison to new stimulus sets (McKim et al., 2016). Possibly, extensively training the novel associations in our paradigm may also produce habitual control, albeit not with the behavioral rigidity elicited by the familiar associations that have been associated with go and stop actions over the course of development.

In both scientific reports and diagnostic criteria, ADHD is characterized by pronounced deficits in inhibitory control (Wodka et al., 2007; American Psychiatric Association, 2013). When taken together with the reward-related irregularities, we posited that ADHD may also be associated with an impaired motivational control system favoring habits over goal-directed behaviors. Our results do not support this hypothesis with our primary analyses, which could be due to a few key factors.

First, our study recruited participants from the general population and obtained a normal distribution of ADHD-related symptom severity, such that most participants in our sample did not reach the clinical threshold for an ADHD diagnosis. This approach contextualizes any potential ADHD-related impairment in motivational processes to a wider audience, thus expanding the applicability of our research. Consequentially, we are unable to sufficiently represent those who are most debilitated by the symptoms in question: individuals who meet the clinical threshold for ADHD. Any potential ADHD-related

effect may therefore be weakened by the large proportion of individuals who present symptoms below the clinical threshold at magnitudes that do not impair daily functioning. Indeed, a study that recruited adults from the general population to examine ADHD symptomology-related inhibitory control disparities found only a modest association between symptom severity and Go/NoGo task accuracy with 440 participants (Polner et al., 2015). A study with a larger sample size ($n = 1156$) obtained from the general population pinpointed Go/NoGo impairments due to high ADHD-like symptoms, though these effects were sensitive to variations in task structure (e.g., speed and reward structure) (Kuntsi et al., 2009). Taken together with our results, although the ADHD–Go/NoGo impairment association is well-documented in clinical presentations of ADHD, symptom-based approaches may not be sensitive to such effects in the general population. Nonetheless, although there may be disorder-specific factors playing a role in behavioral flexibility that are undetected here, we had reasoned that sampling indiscriminately – that is, without diagnostic cutoffs – could expand the generalizability of potential symptom-related anomalies to the public.

An alternative explanation for the absence of a strong link between motivational control and ADHD symptomology is the notion that individuals with ADHD-like symptoms may also have compensatory mechanisms that promote adaptive behavioral output. For instance, despite the strong evidence of response inhibition deficits in ADHD, attention compensation supported by parietal brain activity has been documented, resulting in comparable Go/NoGo task performance (Ersche et al., 2017). Another possibility is that individuals with ADHD may adopt habitual or goal-directed control in different circumstances. A design that capitalizes on varying task difficulty or cognitive demands may be able to reflect such shifts in habitual and goal-directed processes that are sensitive to individual differences. Brain maturation is another candidate for behavioral similarities in ADHD and NT populations. ADHD is associated with a delayed maturation of the prefrontal cortex (Shaw et al., 2007), a region that is critical for error detection, reversal learning, and conflict monitoring. These processes are crucial for optimal Go/NoGo task performance (Garavan et al., 2002; Zhang et al., 2016), especially one involving changes to color-response mappings. Accordingly, adults with ADHD may produce signs of intact Go/NoGo performance due to the maturations in prefrontal regions, compensating for potential impairments that may have been evident with a less mature cortex (Carmona et al., 2012). Another potential compensatory mechanism may be driven by ADHD medications that act on the brain's dopaminergic systems. We did not ascertain whether our participants – with or without ADHD – were taking ADHD medication. Methylphenidate, for instance, has been reported to enhance executive function in individuals with ADHD, as well as in NTs (Schweitzer et al., 2004; Linssen et al., 2014; Moeller et al., 2014). These beneficial effects of ADHD medication on executive function have also been shown to extend beyond methylphenidate (Hosenbocus and Chahal, 2012). Our sample of adults with varying degrees of ADHD-related symptoms may be recruiting similar compensatory mechanisms that aid in maintaining goal-directed control. Future research that captures

developmental and pharmacological aspects of ADHD and goal-directed control may elucidate which of these mechanisms plays a critical role in adaptive motivational control.

We reasoned that because hyperactive ADHD presentation is associated with the number of impulsivity-related items endorsed on the ASRS (Kessler et al., 2005a), participants exhibiting high hyperactivity may execute quicker, impulsive Go actions. Our green-Go RT data supported our hypothesis, in that hyperactivity scores correlated with quicker responses to the well-learned habit eliciting stimulus. It should be noted that this finding was the result of an exploratory analysis. Nonetheless, our finding of a significant response latency and hyperactivity association bridges the fields of motivation and ADHD. Impulsivity, a core element of the hyperactive presentation of ADHD, is also associated with reflexive behaviors to cues and heightened variability in response latency (Kirkeby and Robinson, 2005). The heightened pre-potency to respond to habitual cues tracked by our hyperactivity scale may suggest an overlap in the motivational and inhibitory mechanisms underlying hyperactivity in ADHD, potentially explaining the lapses in behavioral output that result in higher RT and accuracy variability (Kirkeby and Robinson, 2005; Tamm et al., 2012). In other words, if hyperactivity predicts quicker responses to well-learned stimuli and high RT variability, this effect may be due to motivational and motor processes that are activated depending on past experience with the cue at hand. Future research will be imperative in effectively dissociating the motivational, attentional, and inhibitory processes that underlie response latency variability in ADHD.

In addition to the analyses reported here, an alternative method of exploring Go and NoGo performance is via signal detection. In a typical signal detection analysis, hits, misses, false alarms, and correct rejection values are used to derive d' – an estimate of response bias (Stanislaw and Todorov, 1999). Importantly, in each run of our task, a color-response mapping (e.g., green-Go) would only provide two of the four values that comprise a d' score (i.e., hits and misses, but not false alarms or correct rejections for this color). The remaining parameters would need to be extracted from the “incongruent” run (green-NoGo), which would make it difficult to obtain accurate response bias information. However, one can indeed investigate signal detection based on familiarity and congruency of the color-response mappings (e.g., where green-Go and red-NoGo together are coded as d' _familiar_congruent, and green-NoGo, red-Go are together coded as d' _familiar_incongruent). When performed as such, response bias results mirror our NoGo and Go accuracy findings reported here, in that (1) participants show high response bias when the color-response mappings are familiar and congruent with daily experiences, (2) response bias is significantly lower when familiar stimuli are mapped onto incongruent responses, (3) the two novel color-response mappings are similar in the elicited response bias, and (4) response bias does not show significant associations with ADHD-related individual difference measures.

Limitations

We acknowledge several limitations in the present study that should be considered in future investigations.

Although we were able to generalize our findings to a wider audience by recruiting without diagnostic cutoffs, we did not survey participants for history of psychiatric illnesses or psychoactive medication use. Several psychiatric conditions have been documented to affect motivational control (Griffiths et al., 2014). Furthermore, ADHD medications have been shown to improve executive function (Hosenbocus and Chahal, 2012; Linssen et al., 2014; Moeller et al., 2014), which may be related to the expansion of cognitive resources necessary to maintain goal-directed control. An interesting avenue to explore in future ADHD research may be the roles of psychiatric comorbidities and treatment history in the expression of habits.

Our study's primary hypotheses regarding habitual control and ADHD symptomology were motivated by reports of reward circuitry dysfunction in ADHD (Ceceli et al., 2019). However, we did not collect neural data that may speak to the potential links between ADHD symptomology and habitual control as mediated by neural function. The brain systems of reward processing and learning are outside the scope of our study, but the mechanisms underlying motivational control as related to ADHD symptoms may be effectively elucidated by a neurobiological approach. Future research that examines the potential disparities in the ADHD brain related to motivational control may advance our understanding of the disorder's pathophysiology.

We adopted a within-subject design to tackle the expression and disruption of habits over the course of two sessions. This design permitted us to compare habit expression and disruption at an individual basis while improving statistical power. However, it can be argued that administering a task twice to the same set of participants may introduce training effects. Our second session data suggest that participants did not significantly improve their performance in the face of congruent associations by merely undergoing the task in the previous session. However, the definitive method to circumvent potential training effects would be to apply a between-subjects design, in which separate sets of participants undergo the feedback and no-feedback sessions. We report in another study that adopts a between-subject design a similar pattern of results – motivational enhancement indeed disrupts the expression of well-learned habits (Ceceli et al., 2019).

CONCLUSION

Attention deficit-hyperactivity disorder is a heterogeneous psychiatric condition with debilitating consequences to behavior, neural processing, and well-being. In this study, we aimed to reveal the potential irregularities in managing well-learned habits by sampling symptom severity information from the general population. Although we did not find a strong association between motivational control deficits and ADHD-related symptoms, our data replicate a previous report of well-learned habit expression and disruption, and allude to a link between hyperactivity and pre-potency to respond to well-learned Go

stimuli. Taken together with previous reports of compensatory mechanisms aiding in Go/NoGo task performance in ADHD, delay in cortical maturation in ADHD yielding differential inhibitory processes across children and adults, and our sample largely comprising subclinical ADHD presentations, a full understanding of the potential link between ADHD and motivational control may require a neurobehavioral and developmental approach.

DATA AVAILABILITY

The datasets and scripts associated with this study can be found in the **Supplementary Materials**.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Rutgers University Institutional Review Board with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Rutgers University Institutional Review Board.

REFERENCES

- Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q. J. Exp. Psychol. Sect. B* 34, 77–98. doi: 10.1080/14640748208400878
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edn. Washington, DC: American Psychiatric Association.
- Antrop, I., Stock, P., Verté, S., Wiersema, J. R., Baeyens, D., and Roeyers, H. (2006). ADHD and delay aversion: the influence of non-temporal stimulation on choice for delayed rewards. *J. Child Psychol. Psychiatry* 47, 1152–1158. doi: 10.1111/j.1469-7610.2006.01619.x
- Balleine, B. W., and O'Doherty, J. P. (2009). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35, 48–69. doi: 10.1038/npp.2009.131
- Carmona, S., Hoekzema, E., Ramos-Quiroga, J. A., Richarte, V., Canals, C., Bosch, R., et al. (2012). Response inhibition and reward anticipation in medication-naïve adults with attention-deficit/hyperactivity disorder: a within-subject case-control neuroimaging study. *Hum. Brain Mapp.* 33, 2350–2361. doi: 10.1002/hbm.21368
- Castellanos, F. X., and Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nat. Rev. Neurosci.* 3, 617–628. doi: 10.1038/nrn896
- Ceceli, A. O., Myers, C. E., and Tricomi, E. (2019). Demonstrating and disrupting well-learned habits. *bioRxiv* [Preprint]. doi: 10.1101/745356
- Ceceli, A. O., and Tricomi, E. (2018). Habits and goals: a motivational perspective on action control. *Curr. Opin. Behav. Sci.* 20, 110–116. doi: 10.1016/j.cobeha.2017.12.005
- Costa Dias, T. G., Wilson, V. B., Bathula, D. R., Iyer, S. P., Mills, K. L., Thurlow, B. L., et al. (2013). Reward circuit connectivity relates to delay discounting in children with attention-deficit/hyperactivity disorder. *Eur. Neuropsychopharmacol.* 23, 33–45. doi: 10.1016/j.euroneuro.2012.10.015
- Dickinson, A., and Balleine, B. (1994). Motivational control of goal-directed action. *Anim. Learn. Behav.* 22, 1–18. doi: 10.3758/BF03199951

AUTHOR CONTRIBUTIONS

AC and ET designed the experiments. AC and GE coded the experimental paradigm and collected the data. All authors contributed to the data analysis and manuscript preparation.

FUNDING

This work was supported by a grant from the National Science Foundation (BCS1150708) awarded to ET.

ACKNOWLEDGMENTS

We would like to express our gratitude to the Learning and Decision Making Lab of Rutgers University-Newark for their helpful comments, and Miriam Rosenberg-Lee for guidance during task-design.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01997/full#supplementary-material>

- Dillo, W., Göke, A., Prox-Vagedes, V., Szycik, G. R., Roy, M., Donnerstag, F., et al. (2010). Neuronal correlates of ADHD in adults with evidence for compensation strategies – a functional MRI study with a Go/No-Go paradigm. *Ger. Med. Sci.* 8:Doc09. doi: 10.3205/000098
- Ersche, K. D., Lim, T.-V., Ward, L. H. E., Robbins, T. W., and Stoehl, J. (2017). Creature of habit: a self-report measure of habitual routines and automatic tendencies in everyday life. *Pers. Individ. Differ.* 116, 73–85. doi: 10.1016/j.paid.2017.04.024
- Frodl, T., and Skokauskas, N. (2012). Meta-analysis of structural MRI studies in children and adults with attention deficit hyperactivity disorder indicates treatment effects. *Acta Psychiatr. Scand.* 125, 114–126. doi: 10.1111/j.1600-0447.2011.01786.x
- Furukawa, E., Bado, P., Tripp, G., Mattos, P., Wickens, J. R., Bramati, I. E., et al. (2014). Abnormal striatal BOLD responses to reward anticipation and reward delivery in ADHD. *PLoS One* 9:e89129. doi: 10.1371/journal.pone.0089129
- Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., and Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. *Neuroimage* 17, 1820–1829. doi: 10.1006/nimg.2002.1326
- Griffiths, K. R., Morris, R. W., and Balleine, B. W. (2014). Translational studies of goal-directed action as a framework for classifying deficits across psychiatric disorders. *Front. Syst. Neurosci.* 8:101. doi: 10.3389/fnsys.2014.00101
- Haber, S. N. (2003). The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat.* 26, 317–330. doi: 10.1016/j.jchemneu.2003.10.003
- Hosenbocus, S., and Chahal, R. (2012). A review of executive function deficits and pharmacological management in children and adolescents. *J. Can. Acad. Child Adolesc. Psychiatry* 21, 223–229.
- Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., et al. (2005a). The World Health Organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. *Psychol. Med.* 35, 245–256. doi: 10.1017/S0033291704002892
- Kessler, R. C., Adler, L. A., Barkley, R. A., Biederman, J., Conners, C. K., Faraone, S. V., et al. (2005b). Patterns and predictors of attention-deficit/hyperactivity disorder persistence into adulthood: results from the national comorbidity

- survey replication. *Biol. Psychiatry* 57, 1442–1451. doi: 10.1016/j.biopsych.2005.04.001
- Kirkeby, B. S., and Robinson, M. D. (2005). Impulsive behavior and stimulus–response variability in choice reaction time. *J. Res. Personal.* 39, 263–277. doi: 10.1016/j.jrp.2004.04.001
- Kluger, A. N., and DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol. Bull.* 119, 254–284. doi: 10.1037/0033-2909.119.2.254
- Kuntsi, J., Wood, A. C., Meere, J. V. D., and Asherson, P. (2009). Why cognitive performance in ADHD may not reveal true potential: findings from a large population-based sample. *J. Int. Neuropsychol. Soc.* 15, 570–579. doi: 10.1017/S135561770909081X
- Linssen, A. M. W., Sambeth, A., Vuurman, E. F. P. M., and Riedel, W. J. (2014). Cognitive effects of methylphenidate in healthy volunteers: a review of single dose studies. *Int. J. Neuropsychopharmacol.* 17, 961–977. doi: 10.1017/S1461145713001594
- Marx, I., Höpcke, C., Berger, C., Wandschneider, R., and Herpertz, S. C. (2013). The impact of financial reward contingencies on cognitive function profiles in adult ADHD. *PLoS One* 8:e67002. doi: 10.1371/journal.pone.0067002
- McKim, T. H., Bauer, D. J., and Boettiger, C. A. (2016). Addiction history associates with the propensity to form habits. *J. Cogn. Neurosci.* 28, 1024–1038. doi: 10.1162/jocn_a_00953
- Meule, A. (2017). Reporting and interpreting task performance in go/no-go affective shifting tasks. *Front. Psychol.* 8:701. doi: 10.3389/fpsyg.2017.00701
- Moeller, S. J., Honorio, J., Tomasi, D., Parvaz, M. A., Woicik, P. A., Volkow, N. D., et al. (2014). Methylphenidate enhances executive function and optimizes prefrontal function in both health and cocaine addiction. *Cereb. Cortex* 24, 643–653. doi: 10.1093/cercor/bhs345
- Montague, W. E., and Webber, C. E. (1965). Effects of knowledge of results and differential monetary reward on six uninterrupted hours of monitoring. *Hum. Factors* 7, 173–180. doi: 10.1177/00187208650070209
- Natsheh, J. Y., and Shiflett, M. W. (2015). The effects of methylphenidate on goal-directed behavior in a rat model of ADHD. *Front. Behav. Neurosci.* 9:326. doi: 10.3389/fnbeh.2015.00326
- Natsheh, J. Y., and Shiflett, M. W. (2018). Dopaminergic modulation of goal-directed behavior in a rodent model of attention-deficit/hyperactivity disorder. *Front. Integr. Neurosci.* 12:45. doi: 10.3389/fnint.2018.00045
- Norman, L. J., Carlisi, C., Lukito, S., Hart, H., Mataix-Cols, D., Radua, J., et al. (2016). Structural and functional brain abnormalities in attention-deficit/hyperactivity disorder and obsessive-compulsive disorder: a comparative meta-analysis. *JAMA Psychiatry* 73, 815–825. doi: 10.1001/jamapsychiatry.2016.0700
- O'Doherty, J. P. (2016). Multiple systems for the motivational control of behavior and associated neural substrates in humans. *Curr. Top. Behav. Neurosci.* 27, 291–312. doi: 10.1007/7854_2015_386
- O'Doherty, J. P., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452–454. doi: 10.1126/science.1094285
- Plichta, M. M., and Scheres, A. (2014). Ventral–striatal responsiveness during reward anticipation in ADHD and its relation to trait impulsivity in the healthy population: a meta-analytic review of the fMRI literature. *Neurosci. Biobehav. Rev.* 38, 125–134. doi: 10.1016/j.neubiorev.2013.07.012
- Polner, B., Aichert, D., Macare, C., Costa, A., and Ettinger, U. (2015). Gently restless: association of ADHD-like traits with response inhibition and interference control. *Eur. Arch. Psychiatry Clin. Neurosci.* 265, 689–699. doi: 10.1007/s00406-014-0531-7
- Qiu, A., Crocetti, D., Adler, M., Mahone, E. M., Denckla, M. B., Miller, M. I., et al. (2009). Basal ganglia volume and shape in children with attention deficit hyperactivity disorder. *Am. J. Psychiatry* 166, 74–82. doi: 10.1176/appi.ajp.2008.08030426
- Rosch, K. S., Mostofsky, S. H., and Nebel, M. B. (2018). ADHD-related sex differences in fronto-subcortical intrinsic functional connectivity and associations with delay discounting. *J. Neurodev. Disord.* 10:34. doi: 10.1186/s11689-018-9254-9
- Schulz, K. P., Fan, J., Magidina, O., Marks, D. J., Hahn, B., and Halperin, J. M. (2007). Does the emotional go/no-go task really measure behavioral inhibition? convergence with measures on a non-emotional analog. *Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol.* 22, 151–160. doi: 10.1016/j.acn.2006.12.001
- Schweitzer, J. B., Lee, D. O., Hanford, R. B., Zink, C. F., Ely, T. D., Tagamets, M. A., et al. (2004). Effect of methylphenidate on executive functioning in adults with attention-deficit/hyperactivity disorder: normalization of behavior but not related brain activity. *Biol. Psychiatry* 56, 597–606. doi: 10.1016/j.biopsych.2004.07.011
- Shaw, P., Eckstrand, K., Sharp, W., Blumenthal, J., Lerch, J. P., Greenstein, D., et al. (2007). Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19649–19654. doi: 10.1073/pnas.0707741104
- Sonuga-Barke, E. J., Taylor, E., Sembi, S., and Smith, J. (1992). Hyperactivity and delay aversion—I. The effect of delay on choice. *J. Child Psychol. Psychiatry* 33, 387–398. doi: 10.1111/j.1469-7610.1992.tb00874.x
- Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* 31, 137–149. doi: 10.3758/BF03207704
- Ströhle, A., Stoy, M., Wrase, J., Schwarzer, S., Schlagenhauf, F., Huss, M., et al. (2008). Reward anticipation and outcomes in adult males with attention-deficit/hyperactivity disorder. *Neuroimage* 39, 966–972. doi: 10.1016/j.neuroimage.2007.09.044
- Tabachnick, B. G., and Fidell, L. S. (2007). *Using Multivariate Statistics*, 5th Edn. Boston, MA: Allyn & Bacon.
- Tamm, L., Narad, M. E., Antonini, T. N., O'Brien, K. M., Hawk, L. W., and Epstein, J. N. (2012). Reaction time variability in ADHD: a review. *Neurotherapeutics* 9, 500–508. doi: 10.1007/s13311-012-0138-5
- Tricomi, E., Balleine, B. W., and O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29, 2225–2232. doi: 10.1111/j.1460-9568.2009.06796.x
- Umemoto, A., and Holroyd, C. B. (2015). Task-specific effects of reward on task switching. *Psychol. Res.* 79, 698–707. doi: 10.1007/s00426-014-0595-z
- von Rhein, D., Beckmann, C. F., Franke, B., Oosterlaan, J., Heslenfeld, D. J., Hoekstra, P. J., et al. (2017). Network-level assessment of reward-related activation in patients with ADHD and healthy individuals. *Hum. Brain Mapp.* 38, 2359–2369. doi: 10.1002/hbm.23522
- von Rhein, D., Cools, R., Zwiers, M. P., van der Schaaf, M., Franke, B., Luman, M., et al. (2015). Increased neural responses to reward in adolescents and young adults with attention-deficit/hyperactivity disorder and their unaffected siblings. *J. Am. Acad. Child Adolesc. Psychiatry* 54, 394–402. doi: 10.1016/j.jaac.2015.02.012
- Wilbertz, G., Tebartz van Elst, L., Delgado, M. R., Maier, S., Feige, B., Philipsen, A., et al. (2012). Orbitofrontal reward sensitivity and impulsivity in adult attention deficit hyperactivity disorder. *Neuroimage* 60, 353–361. doi: 10.1016/j.neuroimage.2011.12.011
- Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., and Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biol. Psychiatry* 57, 1336–1346. doi: 10.1016/j.biopsych.2005.02.006
- Wodka, E. L., Mahone, E. M., Blankner, J. G., Larson, J. C. G., Fotedar, S., Denckla, M. B., et al. (2007). Evidence that response inhibition is a primary deficit in ADHD. *J. Clin. Exp. Neuropsychol.* 29, 345–356. doi: 10.1080/13803390600678046
- Wodushak, T. R., and Neumann, C. S. (2003). Inhibitory capacity in adults with symptoms of attention deficit/hyperactivity disorder (ADHD). *Arch. Clin. Neuropsychol.* 18, 317–330. doi: 10.1016/S0887-6177(02)00152-X

- Yee, D. M., Krug, M. K., Allen, A. Z., and Braver, T. S. (2016). Humans integrate monetary and liquid incentives to motivate cognitive task performance. *Front. Psychol.* 6:2037. doi: 10.3389/fpsyg.2015.02037
- Young, M. E., Sutherland, S. C., and McCoy, A. W. (2018). Optimal go/no-go ratios to maximize false alarms. *Behav. Res. Methods* 50, 1020–1029. doi: 10.3758/s13428-017-0923-5
- Zhang, Z., Mendelsohn, A., Manson, K. F., Schiller, D., and Levy, I. (2016). Dissociating value representation and inhibition of inappropriate affective response during reversal learning in the ventromedial prefrontal cortex. *eNeuro* 2:ENEURO.72–15.2015. doi: 10.1523/ENEURO.0072-15.2015

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ceceli, Esposito and Tricomi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Binary Theorizing Does Not Account for Action Control

Bernhard Hommel*

Cognitive Psychology Unit, Institute of Psychology, Leiden University, Leiden, Netherlands

OPEN ACCESS

Edited by:

John A. Bargh,
Yale University,
United States

Reviewed by:

Bas Verplanken,
University of Bath,
United Kingdom
Maurizio Tirassa,
University of Turin, Italy

*Correspondence:

Bernhard Hommel
hommel@fsw.leidenuniv.nl

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 03 May 2019

Accepted: 28 October 2019

Published: 14 November 2019

Citation:

Hommel B (2019) Binary Theorizing
Does Not Account for Action Control.
Front. Psychol. 10:2542.
doi: 10.3389/fpsyg.2019.02542

Everyday thinking and scientific theorizing about human action control are equally driven by the apparently obvious contrast between will and habit or, in their more modern disguise: intentional and automatic processes, and model-based and model-free action planning. And yet, no comprehensive category system to systematically tell truly willed from merely habitual actions is available. As I argue, this is because the contrast is ill-conceived, because almost every single action is both willed and habitual, intentional and automatic, and model-based and model-free, simply because will and habit (and their successors) do not refer to alternative modes or pathways of action control but rather to different phases of action planning. I further discuss three basic misconceptions about action control that binary theorizing relies on: the assumption that intentional processes compete with automatic processes (rather than the former setting the stage for the latter), the assumption that action control is targeting processes (rather than representations of action outcomes), and the assumption that people follow only one goal at a time (rather than multiple goals). I conclude that (at least the present style of) binary theorizing fails to account for action control and should thus be replaced by a more integrative view.

Keywords: action control, dual-route models, goal, automaticity and control, intention

BINARY THEORIZING ON ACTION CONTROL

Will vs. Habit

The study of action control was driven by binary theorizing right from the start. In his first systematic analysis of the human will, Ach (1910) postulated that will can be best studied by analyzing the degree to, and the conditions under which, it can overcome what Ach considered its natural opponent: acquired habits. To achieve that, he developed what he called the *combined method* (“kombiniertes Verfahren”), which first established a particular habit, defined as a set of stimulus-response associations reflecting a particular stimulus-response rule, and then changed the instruction in such a way that participants were now to respond differently to the previously acquired stimulus set (see Hommel, 2000a). For instance, participants may first learn to read through lists of nonsense syllables that were followed by a rhyme (e.g., zup → tup, tel → mel) over an extended time period and then respond to the same stimulus syllables by changing the order of the letters (e.g., zup → puz, tel → let). As predicted, participants were slower and produced more errors when applying the new instruction to a stimulus set that was previously related to different responses than when working with a new set. The idea was that being exposed to lists constructed according to the first rule created stimulus-response habits that would need to be overcome in order to successfully apply the second rule. Accordingly, the degree to which

participants were able to overcome the previously acquired habit (i.e., the difference in performance on old versus new sets) was taken to measure “willpower,” which was shown to differ between individuals (which was taken to diagnose individual willpower) and to vary systematically with the practice given on old sets (which was taken to increase the strength of the habit).

It is easy to see that this pioneering approach has survived until today, even though researchers less frequently take the effort to induce habits experimentally anymore: they often exploit existing habits, such as the tendency to read words even if one is to name their color, as in the notorious Stroop task (Stroop, 1935; even though Stroop himself did analyze the impact of experimental training). Like in Ach's studies, the degree to which performance is impaired with stimulus sets that are assumed to activate the hypothetical habit (such as words denoting response-incompatible colors in the Stroop task) as compared to suitable control sets (such as nonwords, non-color words, or words denoting response-compatible colors) is taken to reflect the strength or weakness of willpower, which meanwhile has been relabeled as “cognitive control” or “executive function”—presumably in an attempt to get rid of the phenomenological connotations of the will concept (Goschke, 2003).

Working with binary oppositions such as will and habit has been taken to reflect human nature (Newell, 1973; Melnikoff and Bargh, 2018), and so it comes as no surprise that the will/habit couple has survived in various disguises until today. Its long-standing history tends to be systematically underestimated by available reviews, which for instance have dated back its introduction into theorizing about action control to the work of Tolman (1948, see Dolan and Dayan, 2013), Atkinson and Shiffrin (1968, see Monsell and Driver, 2000), or Dickinson (1985, see Gillan et al., 2015)—thus rather generously neglecting the pioneering study on the phenomenology of will by Michotte and Prüm (1911); the first systematic experimental program on studying will and habit by Ach (1910, 1935), which spanned no less than 30 years; the first approach questioning the goal-independence of habits by Lewin (1922a,b); and the other 200 or so studies on action control summarized in Ach (1935) already.

The basic thought underlying the opposition between will and habit is that some responses are so strongly associated with particular stimuli that encountering the stimulus is sufficient to activate the response. This holds for rhyming in Ach's studies—seeing a nonsense syllables triggers the overlearned rhyming response, reading in the Stroop task—seeing the word is sufficient to trigger some reading tendency, and performing a left or right response in the Simon task (Simon, 1969)—processing a left or right stimulus triggers a spatially corresponding action (Kornblum et al., 1990). The basic setup of all tasks investigating the interplay between will and habit puts the two against each other, just as recommended by Ach (1910), by instructing individuals to carry out a relatively uncommon or counterintuitive action B to a particular stimulus that is assumed to be strongly associated with another action A. If then any experimental evidence can be found that action A was activated to at least some measurable degree, the participant is thought to have experienced an action-control problem that was due to the fact that practice established an

association between A and the stimulus, so that encountering the stimulus would activate action A even under circumstances where A is not appropriate and not wanted.

Very soon after Ach's claims that stimulus-response associations can challenge and may even outcompete the processes controlled by the actual goal, Lewin (1922a,b) reported findings calling for a more moderate view. On the one hand it was possible to counteract an intense intention with a habit that relied on few, sometimes just one repetition but, on the other, 300 repetitions were insufficient to have any impact. According to Lewin (1928), the key to understand the impact of habits has to do with their specific role in the current action plan. On the one hand, habitual actions do not represent real alternatives to intentional actions, in the sense that people would face difficulties to decide whether they should name the color of a Stroop word or read it. Lewin suggests that the intention to open a door that requires pushing the handle up, rather than down, will not be hindered by the thousands or so previous repetitions of opening doors by pushing the handle down. On the other hand, however, habitual actions do have the potency to interfere if they are embedded into a larger action context, such as if one is to open the door on one's way to get a glass of water from the other room.

The same principle seems to apply to the Stroop effect, which is very pronounced (often >100 ms effect size) if the response set consists of spoken color words (i.e., the responses that reading the words would produce) but often dramatically shrinks or disappears with keypressing responses (e.g., McClain, 1983)—and even the effects that keypressing responses sometimes do produce seem to be artifacts due to task-irrelevant but spontaneously occurring internal naming strategies (Martin, 1978; Mascolo and Hirtle, 1990). In other words, the Stroop effect is likely to depend on introducing an obvious contradiction by requiring participants to attend to, and actually generate color words and at the same time nominally declaring color words task-irrelevant and to-be-ignored. Another obvious contradiction results from the fact that, in the standard Stroop task (as well as in other tasks following the same rationale), violating the instruction by reading the word actually pays off in 50% of the trials. This means that, on average, participants are rewarded for unintentionally or intentionally reading the word, especially given that word-reading is faster and requires less effort—just because of the more elaborate practice. That this is an important ingredient of the task is obvious from the finding that the size of the effect varies systematically with the percentage of the payoff: it becomes stronger if payoff increases and weaker if it decreases (e.g., Logan and Zbrodoff, 1979). This suggests that the impact of habitual action tendencies is anything but non-intentional, and clearly very sensitive to the expected outcome—a theme I will get back to below.

Controlled vs. Automatic

As pointed out by Goschke (2003), theories on action control have seen a rather dramatic conceptual overhaul since the early days of Michotte, Ach, and Lewin. While the pioneering approaches were still strongly connected to the phenomenology of willing and acting, understanding which was even an explicit theoretical

aim of Michotte and Ach, later theorizing preferred a less “subjective” terminology that was inspired by the increasingly popular computer metaphor for the description and analysis of human cognition in the 1950s and 1960s (Broadbent, 1958; Neisser, 1967). This terminological preference favored less colorful concepts like “controlled” versus “automatic” processing over the old-fashioned terms will and habit. Even though the basic idea was the same, the explanations changed in flavor: whereas the older approaches tried to explain the strong impact of habits by referring to an assumed cause—the strong stimulus-response association driving the habitual action, the new generation of processing theories tended to emphasize different degrees of speed and efficiency of the underlying processes (even though some studies still tested the practice-dependency of automaticity directly: e.g., Schneider et al., 1984; Smith and Lerner, 1986; MacLeod and Dunbar, 1988). For instance, the observation that responses are easier to perform in response to particular stimuli than others (e.g., left rather than write keypresses to stimuli appearing on the left) was explained by postulating the existence of a particular “population stereotype” (Fitts, 1951). At the surface, accounts of this sort do not seem to go beyond redescribing the actual finding in theoretically sounding terms, but they often implicitly rely on associationist logic: one way or another, such shared stereotypes must emerge from shared practices and training, which implies that stereotype is just another word for an associative structure linking stimuli to particular responses.

In other approaches the correspondence between controlled versus automatic processes on the one hand and will versus habit on the other is even more opaque. For instance, in their comprehensive model of stimulus-response compatibility, Kornblum et al. (1990) attribute the impact of what previously counted as habit to automaticity. It is automaticity that does the major trick in the explanation of why irrelevant stimuli seem to be able to trigger responses that conflict with the actually intended action, like in a Stroop task. Where automaticity comes from is a topic that the authors explicitly neglect: they briefly consider the possibility that training plays a role but then choose “not to make practice a major focus or concern” in their model (p. 263). Again, this renders the major theoretical contribution to the question of *why* irrelevant stimuli can trigger conflicting responses a mere reformulation of the empirical observation in theoretical terms¹.

These and other theoretical developments indicate that the systematic replacement of the will/habit concept by the controlled/automatic concept has tempted at least more cognitively oriented theorists as cited above² to refocus the theoretical attention

away from the possible causes of the impact of the relevant information on action control to the consequences—away from the possible role of overlearning to the resulting automaticity. As a consequence, in these approaches automaticity was no longer defined with respect to its origin, such as the amount of training necessary to achieve it, but with respect to its opponent: the intention or control process. Note that this is a dangerous theoretical twist. The explananda targeted by control/automaticity theories derive from empirical observations that some behavior or some aspects of behavior do not fully comply with the instructions given to the investigated participants: they tend to read words rather than naming their color and press the key that spatially corresponds to the stimulus even if they should do the opposite. A certain lack of control is thus inherent in these observations, which renders the attempt to explain the observations by referring to automaticity circular: if automaticity is only defined by the absence of control, and if control is defined by compliance with the experimental instruction, the observed behavior must be automatic. In other words, automaticity cannot be an explanation because it is an integral component of the description of the to-be-explained phenomenon—automaticity is an explanandum, not an explanans!

These terminological confusions aside, it is fair to say that true automaticity has yet to be demonstrated. Kornblum et al. (1990) suggest applying the definition of Kahneman and Treisman (1984, p. 43), according to whom a strongly automatic process is one that is “neither facilitated by focusing attention on [its object] nor impaired by diverting attention from [it],” whereas “a partially automatic process is one that is normally triggered without attention directed at its object but is facilitated by having attention focused on it” (Kornblum et al., 1990, p. 261). “According to this view,” Kornblum et al. (1990) continue, “an automatic process could under some conditions be attenuated or enhanced. However, under no conditions could it be ignored or bypassed.” I have already mentioned evidence suggesting that even the Stroop effect, thought to be one of the milestones of demonstrating true automaticity, can disappear by simply changing the response set. However, such evidence might be discounted by considering a role of attention, which might be drastically reduced by this change and thus make the automaticity only partial. Moreover, Kornblum et al. claim true automaticity only for feature-overlap between stimuli and responses, which arguably is reduced, in some sense even eliminated by changing the response set in a Stroop task. However, automaticity can be shown to not exist even without changing the responses.

For instance, Valle-Inclán and Redondo (1998) presented participants with a Simon task, in which they responded to red and green colored circles by pressing the left and right response keys, respectively. In one condition, participants received the stimulus-response mapping first and were then presented with the lateralized color circle. Electrophysiological recordings showed that the presentation of the stimulus led to an increased activation in the cortical hemisphere opposite to its location—a classical lateralized readiness potential that is thought to represent response activation of the contralateral response hand (Eimer, 1995). This potential was even seen if the actual response

¹According to Lewin (1931), the idea that categorizing a particular phenomenon is sufficient to explain it is a reflection of what he called Aristotelian psychology (a theoretical attitude that is very typical for stage approaches to studying human information processing: Hommel, in press), which he contrasts with Galilean psychology that seeks to unravel the actual functional mechanisms.

²This is not to say that attempts to systematically control the degree of automaticity acquired through experimental practice no longer exist. The learning-theoretical tradition to make training/exercise part of the experimental design has survived especially in the cognitive neurosciences (e.g., Schwabe and Wolf, 2009) and applied areas related to lifestyle issues and addiction (e.g., Watson et al., 2014; Lin et al., 2016).

required movement of the other hand, suggesting that it indicated the potency of the stimulus to automatically activate the spatially corresponding response hand. In another condition, the stimulus appeared first, and only thereafter the stimulus-response mapping was presented. If, according to the definition of Kahneman and Treisman and Kornblum et al., the association between stimulus location and response would be strongly automatic, the presentation of the stimulus should generate the same electrophysiological response as in the other condition. If the association would be partially automatic, the stimulus might show a reduced electrophysiological response. However, the findings showed no response whatsoever. If anything, this suggests that implementing the instruction is a precondition for automatic responses to occur, which means that they are neither fully nor partially automatic (cf., Trafimov, 2018) but what Bargh (1989) has called conditionally automatic.

A key problem with dealing with the concept of automaticity is that it remains a moving target in the literature. For instance, some authors (like Kahneman and Treisman, 1984) speak of automatic processes while others speak of automatic actions (e.g., Wheatley and Wegner, 2001). Some authors have argued that automatic processes need to meet all criteria for automaticity to deserve this label (what Moors and de Houwer, 2006, call the “all-or-none view”; e.g., Johnson and Hasher, 1987), while others were more liberal, allowing for various combinations of some of the criteria (e.g., Bargh, 1994; Moors and de Houwer, 2006), and the fact that the discussed criteria themselves vary extensively from author to author (see Melnikoff and Bargh, 2018) did not help to find a broad consensus either. For instance, while Kahneman and Treisman considered a process automatic if it is “neither facilitated by focusing attention on [its object] nor impaired by diverting attention from [it],” Bargh (1994) suggested a combination of a lack of awareness and intentionality, high efficiency, and a lack of motivation (a criterion that appeals to the desire criterion that I will criticize below), and Moors and de Houwer (2006) extend this list to eight criteria, according to which automaticity might refer to processes that are unintentional, uncontrollable, goal independent, autonomous, purely stimulus driven, unconscious, efficient, and fast.

I will not provide point-to-point reviews of these criteria but do like to set the stage for the following discussion by means of two comments: first, the sheer number and variability of suggested criteria for sorting processes into automatic versus intentional ones, together with the fact that authors increasingly give up the idea that automaticity criteria might converge onto any coherent category (Bargh, 1994; Moors and de Houwer, 2006; Melnikoff and Bargh, 2018), undermine the original idea that cognitive processes can be categorized into two non-overlapping categories. Second, the criteria that have been suggested so far undoubtedly relate to measurable features of processes but there are reasons to doubt whether they even speak to the question of willed vs. non-willed behavior. As I will elaborate below, this is because: (1) goals and intentions control *outcomes* of behavior but not the *processes* producing it, which renders the connection between action control and criteria like controllability or autonomy questionable;

(2) selecting an action emerges from the *goal-driven but fully automatic competition* between automatically executed action tendencies, which undermines the very idea that processes might be non-automatic in principle; and (3) the selection value that processes bring to this competition may well refer to the efficiency and speed of the action that this process represents, suggesting that the relevance of these criteria in action selection should be considered a sign of intentionality rather than the opposite.

Model-Based vs. Model-Free

The most recent version of will/habit thinking comes in the disguise of models contrasting model-based and model-free systems. This contrast refers to two kinds of modeling reinforcement learning (e.g., Sutton and Barto, 2017): model-based learning is assumed to rely on a state-transition model, which accumulates knowledge about the current state, the possible actions this state allows, and the state that would follow when taking this action, and a reward model that connects end-states with particular rewards. Hence, this kind of learning is based on a kind of model of the environment, which allows forward-planning and reward-maximization even when the environment changes. Model-free learning, in contrast, does not consider sequential dependencies like state-action-outcome relationships or rewards but relies on stored selection values for all previously experienced state-action contingencies.

It is fair to say that there is no coherent theory integrating the available thoughts about how these systems work and how they interact, and it is also fair to say that quite a bit of confusion exists regarding what the terms model-based and model-free imply. One idea is that the goal-related model-based system stores contingencies between actions and outcomes while the automatic, model-free system stores stimulus-response associations (Dickinson and Balleine, 1994). According to this idea, model-based action implies consideration of the expected outcome whereas model-free action is driven by some contextual cue—a metamorphosis of the traditional habit. Others have criticized this conceptual opposition. For instance, Miller et al. (2019) have argued that the original idea assumes that habits are outcome-blind (“value-free”), whereas modern reinterpretations (e.g., Daw et al., 2005) imply that habits and model-free actions are driven by a reward-maximization process, that is, a process that depends directly on potential outcomes. Given that habit strength, the parameter that conventional habit theorists consider to be crucial for the probability to select a stimulus-response association, can well be considered a kind of selection value, the difference between value-free and value-based modeling might be less dramatic than Miller et al. (2019) assume. However, in their review on habits, Wood and R  nger (2016) question whether habits can be equated with model-free learning in view of suggestions that habits are acquired through model-based processes (Dezfouli and Balleine, 2012) and failures to find relationships between the strength of model-free learning and habit formation in individual-difference studies (Friedel et al., 2014; Gillan et al., 2015). Hence, it is clear that the model-free/model-based framework is still under development and it remains to be seen whether a systematic connection between

model-based/model-free learning on the one hand and will/habit on the other will emerge. In any case, model-free action is considered to be insensitive to current action goals, whereas model-based algorithms are assumed to compute transition probabilities (e.g., an agent's likelihood of being in a wanted state after having performed a given action), which are used to compute the expected value of actions by comparing the states they are predicted to produce to the states the agent wants to establish. Some approaches assume that the two systems compete for action control (e.g., Gillan et al., 2015), while others assume that they can be integrated (Krueger and Griffiths, 2018). Some authors consider the model-based/model-free approach a strongly advanced version of the original will/habit approach (e.g., Dolan and Dayan, 2013), while others consider the two pairs of concepts basically equivalent (e.g., Friedel et al., 2014).

However, the probably most defining two novelties in the context of the model-based/model-free approach are the contrast between action-outcome contingencies, which are related to the model-based/goal-related system, and stimulus-response associations, which are the main ingredients of the model-free/habitual system (De Wit and Dickinson, 2009), and the experimental procedure used to test whether a particular action relies on one or the other system. The latter is based on Heyes and Dickinson's (1990) "desire criterion" of voluntary action, which together with the "belief criterion" serves as diagnostic indicator of whether a particular action is based on a goal. The belief criterion requires the voluntarily acting agent to know about the current action-outcome relation and the desire criterion requires him or her to actually want the current outcome. Given that voluntary action is commonly defined as an activity directed toward the creation of some intended effect, the belief criterion is uncontroversial and explicitly or implicitly shared by any approach to voluntary action control (see Hommel and Wiers, 2017). The role and relevance of the desire criterion is less clear, however. The key procedure to assess whether the desire criterion is fulfilled is test after satiation, which reflects the behaviorist heritage of the model-based/model-free approach and the fact that it is mainly based on experiments carried out with rodents. For instance, participants who like popcorn would be tested for popcorn-related actions before and after receiving the opportunity to eat as much popcorn as they like (e.g., Watson et al., 2014). If they would show similar attentional and behavioral biases toward popcorn after the satiation procedure as they showed before the procedure, the corresponding behavioral tendency would be considered to rely on the model-free system and the stimulus-response associations it contains. The rationale for that conclusion seems straightforward: the satiation procedure should make sure that participants no longer want popcorn, so if they would still be showing popcorn-approaching behavior this cannot rely on an active popcorn-getting goal—leaving a previously acquired popcorn-getting habit as the only option.

But is this rationale watertight? Let us consider why a person might eat popcorn. She may like digesting popcorn, feeling popcorn in her mouth, smelling popcorn, listening to the sound of popcorn being chewed, the attention she attracts

from other popcorn-loving individuals, the satisfaction of having access to one's favorite food, the entertainment of filling time with a liked activity, and more. Liking popcorn is thus not a simple desire for one single aspect of popcorn-eating behavior but rather a complex compound of what one might call desire aspects or subdesires. Which of those would be sated by eating as much popcorn as one likes? Being stuffed with popcorn might make the digesting aspect less attractive, but would it eliminate the joy experienced by any of the other aspects? How reasonable is it to expect that the intentional component of the behavior of a sated popcorn-lover would be identical to the behavior of a popcorn-hater or of one who just does not care about popcorn? I suggest that the fundamental flaw of satiation logic consists in the idea that agents have just one single goal and that this goal is comprehensively captured by the aspect of the goal that the satiation procedure is targeting (Hommel and Wiers, 2017). While it is not impossible that this is indeed the case, it is not very likely either.

Moreover, real human actions do not only rely on more than one goal aspect, they also consist of multiple elements: eating one popcorn consists in locating it in a nearby spot, moving one's hand toward it, opening and closing the hand until the popcorn is being grasped, moving it to one's mouth, opening the mouth, moving the popcorn inside, dropping it, closing the mouth, and starting to chew. Most of the elements of this action pattern have been discussed as the paragon of goal-directed voluntary action in the literature on grasping (e.g., Jeannerod, 1988; Milner and Goodale, 1995), which does not seem to fit with the classification of the entire pattern as a non-intentional stimulus-driven habit. One might object that the grasping part of the action may well be intentional and the popcorn part may not, but this is exactly my point: actions commonly comprise of multiple goals and it is unlikely that any satiation procedure can ever target all of them.

Finally, if all the popcorn-related behavior of the sated popcorn-lover would really be run by the model-free system alone, why would she actually eat the popcorn? Popcorn-lovers are likely to have done many things with popcorn apart from eating: buying and putting it into the bag, carrying it home and putting it into the cupboard, unpacking it and putting it on the table, offering it to others, cleaning the table from it, and throwing the remains into the trash, and so forth. The stimulus popcorn must thus be associated with many different responses, which raises the question which of the corresponding stimulus-response habits are triggered by the popcorn after satiation. What experiments show is that even the most popcorn-loving participants show contextually appropriate behavior even after satiation: they may eat some if they stand in front of it, but they do not clean the table from it, store them, or do other things that would not fit the experimental context and the social situation it creates. If so, satiating the popcorn-lover does not seem to prevent her from showing contextually and socially appropriate popcorn-related behavior, which is not well-covered by calling it model-free.

MISCONCEPTIONS IN BINARY THEORIZING

This brief and incomplete historical tour through some of the highlights of binary theorizing on action control was intended to show that none of the suggested terminological couples really works. Practicing stimulus-response combinations is likely to change the representations thereof, and presumably makes these representations more available under certain circumstances. However, there is still no evidence that stimuli can do what intentions and goals can: to trigger a particular response. What stimuli are capable of is to trigger misleading action tendencies under circumstances that are dictated by the kind and generality of the action goal, and to the degree that they are primed and enabled by the goal, whereas the actual association strength often fails to predict the degree to which representations of stimulus-response combinations affects action control. The opposition of controlled and automatic processes suffers from similar problems and from the lack of convincing demonstrations of true automaticity. The available demonstrations are consistent with the idea that automatic processes are enabled by the goal (as suggested by Exner, 1879; James, 1890; Bargh, 1989; Gollwitzer, 1993), so that it is the goal that eventually determines whether what is considered to be an automatic process has any impact on action selection. If the model-based/model-free approach goes beyond the will/habit approach at all, which is not always clear, it does not make a convincing case that satiation procedures are a diagnostic method to tell truly goal-driven from purely stimulus-driven actions. The main problem is that this approach systematically underestimates the complexity of human action planning, a possible reflection of its behaviorist heritage. One complaint about binary theorizing has been that, even though action-control processes can be easily divided into two categories, the various categories that researchers have created so far do not sufficiently overlap to make a convincing coherent story (Melnikoff and Bargh, 2018). Even though I agree, I would even argue that the criteria offered so far have been ill-conceived and failed to allow sorting processes into non-overlapping categories. The reasons for that, I believe, have to do with some fundamental misconceptions regarding (1) the temporal relationship between the operation of processes assumed to reflect the goal and the operation of processes that are assumed to be automatic; (2) the aspects of actions that control operations keep themselves busy with; and (3) the number of goals involved in action control. In the following, I will discuss each misconception in turn.

The Competition Misconception

When he was laying the ground for modern reaction-time-based analyses of human cognitive processes, Donders (1868/1969) was optimistic to have measured the time demands of what he called the “expression of the will.” By cleverly manipulating the cognitive demands of rather simple reaction-time experiments, and by subtracting the corresponding reaction times, Donders estimated the time demands of what we nowadays

would call “response selection” in a binary-choice task to about 1/28 s. More important than the validity of this estimate is the time point at which Donders thought that the will would express itself: between processing the stimulus information and executing the response. Once we replace the outdated terms “will” and “expression of the will” through their modern successors “goal” and “controlled process,” we can see that the main function of controlled processes are thus assumed to consist in stimulus-response translation. This scenario perfectly fits with most modern action-control approaches, including the model of Kornblum et al. (1990), where the stimulus-guided “identification of the correct response” is actually the only control(led) process. It is this process that is assumed to compete with the habitual, automatic, or model-free process for controlling the eventual action.

Even though Donders’ view turned out to provide the basic theoretical template for modern action-control approaches, alternatives were available. In particular, Exner (1879) rejected the idea that the will intervenes between stimulus and response processing. Instead, he argued that preparing for a task or a particular action is accomplished by turning oneself into an automatic system long before the first stimulus appears. It is this automatic state that according to Exner enables humans to act efficiently. Note that the temporal relationship between actual control and automaticity has changed from concurrent competition to a sequence in which control operations set up the stage for automatic processes to take over. Exner’s view provides an excellent theoretical framework for understanding the observations of Valle-Inclán and Redondo (1998) discussed above: automaticity can indeed be demonstrated but it depends on the implementation of the action goal, just as the conditional-automaticity approach has claimed (Bargh, 1989). Hence, rather than competing with habitual, automatic, and model-free processes, goal-related control processes turn the cognitive system into a “prepared reflex,” as Woodworth (1938) has called it (see Hommel, 2000b).

The Process-Control Misconception

One of the oldest theoretical problems that experimental psychology deals with relates to what Turvey (1977) has called “executive ignorance”: how is it possible that humans can carry out intentional actions but, if being asked how they did so, have very little of interest to report? The answer favored by ideomotor theorists since Lotze (1852) and James (1890) consists in the assumption of a mechanism that integrates co-activated representations of the sensory consequences of a movement (reafferent information) and the motor patterns generating these consequences. According to this view, infants and other novices start by motor babbling—performing relatively random movements—and integrate the produced motor patterns with the sensory consequences thereof (i.e., action effects). Once they have experienced action effects they like or find functional in achieving a particular goal, they “imagine,” “expect,” or “predict” these consequences, which functionally translates into reactivating the sensory representations of action effects. Given that these representations have been integrated with the motor patterns that have generated them in the past, reactivating

them will prime and eventually activate the associated motor patterns, which is likely to reproduce the (now intended) sensory consequences.

Recent research has provided strong evidence for the existence of such an ideomotor mechanism, unraveled its neural and functional underpinnings, and its role in the development of intentional action (for reviews, see Hommel, 2009; Shin et al., 2010). However, for present purposes, the only important implication of this research relates to the target of control. If it is true that all that an intentionally acting agent has available are representations of past (and now expected) sensory consequences of movements, it is clear that action planning mainly consists in the activation and maintenance of these representations. In other words, action control deals with and operates on representations of expected sensory outcomes. While this might sound obvious, it is important to emphasize that this does not imply that action control is targeting particular processes. It is in fact the inability to intentionally target particular processes—the executive ignorance—that has provided the main impetus for ideomotor approaches to emerge. It thus makes little sense to compare processes that are thought to be controlled with processes that are thought to be not controlled or, as in most approaches, controlled by external stimuli. Instead, it makes more sense to assume that implementing a particular goal establishes a condition that allows representations of action-outcome relations to compete, and the representation with the closest fit to the intended action effect to win, at least under ideal circumstances (see Hommel and Wiers, 2017, for elaboration). If so, it would only be the implementation of the goal that could meaningfully be referred to as intentional or controlled, while the resulting competition would be fully automatic—just as Exner envisioned.

From this perspective, stimuli might be able to activate particular goals but, once a particular goal is implemented, they would not be able to make an agent perform an action that is entirely unrelated to that goal. And this is indeed what all available purported demonstrations of automaticity show: if a participant commits an error in a manual Stroop task, she is very unlikely to actually speak the word out loud—even though this should theoretically be the strongest habit and the most automatic tendency—but rather press the key that corresponds to the color designated by this word. Note that this error is anything but model-free, as it reflects many aspects of the task instruction, actually results from obviously outcompeting the strongest habit, and takes into account the goal of intending to press keys, rather than to say something or do something else. In other words, the error reflects the consideration of almost all aspects of the goal and the task model—something that arguably undermines all available binary accounts.

The Single-Goal Misconception

Distinguishing between goal-related and automatic processes requires a good understanding of what the current goal actually is. Researchers implicitly or explicitly identify the current goal with reference to the instruction: aspects of the task that were considered relevant in the instruction are assumed to

be represented by the goal whereas aspects of the task that were considered irrelevant are not. If thus evidence for processing the latter can be obtained, this is taken as evidence for control leakage and, thus, automatic processing. Importantly, the logic of this rationale presupposes that people have only one goal at a time, which unfortunately is entirely unrealistic. According to Atkinson and Birch (1970), the stream of human behavior is driven by multiple internal response tendencies that continuously vary in strength. Vallacher and Wegner (1987) have suggested that actions can be described at various levels, due to the concrete action plans being commonly nested into more abstract action plans, which are part of even more abstract plans, etc. Indeed, if a student is participating in a Stroop task, she is unlikely to give up her plans to earn some money, to complete her studies in time, to become a famous scientist, to be a sympathetic person, and to lead a happy life when entering the lab. How are all these goals, small-scale and large-scale, long-term and short-term, reflected in current theorizing on action control? I am afraid they are not.

That this has severe consequences for our understanding of action control can be easily shown. As discussed earlier, tasks that are thought to tap into action control give participants mixed messages about the relevance of processing particular information. In the Stroop task, words are explicitly declared to be irrelevant and yet in a substantial portion of the trials, often up to 50%, processing the word or even reading it pays off, and the argument holds for Simon tasks, flanker tasks, and many other versions of them as well. Mixed messages of this kind are likely to undermine the instructed ignorance to the type of information that the instruction has declared irrelevant. Why would a system that is assumed to be attuned to optimizing reward, as the human cognitive system, not be sensitive to the possibility to receive reward in 50% of the trials? Moreover, researchers commonly try to counteract reward-sensitive strategies by varying the irrelevant information in an unpredictable fashion. This however implies considerable variability with respect to the irrelevant stimulus dimension. Variability implies uncertainty, and the human cognitive system is notoriously interested in reducing uncertainty. This has been emphasized in recent predictive-coding approaches (Friston, 2009) but also featured strongly in the approach of Berlyne (1949, 1960). Berlyne has claimed that one of the major human drives consists in curiosity—a chronic goal that is unlikely to be traded for a Stroop instruction. Curiosity is assumed to be attracted to stimulus aspects of maximal uncertainty, which the cognitive system then tries to reduce by improving its expectations (Sokolov, 1963) or, in more fashionable terms, its predictions (Friston, 2009). If we thus assume that participants bring their curiosity goal to our labs, it should not be overly surprising that they are particularly interested in information satisfying it. If they are, this would not indicate a lack of goal-related action control but rather imply that participants satisfy various goals concurrently. Among other things, this predicts that effects hitherto assumed to reflect a leakage of control decrease as irrelevant information becomes less uncertain—which is exactly what Frings et al. (2019) have observed.

A UNITARY ALTERNATIVE

As I have tried to argue, binary theorizing that divides actions into willed and un-willed categories does not provide us with a useful perspective to understand action control, neither in the disguise of the will/habit opposition, nor in the case of the intentional/automatic opposition, nor with the model-based/model-free opposition.³ There can be little doubt that practice changes the representation of stimulus and action events, that it creates associations between the codes forming these representations, and that these associations have impact on action control. However, there is no systematic evidence suggesting that the amount of practice can predict which actions people choose, or that people choose actions that are unrelated to their current goals. Rather, it seems that goals set the stage for the competition of various, presumably automatic processes. Given that people control goals, rather than processes, it is always possible that one of the processes being involved in the competition turns out to be less functional than others, but this is a normal outcome of processing in a system that is as competitive as the human brain. As argued and developed in some detail elsewhere (Hommel and Wiers, 2017), the time seems ripe to move on toward a more integrative framework of human action control: a framework that embraces the complexity of action control and that goes beyond mere binary categorization, both in terms of functional explanation and with respect to the neural mechanisms. In the following, I will briefly sketch the core concepts of Hommel and Wiers' Unitary Model of Action Control (UMAC; the interested reader is referred to Hommel and Wiers, 2017, for more detail) and relate them to existing dual-route models.

According to UMAC, selecting an action is biased by multiple goals. Goals are functionally represented by one or many selection criteria that serve to provide top-down support for representations of actions that are expected to meet these criteria. For instance, the decision to grasp a cup of coffee on a table by means of one's right hand might be driven by selection criteria that promote actions that involve grasping, actions that serve reaching a cup, actions that are likely to have positive consequences, actions that are easy to perform, and actions that go fast. The selection criteria might be taken to represent multiple goals, like quenching one's thirst for coffee, moving with little effort, having fun, and pushing one's energy, but UMAC does not require the specification or even

the integration of dedicated goals—all that counts are activated selection criteria. Given that many of the criteria will be satisfied by more than one action representation, the (entirely automatic!) competition between suitable representations might be fierce but eventually be gravitating toward the representation of the action that best meets most or all of the criteria. Note that this scenario implies both: that all actions reflect goal states and that all actions are selected automatically. In other words, all actions are both intentional and automatic.

Highly overlearned actions or actions that the agent has preferred to choose under coffee-drinking circumstances may well have a selection benefit in the competition, because they had been learned to have low control demands (i.e., they meet the easy-to-perform criterion particularly well) and to go fast (i.e., they meet the high-speed criterion particularly well). However, it is important to emphasize that the degree of overlearning as such does not render the corresponding action special (or "more automatic") in any way. There would be nothing wrong with calling the corresponding action a habit, simply because the agent tends to prefer this action over others—which is the defining criterion for calling something a "habit" in everyday communication. But the habitual character only exists in the eyes of the observer—the agent simply selects an action that is fast and easy. In other words, the key difference between binary theories and UMAC is that the former assume that particular actions tend to be chosen *because they are habits* that happen to be fast and efficient, whereas the latter (e.g., Moors and de Houwer, 2006) assumes that they are chosen *exactly because they are fast and efficient*. Whereas the former reasoning implies that the selection of a habit is non-intentional, at least under some circumstances, the latter implies that the selection takes place because of the current goals—which of course may involve selection criteria other than my current examples speed and efficiency.

From a UMAC perspective, it makes little sense to develop any binary system to sort actions into two categories. While practicing an action may well increase the likelihood of selecting it in the future, there is no theoretical reason to reserve a dedicated label to overlearned actions. For instance, even if overlearning to open a door by pressing the handle down, to use Lewin's example, will make down-pressing a particularly fast and efficient action that is likely to be a strong competitor for selection under high-speed pressure (a selection criterion that propagates fast and efficient actions), a strong accuracy instruction is likely to render this candidate entirely impotent. Note that this theoretical problem cannot be solved by turning the binary distinction between intentional action and habit into a continuous dimension; it rather highlights the actual status of the word "habit," which should be considered a descriptive term taken from everyday language but not a scientific, and certainly not an explanatory concept (cf., Hommel, in press).

An obvious objection against UMAC might be that it is merely changing the semantics in a way that is impossible to test: every time some seemingly non-intentional behavior can be demonstrated, a new goal might be invented to account for it. This would indeed not do a good service to our

³Note that what I criticize is the way theorists have sorted actions, action-control operations, and related processes into two categories over the last 150 years or so. It is thus *a particular kind of binary theorizing* that I criticize, and my main argument is that the distinctions being drawn between the binary categories make little sense both theoretically and empirically. I would like to emphasize that I am more interested in the flaws in making these distinctions than in the binary nature of the underlying theorizing. Accordingly, theories that would keep that distinction but add further categories would not escape my criticism. Conversely, binary theories that make other distinctions than between willed and un-willed (and related versions) may well escape it, even though I find it difficult to imagine what kind of distinction that might be and even though I would suspect that it would still tempt researchers to categorize actions and related processes rather than understanding their mechanics (the tendency that I criticize in Hommel, in press).

understanding of action control, but fortunately UMAC is not at all immune to empirical test, as the following examples show. First, a key point of UMAC is that implementing an action goal/intention enables (increases the possible impact of) event representations with features from dimensions that either are or seem to be task-relevant. It is this task-relevance that renders the tendency to say “red” in a Stroop task a potent competitor in action selection. A strict automaticity approach could thus easily disconfirm the corresponding UMAC prediction by demonstrating that people say “red” when being faced with the word “RED” in the absence of any task or in a task that neither requires reading nor otherwise dealing with colors or color words. Second, even though it may be difficult to create conditions under which chronic goals like curiosity or novelty-seeking can be entirely switched off, it is certainly possible to create conditions that make that goal more or less relevant, like in a dual-task paradigm with one task emphasizing or not emphasizing novel information. Demonstrating that such a manipulation has no impact on the processing of novel information whatsoever would be difficult to take for UMAC.

Another interesting issue in the comparison of UMAC and strict automaticity approaches relates to the role of external stimuli. Both approaches assume that action alternatives can be activated by processing external stimuli: the automaticity approach assumes that processes and even actions can be triggered by stimuli—where the latter, as I have argued above, is yet to be demonstrated in humans—and UMAC assumes that stimuli activate all representations that feature-overlap with the stimulus on task-relevant dimensions (Hommel, 2004; Hommel and Wiers, 2017). The critical difference between these two theoretical approaches does thus not relate to the possibility of stimulus-induced activation of internal representations but rather to the question whether the degree of this activation is moderated by task-relevance (which UMAC assumes but the automaticity approach does not) and whether activation can result in action, as the automaticity approach assumes, or in competition for action control according to goal criteria, as UMAC suggests.

Yet another difference refers to the role of the context. Many automaticity accounts imply a rather pure, de-contextualized connection between particular stimuli and overlearned responses to these stimuli (e.g., Dickinson, 1985; De Wit and Dickinson, 2009). In contrast, UMAC assumes that the basic representational unit is an event file (Hommel, 2004), which integrates stimuli, actions, and outcomes, as well as internal and external context conditions. This feature allows UMAC to deal with findings as those reported by Neal et al. (2011). These authors found that participants who are used to eating popcorn in the cinema are likely to eat popcorn even if it is stale and even though they report disliking it, but only if it is offered in the cinema but not in a lab room while watching music videos. Even though more research is required to identify further conditions of such observations, UMAC’s assumption that action representations are contextualized and, thus, more likely retrieved and more strongly activated in a context in which they were acquired, is well-equipped to tackle such empirical challenges in principle.

Last but not least, it is important to point out that UMAC does not deny the important role of practice—the key player

of automaticity accounts. According to UMAC, practice can change behavior in various ways that have an impact on action control, that is, on the probability that the event file related to a practiced action is eventually selected for execution. For instance, practice is known to increase the speed and efficiency at which an action is carried out. Increasing practice will thus increase the number of event files that satisfy goals that emphasize or imply speed and efficiency, which will make these event files more likely to outcompete others if and to the degree that these goals are activated. Practice will also lead to a more systematic, sharpened integration of other action effects, so that the experienced popcorn-eater, say, will have learned and will thus anticipate a richer and more specific set of sensory outcomes of popcorn eating than the popcorn greenhorn. This in turn will make the resulting event files more potent competitors under conditions in which goals that are satisfied by such outcomes are activated. For instance, it may take some time to register and appreciate social-improvement signals from other popcorn-eaters in the cinema, so that popcorn eating is more likely to satisfy social goals in the more experienced popcorn-eater. Practice may also increase or reduce the role of context, depending on the kind of experience: if 90% of the event files resulting from one’s street-crossing experience contain a representation of a green light, encountering a green light is likely to play a stronger role in selecting the appropriate action than if light representations in street-crossing event files would be much more varied. UMAC and automaticity accounts do thus not differ with respect to the assumption that practice and learning can have a strong impact on action control, but they rather differ with respect to why and how this impact is thought to be achieved. If, thus, the popcorn-lover keeps eating popcorn even after having finished an XXL tube, this might reflect the ongoing satisfaction of (e.g., tactile, olfactory, or social) goals that are not yet sated, or simply an attempt to fight boredom, rather than a breakdown of intentionality.

CONCLUSION

The unitary account to action control shows that there is no need to heed the conventional distinction between will and habit. In this framework, goals still play an important role, as do automatic processes and practice, but goals and automatic processes do not compete but serve complementary purposes. The next challenge will be to better understand how goals and selection criteria constrain the operation of automatic processes, and when and under which circumstances action representations become relevant competitors in the action-selection process. In any case, I believe that theorizing about action control is ready to take the next step, and that this next step should not consist in inventing yet another binary opposition.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This research was funded by an Advanced Grant of the European Research Council (ERC-2015-AdG-694722) to the author.

REFERENCES

- Ach, N. (1910). *Über den Willensakt und das Temperament*. Leipzig: Quelle & Meyer.
- Ach, N. (1935). "Analyse des Willens" in *Handbuch der biologischen Arbeitsmethoden*. Vol. VI. ed. E. Abderhalden (Berlin: Urban & Schwarzenberg).
- Atkinson, J. W., and Birch, D. (1970). *The dynamics of action*. New York: Wiley.
- Atkinson, R. C., and Shiffrin, R. M. (1968). "Human memory: a proposed system and its control processes" in *The psychology of learning and motivation*. Vol. 2. eds. K. W. Spence and J. T. Spence (New York: Academic Press), 89–195.
- Bargh, J. A. (1989). "Conditional automaticity: varieties of automatic influence in social perception and cognition" in *Unintended thought*. eds. J. S. Uleman and J. A. Bargh (London: Guilford Press), 3–51.
- Bargh, J. A. (1994). "The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition" in *Handbook of social cognition: Basic processes; applications*. eds. R. S. Wyer Jr. and T. K. Srull (Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.), 1–40.
- Berlyne, D. E. (1949). 'Interest' as a psychological concept. *Br. J. Psychol.* 39, 184–195. doi: 10.1111/j.2044-8295.1949.tb00219.x
- Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York: McGraw-Hill.
- Broadbent, D. (1958). *Perception and communication*. London: Pergamon Press.
- Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560
- De Wit, S., and Dickinson, A. (2009). Associative theories of goal-directed behaviour: a case for animal-human translational models. *Psychol. Res.* 73, 463–476. doi: 10.1007/s00426-009-0230-6
- Dezfouli, A., and Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* 35, 1036–1051. doi: 10.1111/j.1460-9568.2012.08050.x
- Dickinson, A. (1985). Actions and habits: the development of behavioural and autonomy. *Philos. Trans. Royal Soc. B* 308, 67–78. doi: 10.1098/rstb.1985.0010
- Dickinson, A., and Balleine, B. W. (1994). Motivational control of goal-directed action. *Anim. Learn. Behav.* 22, 1–18. doi: 10.3758/BF03199951
- Dolan, R. J., and Dayan, P. (2013). Goals and habits in the brain. *Neuron* 80, 312–325. doi: 10.1016/j.neuron.2013.09.007
- Donders, F. C. (1868/1969). *De snelheid van psychische processen*. Gent: University of Gent. English translation: on the speed of mental processes. *Acta Psychol.* 30, 412–431.
- Eimer, M. (1995). Stimulus-response compatibility and automatic response activation: evidence from psychophysiological studies. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 837–854. doi: 10.1037//0096-1523.21.4.837
- Exner, S. (1879). "Physiologie der Grosshirnrinde" in *Handbuch der Physiologie*. 2. Band. 2. Theil. ed. L. Hermann (Leipzig: Vogel), 189–350.
- Fitts, P. M. (1951). "Engineering psychology and equipment design" in *Handbook of experimental psychology*. ed. S. S. Stevens (New York: Wiley), 1287–1340.
- Friedel, E., Koch, S. P., Wendt, J., Heinz, A., Deserno, L., and Schlagenhauf, F. (2014). Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Front. Hum. Neurosci.* 8:587. doi: 10.3389/fnhum.2014.00587
- Frings, C., Merz, S., and Hommel, B. (2019). The impact of stimulus uncertainty on attentional control. *Cognition* 183, 208–212. doi: 10.1016/j.cognition.2018.10.017
- Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* 13, 293–301. doi: 10.1016/j.tics.2009.04.005
- Gillan, C. M., Otto, A. R., Phelps, E. A., and Daw, N. D. (2015). Model-based learning protects against forming habits. *Cogn. Affect. Behav. Neurosci.* 15, 523–536. doi: 10.3758/s13415-015-0347-6
- Gollwitzer, P. M. (1993). Goal achievement: the role of intentions. *Eur. Rev. Soc. Psychol.* 4, 141–185. doi: 10.1080/14792779343000059
- Goschke, T. (2003). "Voluntary action and cognitive control from a cognitive neuroscience perspective" in *Voluntary action: Brains, minds, and sociality*.

ACKNOWLEDGMENTS

The author is grateful to the reviewers and editors for their efforts to help improving this article.

- eds. S. Maasen, W. Prinz, and G. Roth (New York, NY: Oxford University Press), 49–85.
- Heyes, C., and Dickinson, A. (1990). The intentionality of animal action. *Mind Lang.* 5, 87–104. doi: 10.1111/j.1468-0017.1990.tb00154.x
- Hommel, B. (2000a). "Intentional control of automatic stimulus-response translation" in *Interaction between dissociable conscious and nonconscious processes*. eds. Y. Rossetti and A. Revonsuo (Amsterdam: John Benjamins Publishing Company), 223–244.
- Hommel, B. (2000b). "The prepared reflex: automaticity and control in stimulus-response translation" in *Control of cognitive processes: Attention and performance XVIII*. eds. S. Monsell and J. Driver (Cambridge, MA: MIT Press), 247–273.
- Hommel, B. (2004). Event files: feature binding in and across perception and action. *Trends Cogn. Sci.* 8, 494–500. doi: 10.1016/j.tics.2004.08.007
- Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychol. Res.* 73, 512–526. doi: 10.1007/s00426-009-0234-2
- Hommel, B. (in press). Pseudo-mechanistic explanations in psychology and cognitive neuroscience. *Top. Cogn. Sci.*
- Hommel, B., and Wiers, R. W. (2017). Towards a unitary approach to human action control. *Trends Cogn. Sci.* 21, 940–949. doi: 10.1016/j.tics.2017.09.009
- James, W. (1890). *The principles of psychology*. Vol. 2. New York: Dover Publications.
- Jeannerod, M. (1988). *The neural and behavioural organization of goal-directed movements*. New York, NY: Clarendon Press/Oxford University Press.
- Johnson, M. K., and Hasher, L. (1987). Human learning and memory. *Annu. Rev. Psychol.* 38, 631–668. doi: 10.1146/annurev.ps.38.020187.003215
- Kahneman, D., and Treisman, A. (1984). "Changing views of attention and automaticity" in *Varieties of attention*. eds. R. Parasuraman and D. R. Davies (Orlando, FL: Academic Press), 29–61.
- Kornblum, S., Hasbroucq, T., and Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility -a model and taxonomy. *Psychol. Rev.* 97, 253–270. doi: 10.1037/0033-295X.97.2.253
- Krueger, P. M., and Griffiths, T. L. (2018). Shaping model-free habits with model-based goals. Proceedings of the 40th Annual Conference of the Cognitive Science Society. Available at: <https://pdfs.semanticscholar.org/4a31/47b00de1e5fe2625636a01f1b4d6cb31391e.pdf> (Accessed May 3, 2019).
- Lewin, K. (1922a). Das Problem der Willensmessung und das Grundgesetz der Assoziation I. *Psychol. Forsch.* 1, 191–302.
- Lewin, K. (1922b). Das Problem der Willensmessung und das Grundgesetz der Assoziation II. *Psychol. Forsch.* 2, 65–140.
- Lewin, K. (1928). Die Entwicklung der experimentellen Willenspsychologie und die Psychotherapie. *Arch. Psychiat.* 85, 515–537.
- Lewin, K. (1931). The conflict between Aristotelian and Galilean modes of thought in contemporary psychology. *J. Gen. Psychol.* 5, 141–176. doi: 10.1080/00221309.1931.9918387
- Lin, P. Y., Wood, W., and Monterosso, J. (2016). Healthy eating habits protect against temptations. *Appetite* 103, 432–440. doi: 10.1016/j.appet.2015.11.011
- Logan, G. D., and Zbrodoff, N. J. (1979). When it helps to be misled: facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Mem. Cogn.* 7, 166–174. doi: 10.3758/BF03197535
- Lotze, R. H. (1852). *Medizinische Psychologie oder die Physiologie der Seele*. Leipzig: Weidmann'sche Buchhandlung.
- MacLeod, C. M., and Dunbar, K. (1988). Training and Stroop-like interference: evidence for a continuum of automaticity. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 126–135. doi: 10.1037//0278-7393.14.1.126
- Martin, M. (1978). Speech recoding in silent reading. *Mem. Cogn.* 30, 187–200.
- Mascolo, M., and Hirtle, S. C. (1990). Verbal coding and the elimination of Stroop interference in a matching task. *Am. J. Psychol.* 103, 195–215. doi: 10.2307/1423142
- McClain, L. (1983). Effects of response type and set size on Stroop color-word performance. *Percept. Mot. Skills* 56, 735–743.
- Melnikoff, D. E., and Bargh, J. A. (2018). The mythical number two. *Trends Cogn. Sci.* 22, 280–293. doi: 10.1016/j.tics.2018.02.001

- Michotte, A., and Prüm, E. (1911). Le choix volontaire et ses antécédents immédiats. *Arch. Psychol.* 10, 113–299.
- Miller, K. J., Shenhav, A., and Ludvig, E. A. (2019). Habits without values. *Psychol. Rev.* 126, 292–311. doi: 10.1037/rev0000120
- Milner, D., and Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.
- Monsell, S., and Driver, J. (2000). “Banishing the homunculus” in *Control of cognitive processes: Attention and performance XVIII*. eds. S. Monsell, and J. Driver (Cambridge, MA: MIT Press), 3–32.
- Moors, A., and de Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychol. Bull.* 132, 297–326. doi: 10.1037/0033-2909.132.2.297
- Neal, D. T., Wood, W., Wu, M., and Kurlander, D. (2011). The pull of the past: when do habits persist despite conflict with motives? *Personal. Soc. Psychol. Bull.* 37, 1428–1437. doi: 10.1177/0146167211419863
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Newell, A. (1973). “You can’t play 20 questions with nature and win” in *Visual information processing*. ed. W. G. Chase (New York: Academic Press), 283–308.
- Schneider, W., Dumais, S. T., and Shiffrin, R. (1984). “Automatic and control processing and attention” in *Varieties of attention*. eds. R. S. Parasuraman, and D. R. Davies (New York: Academic Press), 1–27.
- Schwabe, L., and Wolf, O. T. (2009). Stress prompts habit behavior in humans. *J. Neurosci.* 29, 7191–7198. doi: 10.1523/jneurosci.0979-09.2009
- Shin, Y. K., Proctor, R. W., and Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychol. Bull.* 136, 943–974. doi: 10.1037/a0020541
- Simon, J. R. (1969). Reactions toward the source of stimulation. *J. Exp. Psychol.* 81, 174–176. doi: 10.1037/h0027448
- Smith, E. R., and Lerner, M. (1986). Development of automatism of social judgments. *J. Pers. Soc. Psychol.* 50, 246–259. doi: 10.1037/0022-3514.50.2.246
- Sokolov, E. N. (1963). *Perception and the conditioned reflex*. New York: MacMillan.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 28, 643–662.
- Sutton, R. S., and Barto, A. G. (2017). *Reinforcement learning: An introduction*, 2nd Edn. MIT Press.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychol. Rev.* 55, 189–208. doi: 10.1037/h0061626
- Trafimov, D. (2018). “The automaticity of habitual behaviours: inconvenient questions” in *The psychology of habit: Theory, mechanisms, change, and contexts*. ed. B. Verplanken (Heidelberg: Springer), 379–395.
- Turvey, M. T. (1977). “Preliminaries to a theory of action with reference to vision” in *Perceiving, acting and knowing: Toward an ecological psychology*. eds. R. Shaw, and J. Bransford (Hillsdale, NJ: Lawrence Erlbaum Associates).
- Vallacher, R. R., and Wegner, D. M. (1987). What do people think they’re doing? Action identification and human behavior. *Psychol. Rev.* 94, 3–15. doi: 10.1037/0033-295X.94.1.3
- Valle-Inclán, F., and Redondo, M. (1998). On the automaticity of ipsilateral response activation in the Simon effect. *Psychophysiology* 35, 366–371. doi: 10.1111/1469-8986.3540366
- Watson, P., Wiers, R. W., Hommel, B., and de Wit, S. (2014). Working for food you don’t desire: cues interfere with goal-directed food-seeking. *Appetite* 79, 139–148. doi: 10.1016/j.appet.2014.04.005
- Wheatley, T. P., and Wegner, D. M. (2001). “Automaticity in action” in *International encyclopedia of the social and behavioral sciences*. eds. N. J. Smelser and P. B. Baltes (London: Pergamon), 991–993.
- Wood, W., and Rünger, D. (2016). Psychology of habit. *Annu. Rev. Psychol.* 67, 289–314. doi: 10.1146/annurev-psych-122414-033417
- Woodworth, R. S. (1938). *Experimental psychology*. New York: Holt, Rinehart & Winston.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Hommel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Model-Based and Model-Free Social Cognition: Investigating the Role of Habit in Social Attitude Formation and Choice

Leor M. Hackel^{1*}, Jeffrey J. Berg², Björn R. Lindström³ and David M. Amodio^{2,3*}

¹ Department of Psychology, University of Southern California, Los Angeles, CA, United States, ² Department of Psychology, New York University, New York, NY, United States, ³ Department of Psychology, University of Amsterdam, Amsterdam, Netherlands

OPEN ACCESS

Edited by:

John A. Bargh,
Yale University, United States

Reviewed by:

Kent Berridge,
University of Michigan, United States
Jon Juvina,
Wright State University, United States

*Correspondence:

Leor M. Hackel
lhackel@usc.edu
David M. Amodio
david.amodio@nyu.edu

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 20 May 2019

Accepted: 31 October 2019

Published: 21 November 2019

Citation:

Hackel LM, Berg JJ,
Lindström BR and Amodio DM (2019)
Model-Based and Model-Free Social
Cognition: Investigating the Role
of Habit in Social Attitude Formation
and Choice. *Front. Psychol.* 10:2592.
doi: 10.3389/fpsyg.2019.02592

Do habits play a role in our social impressions? To investigate the contribution of habits to the formation of social attitudes, we examined the roles of model-free and model-based reinforcement learning in social interactions – computations linked in past work to habit and planning, respectively. Participants in this study learned about novel individuals in a sequential reinforcement learning paradigm, choosing financial advisors who led them to high- or low-paying stocks. Results indicated that participants relied on both model-based and model-free learning, such that each type of learning was expressed in both advisor choices and post-task self-reported liking of advisors. Specifically, participants preferred advisors who could provide large future rewards as well as advisors who had provided them with large rewards in the past. Although participants relied more heavily on model-based learning overall, they varied in their use of model-based and model-free learning strategies, and this individual difference influenced the way in which learning related to self-reported attitudes: among participants who relied more on model-free learning, model-free social learning related more to post-task attitudes. We discuss implications for attitudes, trait impressions, and social behavior, as well as the role of habits in a memory systems model of social cognition.

Keywords: social, cognition, attitude, learning, habit, model-free, model-based, computational

MODEL-BASED AND MODEL-FREE SOCIAL COGNITION

Human thriving depends on social relationships, and the impressions we form of new acquaintances are essential guides to our social behavior (Fitzsimons and Anderson, 2013). We befriend people who are kind, hire people who are competent, avoid those who are domineering, or seek counsel from those who are empathic. In this way, impression formation often serves our goals (Brewer, 1988; Fiske and Neuberg, 1990; Bargh and Ferguson, 2000), as we use our knowledge of other people – their traits, mental states, and behaviors – to predict their actions and decide whether to interact with them (Heider, 1958; Tamir and Thornton, 2018).

Yet, while goals drive much of human behavior, this is not always the case. Habits, in particular, are responses that occur automatically and independent of our goals, often representing a highly-repeated behavior that was once goal-directed but that persists and is expressed even when the goal has changed (Wood and Rünger, 2016; Robbins and Costa, 2017). Habits likely

explain many behaviors, from benign compulsions like biting one's nails to more harmful acts like mindlessly reaching for a cigarette. Here, we asked whether habit-like processes may also contribute to social cognition – how we learn about, interact with, and evaluate other people – and thus help explain social behaviors that appear to occur independently of, or in opposition to, one's goals.

Multiple Systems for Social Learning

Research on impression formation has, to date, primarily emphasized conceptual forms of learning that give rise to goal-directed behavior; that is, acquiring conceptual knowledge about a person's traits and behavior (Uleman and Kressel, 2013). Early theories of impression formation focused on instructed forms of learning, in which we learn about a person from descriptions shared by others (Asch, 1946; Wyer and Carlston, 1979). If we are told that Bob is generous and friendly, we may infer that he's a good person. We can also learn about other people through observation and the use of attributional processing (Heider, 1958; Jones and Davis, 1965; Rydell and McConnell, 2006). If we see Jane offer money to a homeless person, we may infer from her actions that she is generous; if we see Jane choose a high-performing stock, we may infer that she is competent. These conceptual inferences can give rise to goal-directed behaviors, like choosing to spend time with someone who is generous or to hire someone who is competent.

More recent research has shown that social attitudes and impressions can also be formed through reward-based instrumental learning in direct social interaction – trial-and-error learning in which people make choices and receive feedback (Hackel et al., 2015). For instance, one might choose a lunch partner and experience rewards when they share their food, or one might hire a financial advisor and experience rewards when their advice pays off. Through this feedback, one can learn the reward value of an individual while also inferring aspects of their character traits (Hackel et al., 2015). Unlike instructed and observational forms of learning, which are typically passive (e.g., reading about another person), instrumental learning is active: it concerns feedback from another person regarding one's own actions. If, on most days, Bob's greeting to Jane is met with a smile, he will associate reward with his behavior toward Jane in addition to inferring that she is friendly.

Instrumental learning thus represents a distinct mode of learning in social interactions relative to conceptual knowledge (Amodio, 2019). Instead of inferring other people's qualities in order to decide how to interact with them, instrumental learning involves learning the reward value of social interaction through direct action and feedback. That is, in traditional impression formation approaches, Bob learns to interact with Jane because he infers she is friendly, and he wants to be around friendly people. In instrumental learning, Bob learns to interact with Jane because he previously did so and received rewarding outcomes, such as social rewards like smiles and compliments or material rewards like money and food. He may like Jane as a result of those rewards, rather than as a result of qualities he attributes to her. Thus, instrumental learning directly informs how we should interact with others given the rewards they provide. In this way,

preferences acquired through instrumental learning may be more directly tied to behavior.

A Role for Habits in Social Cognition?

Over time, instrumentally learned responses may be automatized into habits (Thorndike, 1911; Robbins and Costa, 2017). Although people may initially perform an action deliberately to achieve a goal, rewards can “stamp in” an association between a stimulus (or context) and a response, such that people later perform the response automatically. In contrast to skills, which are goal-directed action routines triggered intentionally, habits reflect a well-learned response that unfolds even when it is not consistent with a goal, and it persists even when its expression is no longer rewarded (Balleine and Dickinson, 1998; Tricomi et al., 2009; Wood and R  nger, 2016; Wood, 2017). Nevertheless, habits can be adaptive, initiating an important behavior that we might otherwise forget in the pursuit of another goal, such as grabbing our keys when rushing out the door to get to work in the morning.

Habits differ from other forms of unintentional learning that may contribute to impression formation. For example, spontaneous trait impressions (STIs) form when a perceiver is simply asked to read and memorize a set of trait-implicating sentences (Winter and Uleman, 1984; Carlston and Skowronski, 1994). People may be unaware that they formed an impression, yet STIs become evident in measures of cued recall and may subsequently influence judgment (Moskowitz and Roman, 1992). There is also evidence that evaluative conditioning, in which a neutral social target is paired repeatedly with either positive or negative images (Walther, 2002; Olson and Fazio, 2006), may even occur when such images are presented subliminally (e.g., De Houwer et al., 1997; Hofmann et al., 2010; but see Sweldens et al., 2014). However, both forms of learning involve passive exposure to stimuli and the formation of conceptual associations, likely supported by a semantic/conceptual associative memory system (Amodio and Berg, 2018; Amodio, 2019), in contrast to the active process of action-outcome learning involved in instrumental habit formation.

Examining Habit Formation Through Reinforcement Learning

A major challenge in the study of habits in humans is that it is often difficult to discern habits from other, goal-directed processes in behavior. However, this distinction has recently been linked to two forms of behavior within a computational account of reinforcement learning (Daw et al., 2011). Broadly, reinforcement learning algorithms describe how an agent learns the value of different actions with different states of the world by making choices and experiencing rewards (Sutton and Barto, 1998). According to this account, two types of computations can underlie reinforcement learning: Agents can engage in *model-based learning*, in which they consider the likely outcomes of their actions given knowledge about their environment, and also in *model-free learning*, in which they associate actions directly with reward value and repeat previously rewarded actions (Daw et al., 2011; Doll et al., 2015). Model-based learning is thus

prospective and goal-oriented, sensitive to both environmental contingencies (e.g., how to get to a reward) and expected outcomes (e.g., whether a desirable reward will be attained) – like a hungry mouse considering how to navigate a maze to reach the room with the tastiest cheese. In contrast, model-free learning is retrospective, relying on a past history of rewards for an action; it requires no internal model of one's environment and is insensitive to the outcomes an action will presently bring. A model-free learner stores cached values for previously performed actions and selects actions with the highest cached value.

Because model-free learning is computationally simpler but less flexible than model-based learning, it may give rise to behavior that has features of habits. For instance, an animal might continue to press a food lever despite being fully sated because this action was previously rewarded and thus associated with high reward value (Dickinson and Balleine, 1994; Daw et al., 2011). Although a model-free learner could eventually learn to adapt to the new value, it would persist in pressing the lever until learning takes place in its newly satiated state. In contrast, a model-based learner should not require this learning at all; instead, it should plan ahead to the likely outcome of the lever press, realize that it does not desire that outcome, and avoid the action from the start. Given these characteristics, the model-based/model-free distinction has been used recently to probe the role of habits in a range of learning contexts in humans. For instance, individuals who engage in greater model-based learning show less persistence in a devaluation task – a classic marker of habits (Gillan et al., 2015). Yet, to date, this approach has not been applied to questions on the formation of social impressions through direct social interactions with other people.

Model-Free Learning in Social Cognition

How might a model-based/model-free account relate to social impressions? When other people provide us with material feedback (like a gift) or social feedback (like a smile or a compliment), we experience this feedback as rewarding; as a result, this feedback can reinforce our social choices and draw us back to the same partners again in the future (Jones et al., 2011; Lin et al., 2011; Lindström et al., 2014; Hackel et al., 2015; Lindström and Tobler, 2018). If people learn from this feedback in a model-free manner, specifically, they might return to interaction partners previously associated with high reward regardless of whether those partners will currently provide desirable outcomes. This pattern would resemble a traditional definition of habit.

Some existing work hints at the possibility that reward feedback gives rise to social preferences that persist in a habit-like manner. In research by Hackel et al. (2015), participants played an economic game in which they chose partners who could share money; partners varied in the average *amount* they shared (indicating reward value) and average *proportion* they shared (indicating generosity). During initial learning, it was economically advantageous for participants to prefer individuals who provided large rewards, regardless of their generosity. However, when participants were later asked to choose one of

these partners to work with in a non-economic puzzle-solving task – a context where generosity, but not previous reward value, is advantageous – participants' choices were still influenced by partners' past reward value in addition to their generosity. This persistent influence of past reward – even when reward value no longer informed desired outcomes – suggests that participants may have developed model-free reward associations that guided subsequent social preferences. Nevertheless, past work has not directly tested this possibility by dissociating model-based and model-free learning in social interaction.

Study Overview

The present research was designed to provide initial evidence for model-free learning in social impression formation. To this end, we administered a sequential choice task commonly used to dissociate model-based and model-free learning (Kool et al., 2016; Kool et al., 2017; see also Daw et al., 2011), adapted to examine social partner choice and attitudes. On each round, participants chose financial advisors who had supposedly invested in one of two stocks; participants then received a payout from that advisor's stock. We examined the extent to which participants chose advisors based on model-based and model-free reinforcement, and further examined whether these forms of learning predicted participants' subjective attitude toward each advisor.

MATERIALS AND METHODS

Participants

Sixty-nine participants (42 male, 27 female) were recruited via Amazon Mechanical Turk (AMT), in exchange for \$3.50 for study completion, plus a monetary bonus based on their task performance. A sample size of 65 participants was chosen *a priori*; an additional four participants completed the task due to an error in which an extra set of slots was posted. Data collection was completed before analysis. Participants were eligible if they were located in the United States, completed at least one prior AMT study, and had approval rates of at least 95%. Informed consent was obtained from all participants in accordance with the guidelines of the New York University Committee on Activities Involving Human Subjects. We excluded data from participants who did not respond in time to either the first or second stage of a trial on more than 20% of trials (Kool et al., 2017). This rule excluded data from four participants, leaving data from 65 participants in analyses.

Procedure

Participation took place via Psiturk, an online platform for cognitive tasks (Gureckis et al., 2016). After providing consent, participants read a self-guided description of the study, which included practice trials, and completed the main experimental task. Next, participants completed self-reported evaluation items and a demographics questionnaire. Lastly, participants were informed of their bonus compensation for participating and then completed a debriefing procedure that included a suspicion probe and an explanation of study goals. All data exclusions, all

manipulations, and all measures included in this research are fully reported in this article.

Two-Step Task

We adapted a sequential learning task (Kool et al., 2016, 2017) designed to dissociate model-free and model-based learning (Figure 1). In our adaptation, participants were told they would learn about choices made by four AMT workers who previously participated in a financial decision-making study (see **Supplementary Material** for full task instructions). According to this cover story, these previous workers were assigned the role of “Financial Advisor,” in which they chose (only) one of two stocks (“Axiom” and “Zephyr”) to invest in for the duration of the study. These Advisors then earned money based on the performance of their chosen stock, which fluctuated throughout the study and could change from one round of “dividends” to the next.

Next, participants were assigned to the role of the “Client,” in which they would make a series of decisions about which Advisor to hire. Participants learned they would earn points based on the performance of the stock chosen by their hired Advisor on each round. Participants were explicitly told that the performance of the stocks would change over time (“a stock that was bad at the beginning of the game might start performing well, and a stock that initially pays well might perform poorly later on”), and that they should try to hire Advisors with the better performing stock at that particular moment. Moreover, participants were informed that they would receive a monetary bonus for their performance in the task, with better performance (in terms of points earned) equating to a larger bonus.

Return on each trial, participants began in one of two randomly chosen first-stage states. In these states, participants were presented with one of two pairs of Advisors, represented by distinct cartoon avatars (Figure 1). Avatars were randomly

assigned to different roles across participants (i.e., which stock they were linked with) and were equally likely to appear on the left or right side of the screen. Participants chose one of the two Advisors via button response and then transitioned deterministically to one of the two stocks, which comprised the second-stage states. That is, participants could reach either of the two stocks from each of the first-stage states; one Advisor in each pair always invested in the Axiom stock and the other Advisor in the given pair always invested in the Zephyr stock.

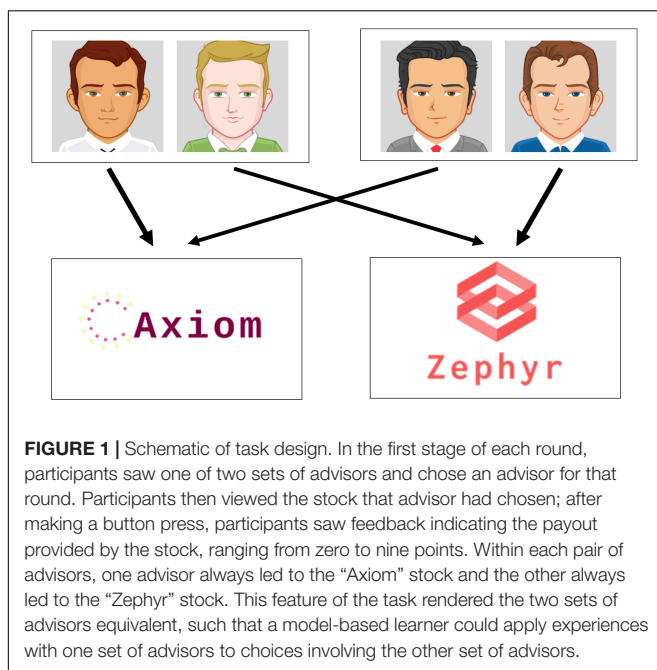
When they reached the second-stage state, participants were instructed to press the spacebar to reveal the performance of the stock in which the chosen Advisor invested. If participants did not respond in time to either the first- or second-stage states, no reward was provided and participants moved to the next trial. The number of points obtained for each stock fluctuated slowly and stochastically over the course of the task, varying according to a Gaussian random walk ($SD = 2$) with reflecting bounds at 0 and +9 points. The drifting nature of the reward feedback encouraged continuous learning throughout the task.

Importantly, the two first-stage states were equivalent in terms of the stocks they could lead to: within each pair of advisors, one Advisor always invested in the Axiom stock, whereas the other Advisor always invested in the Zephyr stock. This design allows for the separation of model-free and model-based control. Given that both stocks can be reached from each pair of Advisors, the stock reached from one set of advisors can be used by a model-based learner to update preferences regardless of which set of advisors is encountered on the next trial. For instance, if an Advisor in one pair invested in the Axiom stock and this stock paid out a large number of points on that trial, a model-based learner should subsequently be more likely to choose the Advisor in the other pair that also invests in the Axiom stock. That is, a model-based learner can generalize across equivalent first-stage choice options due to its exploitation of the overarching task structure. Conversely, model-free learners would not generalize across equivalent first-stage choice options, as they simply rely on directly-experienced action-outcome associations – the outcomes experienced following a choice in one pair of advisors should not affect preferences for the advisors in the second pair, and vice-versa.

Participants were trained extensively on the deterministic transitions (i.e., which financial advisor in a given pairing invested in which of the two stocks) prior to completing the experimental trials, such that 80% accuracy across 15 consecutive trials was required to advance to the main task. Participants did not receive explicit instructions on which advisor led to which stock, but rather were required to learn these transitions through experience. After this training phase, participants completed 150 trials of the main task, split evenly between the two first-stage states. The response deadline in both stages was 1500 ms and feedback was presented for 1000 ms.

Post-task Evaluations

Following the two-step task, participants responded to a series of self-report items which pertained to participants’ evaluations (or “liking”) of the different Advisors encountered during the two-step task. Participants were presented with the avatar of



each financial advisor, one at a time, and rated how much they liked the advisor using a seven-point scale (from 1 = “Do not like them at all” to 7 = “Like them a lot”). Finally, participants were also asked to estimate how valuable, on average, each of the two stocks were over the course of the learning task (see **Supplementary Material**).

Computational Model

In order to determine the degree to which participants employed model-based and model-free learning, we fit data from the learning phase to a computational model of reinforcement learning used in previous work (Kool et al., 2017). Doing so allowed us to estimate latent variables related to social learning for each subject (Hackel and Amodio, 2018), which we then used as input in our analyses.

The model contains a hybrid of model-free learning and model-based learning for selecting advisors (see **Supplementary Material** for additional details and **Supplementary Table S1** for parameter fits). The model-free system stores values for advisors at the first stage and for stocks at the second stage based on prior reward feedback. The model-based system computes the value of selecting each advisor at the time of choice, combining knowledge about how advisors lead to stocks with the expected payoff of each stock (acquired through model-free learning at the second stage). A model-based learner thus prospectively plans toward a goal: he or she selects an advisor based on the stock the advisor will lead to, in light of the reward expected from each stock. In contrast, a model-free learner selects advisors based on the rewards those advisors have led to in the past.

Critically, the model includes a weighting parameter (w) that indicates the relative influence of model-based and model-free learning in choice, ranging between 0 (purely model-free) and 1 (purely model-based). This parameter can serve as an individual difference measure of the extent to which a participant engaged in model-based or model-free learning. We fit this model for each participant using maximum *a posteriori* (MAP) estimation, with empirical priors used in previous work (Gershman, 2016; Kool et al., 2017). Doing so allowed us to estimate each participant's w parameter (mean = 0.83), indicating the extent to which they relied on model-based vs. model-free learning. We used this parameter in subsequent analyses examining individual differences in the use of these learning strategies.

RESULTS

Model-Free and Model-Based Social Learning

To what extent did participants engage in model-based and model-free social learning? To answer this question, we examined choices in the learning phase, drawing on the following logic of the task. As noted above, the two sets of advisors in the task are equivalent, such that one advisor from each set leads to a particular stock. As a result, a model-based learner would generalize experiences with one set of advisors to the other set. For instance, imagine a participant who sees the first pair of

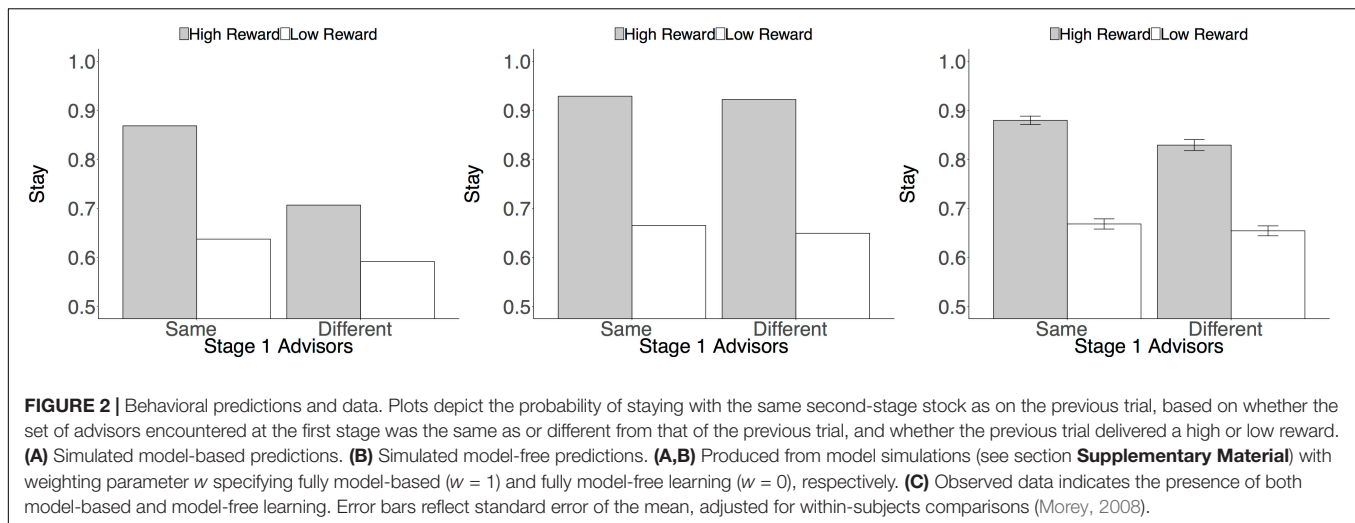
advisors, picks the advisor that leads to the “Axiom” stock, and receives a large reward. On the next round, a model-based learner would try to return to the “Axiom” stock regardless of whether they see the same pair of advisors or a different pair of advisors. In contrast, a model-free learner updates values for individual advisors and chooses advisors based on these values. A model-free learner would therefore repeat their choice on the next trial if presented with the same advisors but would do so to a lesser extent if presented with different advisors. That is, the model-free learner would fail to generalize across sets of advisors.

Drawing on this task logic, we fit learning phase data to a lagged regression model predicting, on a trial-by-trial basis, whether or not participants repeated their most recent choice of Stage 2 stocks (1 = stay, 0 = switch). This analysis provides a model-agnostic way to test the qualitative behavioral predictions of the model-free/model-based account of learning. Following Kool et al. (2016), predictors included the reward earned on the previous trial (standardized, within-subject, to z-scores), whether or not the previous trial started with the same set of advisors (1 = same, -1 = different), and the interaction of these two predictors. A main effect of reward would indicate model-based learning: people return to a high-paying stock regardless of whether they see the same or different advisors on the next trial to get to that stock (simulated data shown in **Figure 2A**). An interaction of reward and start state would indicate model-free learning: people try to return to a high-paying stock, but particularly do so when presented with the same set of advisors, thus repeating the advisor choice that led to the large reward (**Figure 2B**). Models were fit using the lme4 package in R (Bates et al., 2015; R Core Team, 2016). Random variances were allowed for the intercept and all slopes (see **Supplementary Table S2** for all coefficients.).

This analysis revealed a main effect of reward, $b = 1.47$, $SE = 0.07$, $z = 19.80$, $p < 0.001$, consistent with model-based learning: overall, participants returned to second-stage stocks after receiving large rewards. However, the analysis also revealed a Reward \times Start State interaction, $b = 0.22$, $SE = 0.03$, $z = 6.45$, $p < 0.001$, indicating the presence of model-free learning: participants were more likely to return to a high-paying stock when starting with the same advisors at the first stage. Although participants in our sample were highly model-based (mean w parameter in the computational model fits = 0.83), these results support the hypothesis that both model-based and model-free reinforcement learning contributed to social choice (**Figure 2C**).

Post-task Evaluations

If reinforcement learning also gives rise to attitudes, participants might like advisors who can provide reward in the future (model-based value) and advisors associated with past reward (model-free value). To test how learning affects attitudes, we examined participants' self-reported liking of each advisor following the learning task. Using each subject's individual parameter fits in the computational model, we estimated the final model-based and model-free values associated with each advisor for each subject at the end of learning, given the unique series



of stimuli and outcomes viewed by each participant. We then regressed liking ratings simultaneously on each type of value.

Notably, model-based values were identical for advisors who led to the same stock. That is, if the Axiom stock would be expected to deliver 6 points on average at the end of the task, then each advisor who leads to the Axiom stock would have a model-based value of 6 points. If social evaluations reflect model-based learning, participants would therefore like the two advisors who led to the Axiom stock equally. In contrast, model-free values reflect the unique reward history associated with a particular advisor; even for two advisors who led to the Axiom stock, participants might have experienced different reward outcomes with each advisor. If social evaluations reflect model-free learning, people would therefore prefer advisors who provided greater rewards. Finally, this tendency should depend on individual differences in learning, as reflected in the w parameter: individuals who engage in greater model-free learning should especially like advisors associated with high model-free value.

To test these hypotheses, we fit a mixed-effects linear regression predicting post-task liking ratings (**Supplementary Table S3**). Predictors included each participant's final model-free values and model-based values toward each advisor (estimated from the computational model), each participant's w parameter, and the interaction of w with each type of value. Each predictor was standardized to z-scores (within-subject for the value regressors and between-subject for the w parameter). As a result, main effects of value regressors are interpretable relative to the mean level of the w parameter ($w = 0.83$). We included random variances for the intercept and each predictor. The models were fit using the lme4 package and lmerTest packages (Bates et al., 2015; Kuznetsova et al., 2016) in R (R Core Team, 2016).

This analysis yielded a main effect of model-based values, $b = 0.30$, $SE = 0.14$, $t(71.46) = 2.17$, $p = 0.03$, and a marginally significant main effect of model-free values, $b = 0.16$, $SE = 0.09$, $t(162.97) = 1.82$, $p = 0.07$. In other words, at mean levels of the

w parameter, attitudes reflected both kinds of learning: people liked advisors who could lead them to more rewarding stocks and also liked advisors who were uniquely associated with greater reward in the past.

We further examined whether the effects of model-based and model-free learning on reported attitudes varied by participants' individual learning tendencies, as indexed by the w parameter. We found that the w parameter, which represents this individual difference variable, interacted with model-free values, $b = -0.24$, $SE = 0.08$, $t(148.01) = -2.97$, $p = 0.004$. Participants who exhibited relatively greater model-free learning also expressed greater liking of partners who had provided more reward. Simple effects analysis supported this interpretation: for learners relying relatively more on model-free control (centered at the 25th percentile of the w parameter, or $w = 0.70$), model-free values were strongly predictive of attitudes toward advisors, $b = 0.31$, $SE = 0.10$, $t(155.32) = 3.11$, $p = 0.002$, revealing a novel effect of model-free learning on social evaluation. By contrast, for those relying relative more on model-based control (centered at the 75th percentile of the w parameter, or $w = 1$), model-free values were not associated with evaluations, $b = -0.03$, $SE = 0.11$, $t(162.01) = -0.31$, $p = 0.76$. Thus, participants who exhibited model-free learning also liked advisors associated with greater model-free value¹.

Together, these results identify two ways in which reinforcement learning influences social attitudes, one that is goal-directed and one that is habit-like: people like others who are equivalently capable of providing large rewards in the future, and they also like others who have uniquely provided large rewards in the past. Moreover, the influence of past (model-free) reward history depends on individual differences in learning: individuals who weight model-free rewards more strongly during

¹In contrast, we did not observe an interaction between the w parameter and model-based values (see **Supplementary Material**). This finding is consistent with the fact that model-based learning was relatively high across participants, whereas not all participants showed a meaningful degree of model-free learning (i.e., $w < 1$).

learning also have a stronger preference for advisors associated with past rewards.

DISCUSSION

Does habit play a role in social impressions? Our findings demonstrate that, indeed, people form impressions through reward-based reinforcement processes that include model-free learning – a form of learning thought to contribute to habitual behavior. In the sequential learning task used here, participants chose financial advisors based on both model-based and model-free learning. That is, participants chose advisors who could lead them to desirable stocks in the future (model-based) as well as who were associated with high rewards in prior interactions (model-free). Although participants relied far more heavily on model-based (as opposed to model-free learning) in general, this pattern of model-free learning suggests the additional role of a habit-like component of learning and behavior in the context of social impression formation.

Furthermore, participants' learning processes had implications for their explicit social evaluations. Across participants, both model-based and model-free learning predicted self-reported attitudes toward advisors. Moreover, participants varied in their reliance on model-based vs. model-free processing during the learning task, and this individual difference in learning related to differences in evaluation: participants who exhibited greater model-free learning during the investment task showed an effect of model-free learning on self-reported attitudes. Thus, these findings dissociate two routes through which reinforcement learning contributes to attitudes toward social partners, and they highlight the importance of considering individual differences in learning strategies during social interactions to understand the effects of rewards on social attitudes and decisions.

Model-Based and Model-Free Social Cognition

Our central finding – of model-free learning in social impression formation – offers novel theoretical implications for social cognition, learning, and attitudes. First, our findings highlight a role for reward-based reinforcement learning in social interactions. Previous impression formation research demonstrates that people learn about the traits of others in order to predict how others will behave (Heider, 1958). For instance, by observing financial advisors, people can form impressions of an advisor's competence and predict that advisor's future performance (Boorman et al., 2013; Leong and Zaki, 2018). Our results introduce a complementary mode of social learning based on reward: people also learn whom to choose and whom to like through instrumental learning, such as directly choosing an advisor and experiencing rewards as a result.

The observation of model-free social learning, in particular, supports the proposed role of habit in social cognition. In model-free learning, people repeat previously-rewarded choices in a relatively inflexible manner – the hallmark of a habit. Habits may therefore influence social behavior: because habits reflect

routinized responses that operate most adaptively in invariable environments, they may fill in the gaps between goal-directed responses to facilitate social behavior. In some cases, habits may have harmful effects; for example, people may persist in interacting with social partners with whom they had positive past experiences, even when other partners might be equally or more relevant to one's current goals. In other cases, habits may be beneficial, leading an individual to approach a previously-rewarding person while distracted by their pursuit of an unrelated goal – perhaps eliciting help, if needed, or simply avoiding a social faux pas. In both cases, their effects may be subtle, relative to goal-directed responses, yet still crucial to adaptive social function.

Although model-based and model-free learning offer different benefits and costs, their concerted function may promote successful social interactions. Social life offers a wealth of information about other people – their traits, preferences, and emotions – which lets us know whom to interact with and how to interact with them. Through experience, we learn which members of our social networks to turn to for empathy as opposed to fun (Morelli et al., 2017) and which verbal or facial cues predict different emotions for close others (Zaki et al., 2016). Model-free learning offers a computationally simple way to learn how to act around others given this wealth of information, requiring little deliberation (Otto et al., 2013). Yet, at the same time, model-free learning is relatively inflexible, leaving people unable to adapt as contingencies change or to plan ahead in novel settings. By comparison, model-based learning requires greater effort but allows people to adapt to new contingencies and make novel plans – for instance, choosing a gift for another person for the first time given knowledge about their preferences. Both types of learning are functional, with tradeoffs that depend on the particulars of a situation, and thus an important goal of future research will be to explore how these tradeoffs are managed and prioritized across situations.

It is notable that participants' behavior was highly model-based in our study, on average – more so than in past work using this task (Kool et al., 2017; see also Da Silva and Hare, 2019). It is possible that the social framing of the task made it easier for people to reason in a model-based manner, much as people find it easier to reason about social relations than non-social relations (Cosmides, 1989; Mason et al., 2010). Moreover, our instructions framed rewards in terms of stock performance, which offers a familiar and intuitive explanation for drifting outcomes. While it is possible that these features made our instructions clearer relative to past work (Da Silva and Hare, 2019), the familiarity of concepts used in our task framing may have facilitated model-based choices – an interesting possibility for future research.

Finally, and more broadly, this work sheds light on how multiple forms of learning and memory can contribute to social cognition. Based on research in cognitive neuroscience (Squire, 2004; Henke, 2010), Amodio (2019; see also Amodio and Ratner, 2011) theorized that social cognition comprises multiple distinct and interactive learning and memory systems, including habits. Although classic work in social psychology has focused primarily on the roles of conceptual associations and Pavlovian forms

of learning, research has just recently begun to probe the role of reward-based forms of learning in social cognition (Hackel et al., 2015; Lindström and Tobler, 2018). To date, these studies have not distinguished between types of computations that may underlie instrumental learning from rewards. Here, by using a two-step learning task to examine social learning, we were able to dissociate model-based and model-free forms of reward learning and, in doing so, provide new evidence for the role of multiple learning systems, functioning in concert, in social cognition.

Potential Limitations

The goal of this research was to examine learning processes that give rise to habitual behavior. However, there remain open questions about the extent to which model-free learning, as assessed in sequential decision-making (i.e., two-step) tasks, corresponds to traditional definitions of habit. First, questions have been raised as to whether additional strategies may contribute to observed effects of model-free learning in sequential decision tasks (Dezfouli and Balleine, 2012; Da Silva and Hare, 2019; but see Morris and Cushman, 2019), just as other representations may contribute to observed effects of model-based learning (Momennejad et al., 2017; Russek et al., 2017).

Although our task was designed to examine two specific learning processes, it is useful to consider the possibility of alternative ways of representing the task and outcomes that might yield different inferences. For instance, if participants grouped the two “Axiom” advisors under one abstract action representation of “pro-Axiom-choice” (possibly through model-based processes), then putative patterns of model-based learning might actually reflect model-free learning over such groupings; conversely, if participants represented four end states in the task – acting as if there were two distinct Axiom stocks and two distinct Zephyr stocks depending on the advisor chosen – then putative patterns of model-free learning could reflect model-based learning. However, we believe such a four-state task representation is unlikely, given that the instructions and visual display emphasized that there were two end states, each reached from two advisors. For a participant to use a 4-state task representation, they would have to ignore this information and the actual transition structure of the task, associating end-states with actions used to get there, in which case it may not be obvious that this would still be a model-based controller (see Morris and Cushman, 2019, for related discussion). Future work could test whether people generate unexpected task representations and whether these contribute to learning.

More broadly, people may use learning and choice strategies not encapsulated by our task and analyses, moving beyond the two approaches studied here (see **Supplementary Material** for further discussion). For instance, in other settings, people might choose individual advisors based on trait impressions (Hackel et al., 2015) or might learn specific motor actions (Shahar et al., 2019) – such as pushing a particular button or walking toward a colleague’s office – in addition to learning the value of a social partners. Although our experiment did not address these broader theoretical questions regarding

model-based and model-free learning accounts, future research on reinforcement learning in social cognition will benefit from advances in our understanding of these processes as they develop.

Second, there is some debate on whether – and to what extent – model-free learning maps on to traditional definitions of habitual control (Miller et al., 2019; see also Gillan et al., 2015; Sjoerds et al., 2016). Miller et al. (2019) argue that traditional conceptualizations of habits reflect stimulus-response associations devoid of expected value representations (i.e., are *value-free*), whereas model-free algorithms still depend on the expected value representations associated with a learner’s available actions (i.e., are *value-based*). In this view, habits form directly through action repetition within a given context, regardless of reward outcomes. It is possible that both model-free RL and action repetition contribute to behaviors commonly considered habitual (Pauli et al., 2018). These processes might align with a theoretical distinction between “direct” cuing of habits, in which responses are directly associated with context cues, and “motivated” cuing of habits, in which responses depend on the motivation linked to a behavior through past rewards (Wood and Neal, 2007). To complement and extend our findings, future work could consider these varied approaches.

New Questions About Habits in Social Behavior

Our use of the two-step task to probe the role of habits in social cognition raises several new questions regarding other aspects of habits in social life. For instance, a classic marker of a habit is its persistence even when it no longer fulfills a valued goal (Wood and R  nger, 2016). Past work suggests that reward feedback in social interaction can have such a persistent impact (Hackel et al., 2015). Future work should consider tasks traditionally employed to test for this kind of habitual persistence, such as the slips-of-action paradigm (e.g., Gillan et al., 2011; de Wit et al., 2012) or outcome devaluation/revaluation procedures (e.g., Valentin et al., 2007; de Wit et al., 2009; Tricomi et al., 2009; see Foerde, 2018, for review).

Our findings raise further questions regarding the specificity of habits in social impressions, relationships, and behaviors. For example, do people form habits to interact with specific partners in specific contexts? Or do they form habits to approach or avoid social interaction in general? Are there benefits to forming such social habits? Answering these questions promises to illuminate the structure of people’s social lives, much as advances in habit research sheds light on how habits can promote healthy eating, exercising, or studying (Galla and Duckworth, 2015; Lin et al., 2016).

Finally, the implications of our findings extend to other areas of research within social psychology, such as intergroup relations, complementing recent work suggesting that model-free learning may underlie implicit attitudes toward social groups (Kurdi et al., 2019). The concept of habit has previously been invoked in prior theories of social attitudes, such as to describe the phenomenon of implicit prejudice and the

difficulty people have in ridding themselves of it (e.g., “breaking the prejudice habit,” Devine, 1989; Devine et al., 2012). However, this usage has been largely colloquial or metaphorical, as previous research has not used methods capable of assessing habit-like patterns of preference and choice. Our findings suggest that social experiences may indeed give rise to a form of habit, but these are rooted more directly in reward-based action tendencies than in conceptual processes such as stereotypes.

Nevertheless, if some aspects of prejudice are truly habit-like, then they may be extraordinarily difficult to control or eradicate. As such, interventions involving the replacement of a biased thought or action with an egalitarian response (Devine, 1989) or changes in the situational affordances for bias expression (Amodio and Swencionis, 2018) should be more effective than methods for unlearning bias (Lai et al., 2014). Furthermore, an intervention aimed at “unlearning” a habit-like response would require action-based interventions, in contrast to conventional interventions aimed at modifying a person’s beliefs and values. As our conceptualization of habits in social cognition develops, it may begin to elucidate psychological processes in other domains as well.

CONCLUSION

Habits are integral to everyday human behavior, and they may also support our social behaviors. Our findings represent an initial demonstration that habit-like learning processes are also involved in the formation of social preferences and attitudes. These findings expand our understanding of how learning and memory systems support social cognition and provide a foundation for new research on the role of habit in social learning.

REFERENCES

- Amodio, D. M. (2019). Social Cognition 2.0: an interactive memory systems account. *Trends Cogn. Sci.* 23, 21–33. doi: 10.1016/j.tics.2018.10.002
- Amodio, D. M., and Berg, J. J. (2018). Toward a multiple memory systems model of attitudes and social cognition. *Psychol. Inq.* 29, 14–19. doi: 10.1080/1047840x.2018.1435620
- Amodio, D. M., and Ratner, K. G. (2011). A memory systems model of implicit social cognition. *Curr. Dir. Psychol. Sci.* 20, 143–148. doi: 10.1177/0963721411408562
- Amodio, D. M., and Swencionis, J. K. (2018). Proactive control of implicit bias: a theoretical model and implications for behavior change. *J. Personal. Soc. Psychol.* 115, 255–275. doi: 10.1037/pspi0000128
- Asch, S. E. (1946). Forming impressions of personality. *J. Abnorm. Soc. Psychol.* 41, 258–290.
- Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419. doi: 10.1016/s0028-3908(98)00033-1
- Bargh, J. A., and Ferguson, M. J. (2000). Beyond behaviorism: on the automaticity of higher mental processes. *Psychol. Bull.* 126, 925–945. doi: 10.1037//0033-2909.126.6.925
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.
- Boorman, E. D., O’Doherty, J. P., Adolphs, R., and Rangel, A. (2013). The behavioral and neural mechanisms underlying the tracking of expertise. *Neuron* 80, 1558–1571. doi: 10.1016/j.neuron.2013.10.024
- Brewer, M. B. (1988). “A dual process model of impression formation,” in *Advances in Social Cognition*, Vol. 1, eds T. K. Srull, and R. S. Wyer, Jr. (Hillsdale, NJ: Lawrence Erlbaum Associates), 1–36.
- Carlston, D. E., and Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *J. Personal. Soc. Psychol.* 66, 840–856. doi: 10.1037//0022-3514.66.5.840
- Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31, 187–276. doi: 10.1016/0010-0277(89)90023-1
- Da Silva, C. F., and Hare, T. A. (2019). Humans are primarily model-based and not model-free learners in the two-stage task. *BioRxiv*. [preprint]. doi: 10.1101/682922
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans’ choices and striatal prediction errors. *Neuron* 69, 1204–1215. doi: 10.1016/j.neuron.2011.02.027
- De Houwer, J., Hendrickx, H., and Baeyens, F. (1997). Evaluative learning with “subliminally” presented stimuli. *Conscious. Cogn.* 6, 87–107. doi: 10.1006/ccog.1996.0281
- de Wit, S., Corlett, P. R., Aitken, M. R., Dickinson, A., and Fletcher, P. C. (2009). Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *J. Neurosci.* 29, 11330–11338. doi: 10.1523/JNEUROSCI.1639-09.2009
- de Wit, S., Watson, P., Harsay, H. A., Cohen, M. X., van de Vijver, I., and Ridderinkhof, K. R. (2012). Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *J. Neurosci.* 32, 12066–12075. doi: 10.1523/JNEUROSCI.1088-12.2012

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by New York University Committee on Activities Involving Human Subjects. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors developed the theoretical ideas, questions, and approach. LH and JB designed the task. JB collected the data. LH and BL analyzed the data. LH and DA drafted the manuscript, with input and edits by JB and BL.

FUNDING

This research was supported by a grant to DA (VICI 016.185.058) from the Netherlands Organization for Scientific Research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02592/full#supplementary-material>

- Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. *J. Personal. Soc. Psychol.* 56, 5–18. doi: 10.1037//0022-3514.56.1.5
- Devine, P. G., Forscher, P. S., Austin, A. J., and Cox, W. T. (2012). Long-term reduction in implicit race bias: a prejudice habit-breaking intervention. *J. Exp. Soc. Psychol.* 48, 1267–1278. doi: 10.1016/j.jesp.2012.06.003
- Dezfouli, A., and Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* 35, 1036–1051. doi: 10.1111/j.1460-9568.2012.08050.x
- Dickinson, A., and Balleine, B. (1994). Motivational control of goal-directed action. *Anim. Learn. Behav.* 22, 1–18. doi: 10.3758/bf03199951
- Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., and Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nat. Neurosci.* 18, 767–772. doi: 10.1038/nn.3981
- Fiske, S. T., and Neuberg, S. L. (1990). “A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation,” in *Advances in Experimental Social Psychology*, Vol. 23, ed. M. P. Zanna, (New York, NY: Academic Press), 1–74. doi: 10.1016/s0065-2601(08)60317-2
- Fitzsimons, G. M., and Anderson, J. (2013). “Interpersonal cognition: seeking, understanding, and maintaining relationships,” in *Handbook of Social Cognition*, ed. D. Carlston, (New York, NY: Oxford University Press), 590–615.
- Foerde, K. (2018). What are habits and do they depend on the striatum? A view from the study of neuropsychological populations. *Curr. Opin. Behav. Sci.* 20, 17–24. doi: 10.1016/j.cobeha.2017.08.011
- Galla, B. M., and Duckworth, A. L. (2015). More than resisting temptation: beneficial habits mediate the relationship between self-control and positive life outcomes. *J. Personal. Soc. Psychol.* 109, 508–525. doi: 10.1037/pspp0000026
- Gershman, S. J. (2016). Empirical priors for reinforcement learning models. *J. Math. Psychol.* 71, 1–6. doi: 10.1016/j.jmp.2016.01.006
- Gillan, C. M., Otto, A. R., Phelps, E. A., and Daw, N. D. (2015). Model-based learning protects against forming habits. *Cogn. Affect. Behav. Neurosci.* 15, 523–536. doi: 10.3758/s13415-015-0347-6
- Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., et al. (2011). Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am. J. Psychiatry* 168, 718–726. doi: 10.1176/appi.ajp.2011.10071062
- Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., et al. (2016). psiTurk: an open-source framework for conducting replicable behavioral experiments online. *Behav. Res. Methods* 48, 829–842. doi: 10.3758/s13428-015-0642-8
- Hackel, L. M., and Amodio, D. M. (2018). Computational neuroscience approaches to social cognition. *Curr. Opin. Psychol.* 24, 92–97. doi: 10.1016/j.copsyc.2018.09.001
- Hackel, L. M., Doll, B. B., and Amodio, D. M. (2015). Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. *Nat. Neurosci.* 18, 1233–1235. doi: 10.1038/nn.4080
- Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York, NY: Wiley.
- Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nat. Rev. Neurosci.* 11, 523–532. doi: 10.1038/nrn2850
- Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., and Crombez, G. (2010). Evaluative conditioning in humans: a meta-analysis. *Psychol. Bull.* 136, 390–421. doi: 10.1037/a0018916
- Jones, E. E., and Davis, K. E. (1965). “From acts to dispositions: The attribution process in person perception,” in *Advances in Experimental Social Psychology*, Vol. 2, ed. L. Berkowitz, (New York, NY: Academic Press), 219–266. doi: 10.1016/s0065-2601(08)60107-0
- Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., et al. (2011). Behavioral and neural properties of social reinforcement learning. *J. Neurosci.* 31, 13039–13045.
- Kool, W., Cushman, F. A., and Gershman, S. J. (2016). When does model-based control pay off? *PLoS Comput. Biol.* 12:e1005090. doi: 10.1371/journal.pcbi.1005090
- Kool, W., Gershman, S. J., and Cushman, F. A. (2017). Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychol. Sci.* 28, 1321–1333. doi: 10.1177/0956797617708288
- Kurdi, B., Gershman, S. J., and Banaji, M. R. (2019). Model-free and model-based learning processes in the updating of explicit and implicit evaluations. *Proc. Natl. Acad. Sci. U.S.A.* 116, 6035–6044. doi: 10.1073/pnas.1820238116
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2016). *lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models (Lmer Objects of lme4 Package)*. R package (Version 2.0–32).
- Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E. L., Joy-Gaba, J. A., et al. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *J. Exp. Psychol.: Gen.* 143, 1765–1785. doi: 10.1037/a0036260
- Leong, Y. C., and Zaki, J. (2018). Unrealistic optimism in advice taking: a computational account. *J. Exp. Psychol.* 147, 170–189. doi: 10.1037/xge0000382
- Lin, A., Adolphs, R., and Rangel, A. (2011). Social and monetary reward learning engage overlapping neural substrates. *Soc. Cogn. Affect. Neurosci.* 7, 274–281. doi: 10.1093/scan/nsr006
- Lin, P. Y., Wood, W., and Monterosso, J. (2016). Healthy eating habits protect against temptations. *Appetite* 103, 432–440. doi: 10.1016/j.appet.2015.11.011
- Lindström, B., Selbing, I., Molapour, T., and Olsson, A. (2014). Racial bias shapes social reinforcement learning. *Psychol. Sci.* 25, 711–719. doi: 10.1177/0956797613514093
- Lindström, B., and Tobler, P. N. (2018). Incidental ostracism emerges from simple learning mechanisms. *Nat. Hum. Behav.* 2, 405–414. doi: 10.1038/s41562-018-0355-y
- Mason, M. F., Magee, J. C., Kuwabara, K., and Nind, L. (2010). Specialization in relational reasoning: the efficiency, accuracy, and neural substrates of social versus nonsocial inferences. *Soc. Psychol. Personal. Sci.* 1, 318–326. doi: 10.1177/1948550610366166
- Miller, K. J., Shenav, A., and Ludvig, E. A. (2019). Habits without values. *Psychol. Rev.* 126, 292–311. doi: 10.1037/rev0000120
- Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., and Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nat. Hum. Behav.* 1, 680–692. doi: 10.1038/s41562-017-0180-8
- Morelli, S. A., Ong, D. C., Makati, R., Jackson, M. O., and Zaki, J. (2017). Empathy and well-being correlate with centrality in different social networks. *Proc. Natl. Acad. Sci. U.S.A.* 114, 9843–9847. doi: 10.1073/pnas.1702155114
- Morey, R. D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutor. Quant. Methods Psychol.* 4, 61–64. doi: 10.20982/tqmp.04.2.p061
- Morris, A., and Cushman, F. A. (2019). Model-free RL or action sequences? *PsyArXiv*. [preprint].doi: 10.31234/osf.io/k67tm
- Moskowitz, G. B., and Roman, R. J. (1992). Spontaneous trait inferences as self-generated primes: implications for conscious social judgment. *J. Personal. Soc. Psychol.* 62, 728–738. doi: 10.1037//0022-3514.62.5.728
- Olson, M. A., and Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personal. Soc. Psychol. Bull.* 32, 421–433. doi: 10.1177/0146167205284004
- Otto, A. R., Gershman, S. J., Markman, A. B., and Daw, N. D. (2013). The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol. Sci.* 24, 751–761. doi: 10.1177/0956797612463080
- Pauli, W. M., Cockburn, J., Pool, E. R., Pérez, O. D., and O’Doherty, J. P. (2018). Computational approaches to habits in a model-free world. *Curr. Opin. Behav. Sci.* 20, 104–109.
- R Core Team. (2016). *R: A Language and Environment for Statistical Computing (Version 3.3.1)*. Vienna: R Core Team.
- Robbins, T. W., and Costa, R. M. (2017). Habits. *Curr. Biol.* 27, R1200–R1206. doi: 10.1016/j.cub.2017.09.060
- Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., and Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Comput. Biol.* 13:e1005768. doi: 10.1371/journal.pcbi.1005768
- Rydell, R. J., and McConnell, A. R. (2006). Understanding implicit and explicit attitude change: a systems of reasoning analysis. *J. Personal. Soc. Psychol.* 91, 995–1008. doi: 10.1037/0022-3514.91.6.995
- Shahar, N., Moran, R., Hauser, T. U., Kievit, R. A., McNamee, D., Moutoussis, M., et al. (2019). Credit assignment to state-independent task representations and its relationship with model-based decision making. *Proc. Natl. Acad. Sci. U.S.A.* 116, 15871–15876. doi: 10.1073/pnas.1821647116
- Sjoerds, Z., Dietrich, A., Deserno, L., De Wit, S., Villringer, A., Heinze, H. J., et al. (2016). Slips of action and sequential decisions: a cross-validation study of tasks assessing habitual and goal-directed action control. *Front. Behav. Neurosci.* 10:234. doi: 10.3389/fnbeh.2016.00234

- Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem.* 82, 171–177. doi: 10.1016/j.nlm.2004.06.005
- Sutton, R. S., and Barto, A. G. (1998). *Introduction to Reinforcement Learning*. Cambridge, MA: MIT Press.
- Sweldens, S., Corneille, O., and Yzerbyt, V. (2014). The role of awareness in attitude formation through evaluative conditioning. *Personal. Soc. Psychol. Rev.* 18, 187–209. doi: 10.1177/1088868314527832
- Tamir, D. I., and Thornton, M. A. (2018). Modeling the predictive social mind. *Trends Cogn. Sci.* 22, 201–212. doi: 10.1016/j.tics.2017.12.005
- Thorndike, E. (1911). *Animal Intelligence*. New York, NY: Hafner.
- Tricomi, E., Balleine, B. W., and O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29, 2225–2232. doi: 10.1111/j.1460-9568.2009.06796.x
- Uleman, J. S., and Kressel, L. M. (2013). “A brief history of theory and research on impression formation,” in *Oxford Handbook of Social Cognition*, ed. D. E. Carlston, (New York, NY: Oxford University Press), 53–73.
- Valentin, V. V., Dickinson, A., and O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.* 27, 4019–4026. doi: 10.1523/jneurosci.0564-07.2007
- Walther, E. (2002). Guilty by mere association: evaluative conditioning and the spreading attitude effect. *J. Personal. Soc. Psychol.* 82, 919–934. doi: 10.1037/0022-3514.82.6.919
- Winter, L., and Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *J. Personal. Soc. Psychol.* 47, 237–252. doi: 10.1037//0022-3514.47.2.237
- Wood, W. (2017). Habit in personality and social psychology. *Personal. Soc. Psychol. Rev.* 21, 389–403. doi: 10.1177/1088868317720362
- Wood, W., and Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychol. Rev.* 114, 843–863. doi: 10.1037/0033-295x.114.4.843
- Wood, W., and Rünger, D. (2016). Psychology of habit. *Annu. Rev. Psychol.* 67, 289–314.
- Wyer, R. S. Jr., and Carlston, D. E. (1979). *Social Cognition, Inference, and Attribution*. Hillsdale, NJ: Erlbaum Publishers.
- Zaki, J., Kallman, S., Wimmer, G. E., Ochsner, K., and Shohamy, D. (2016). Social cognition as reinforcement learning: feedback modulates emotion inference. *J. Cogn. Neurosci.* 28, 1270–1282. doi: 10.1162/jocn_a_00978

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Hackel, Berg, Lindström and Amodio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



On How Definitions of Habits Can Complicate Habit Research

Jan De Houwer*

Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

OPEN ACCESS

Edited by:

Wendy Wood,
University of Southern California,
United States

Reviewed by:

L. Alison Phillips,
Iowa State University, United States
David M. Amodio,
New York University, United States
Idit Shalev,
Ariel University, Israel

*Correspondence:

Jan De Houwer
jan.dehouwer@ugent.be

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 14 February 2019

Accepted: 08 November 2019

Published: 29 November 2019

Citation:

De Houwer J (2019) On How
Definitions of Habits Can
Complicate Habit Research.
Front. Psychol. 10:2642.
doi: 10.3389/fpsyg.2019.02642

The core message of this paper is that many of the challenges of habit research can be traced back to the presence of causal elements within the definition of habits. For instance, the idea that habits are stimulus-driven implies that habitual behavior is not causally mediated by goal-representations. The presence of these causal elements in the definition of habits leads to difficulties in establishing empirically whether behavior is habitual. Some of these elements can also impoverish theoretical thinking about the mechanisms underlying habitual behavior. I argue that habit research would benefit from eliminating any reference to specific S-R association formation theories from the definition of habits. Which causal elements are retained in the definition of habits depends on the goals of researchers. However, regardless of the definition that is selected, it is good to be aware of the implications of the definition of habits for empirical and theoretical research on habits.

Keywords: habits, automaticity, conceptual analysis, learning, goal-directed behavior

When asked to explain their behavior, lay people often refer to habits. Likewise, when making resolutions for the future, they often express a wish to install new habits or to change old ones. The concept “habit” is popular not only with lay people but also engages academic psychologists (see Wood and Rünger, 2016, for a review). As noted by Gardner (2015), one important difference in the way habits are conceptualized by lay people versus academics is that the former focus on observable aspects of behavior (e.g., the frequency with which a behavior is emitted) whereas the latter focus on the (mental) causes of behavior (e.g., the fact that the behavior is triggered by cues in the environment without being directed at goals; see Wood and Neal, 2007, for a discussion of the interface between habits and goals).

Although the focus on explanation is an undeniable strength of the academic approach to habits, in this paper, I draw attention the downsides of incorporating assumptions about causes into definitions of habits and habitual behavior. In the section “The Conceptual Level: Defining Habits and Habitual Behavior,” I briefly consider some of the definitions of habits that have been put forward by lay people and academics. These definitions have in common that they

have implications for the criteria that are used to distinguish empirically between habitual and non-habitual behavior. In the section “The Empirical Level: Establishing the Presence of Habitual Behavior,” I discuss problems with empirically verifying the causal criteria put forward in the scientific literature on habits. The section “The Theoretical Level: Explaining Habitual Behavior” focuses on the constraints in theorizing that follow from definitions of habits that refer to S-R associations (i.e., links between stimulus and response representations *via* which activation can spread). Finally, I discuss the possible merits of removing causal assumptions from the definition of habits. Many of the challenges that are addressed in this paper have been discussed before by others (e.g., Watson and de Wit, 2018). The current paper aims to go beyond those past contributions by highlighting how these challenges relate to the causal nature of scientific definitions of habits. Based on this insight, new ways of tackling these challenges can be considered.

THE CONCEPTUAL LEVEL: DEFINING HABITS AND HABITUAL BEHAVIOR

The first challenge for any area of research is to reach some level of clarity about and consensus on what is being studied (i.e., what constitutes the explanandum). To the extent that definitions of a research topic diverge, scientific progress is bound to be hampered by misunderstandings and false debates. Although the definition of a concept can change over time and general agreement about definitions is rare in psychological science, there is merit in trying to improve clarity at the conceptual level, if only by creating awareness of the various definitions that have been proposed and the way in which they are related (Machado and Silva, 2007). In this section, I will first consider the different ways in which habits have been defined. This allows me to then highlight the causal nature of those definitions and the implications this has for habit research.

The recent paper of Gardner (2015) provides an excellent starting point for considering the range of definitions of habits that have been proposed. Lay definitions are mainly descriptive, referring to habits as behaviors that are emitted frequently or in a persistent, automatic manner. Scientific definitions of habits, on the other hand, contain explanatory elements. Some of these scientific definitions also refer to habits as (frequent, persistent, or automatic) behaviors but in addition those behaviors are said to have particular causes. These causes can refer to past experience, such as the idea that habits are the result of the repetition of behavior, and/or to underlying mental processes, such as the idea that habits result from the activation of S-R associations without the involvement of goals (i.e., representations of desired end states; see Gardner, 2015, for an overview). Many of these definitions imply that habitual behavior is stimulus-driven, that is, dependent on cues in the current context that trigger the behavior without considerations of the

current outcomes of the behavior¹. Other scientific definitions do not refer to habits as a behavior but as a mental cause underlying behavior. For instance, habits have been defined as behavioral impulses that are instigated by S-R associations or as the S-R associations themselves. Many definitions, however, refer to several of these components. To illustrate, Gardner et al. (2011, p. 175) define habits as “behavioural patterns learned through context dependent repetition: repeated performance in unvarying settings reinforces context-behaviour associations such that, subsequently, encountering the context is sufficient to automatically cue the habitual response.” Wood and Neal (2009, p. 580) define habits as “A type of automaticity characterized by a rigid contextual cuing of behavior that does not depend on people’s goals and intentions. Habits develop as people respond repeatedly in a stable context and thereby form direct associations in memory between that response and cues in the performance context.”

Gardner (2015) already highlighted the fundamental difference between habits as a type of behavior and habits as a underlying determinant of behavior (e.g., an impulse or S-R association). To reduce confusion, in this paper, I will use the term “habitual behavior” to refer to habits as a type of behavior. Importantly, all definitions of habits put forward criteria for distinguishing between habitual and non-habitual behavior. These criteria can

¹Note that the concept of stimulus-driven behavior does not overlap with the concept of respondent behavior that is often used by functional researchers (see Skinner, 1953). Like stimulus-driven behavior, respondent behavior is under the control of stimuli in the environment. However, unlike stimulus-driven behavior, behavior can be called respondent only if it was never before under the control of its consequences. This is an important distinction because it is typically assumed that many stimulus-driven behaviors are originally goal-directed but become stimulus-driven only as the result of the frequent execution of the behavior. Hence, most stimulus-driven behaviors do not qualify as respondent behavior. The distinction between respondent and stimulus-driven behavior is related to the fact that functional psychology focuses on functional causation (A is a function of B) whereas cognitive psychology focuses on mechanistic causation (A triggers B; see Chiesa, 1992, for an excellent discussion). Functional causation does not require contiguous causes (i.e., events in the here and now that put behavior in motion, much like one cogwheel can put another cogwheel in motion) but allows for causes that are present in the past. Hence, if the presence of a behavior in the past has been a function of its consequences (i.e., it was an operant behavior) and if its current presence is a function of its presence in the past (i.e., it is more likely now because it was repeatedly emitted in the past), then the current presence of the behavior is a function of the consequences of the behavior in the past, which is why also the current behavior would qualify as an operant behavior. The concept of stimulus-driven behavior, on the other hand, only takes into account contiguous causes and thus only entities that are present immediately before the behavior is initiated. For cognitive psychologists, these contiguous causes can be events in the current physical environment but also representations at the mental level. A behavior qualifies as stimulus-driven if the only contiguous cause of the behavior is (the representation of) a stimulus in the environment without the involvement of representations of goals. In sum, whereas the concept of respondent behavior is inherently functional in nature, the concept of stimulus-driven behavior is inherently mental in that it refers to the (absence of a) mechanistic causal impact of goal representations (see De Houwer, 2011; Hughes et al., 2016, for a discussion of the relation between functional and cognitive psychology). Within functional psychology, one could in principle study how the frequency of reinforcement in the past changes the moderators of behavior in the present (e.g., Barnes-Holmes et al., 2017).

refer to more or less observable characteristics of behavior (e.g., frequency, persistence, automaticity); to assumptions about the experiences that cause this behavior (e.g., repetition of a behavior in a context); and/or to assumptions about the mental processes and representations that cause the behavior (e.g., the activation of an impulse *via* the operation of an S-R association).

In this paper, I focus on the implications of the criteria that habit researchers use to distinguish habitual from non-habitual behavior. Although Gardner (2015) correctly points out that scientists should move beyond mere description of behavior and consider the causes of behavior, there are downsides to incorporating causal elements within scientific definitions of to-be-explained phenomena. First, it can hamper attempts to verify empirically whether the phenomenon is present (i.e., to determine whether a behavior qualifies as habitual), which leads to difficulties in studying the phenomenon. Causality can never be observed directly but must always be inferred from observable events. This problem is exacerbated when the causes themselves are unobservable, as is the case with many mental processes and representations (e.g., S-R associations in memory). Second, defining phenomena in terms of their causes confounds the explanandum (that which needs to be explained) with the explanans (that by which the explanandum is explained; Hempel, 1970). In other words, it implies *a priori* assumptions about the causes of the phenomenon. This is less problematic when those *a priori* assumptions turn out to be justified. However, if those assumptions are incorrect, then research based on this definition does not necessarily inform us about the phenomenon, thus hampering the cumulative nature of research. Moreover, an *a priori* commitment to certain causes of a phenomenon may prevent researchers from considering the role of other potential causes of the phenomenon, thereby reducing theoretical diversity and ultimately hampering theoretical progress.

In the remainder of this paper, I discuss these challenges at the empirical and theoretical level, as well as possible ways to deal with those challenges. Rather than providing a systematic review of the literature in order to assess the exact extent to which problems at the empirical and theoretical level arise in habit research, I will focus on developing the conceptual argument and will merely provide examples of the problems that can arise. The examples that I provide come from behavioral research on habits in humans. The conceptual issues that I address also apply to neuroscientific research on habits in humans but this research will not be covered in this paper.

THE EMPIRICAL LEVEL: ESTABLISHING THE PRESENCE OF HABITUAL BEHAVIOR

For many psychologists, the defining characteristic of habits is that they are stimulus-driven (Gardner, 2015; Wood and R  nger, 2016). This idea introduces several causal assumptions

within the definition of habitual behavior. In the following paragraphs, I will highlight these causal assumptions, as well as the challenges they create for establishing that behavior is habitual in the sense of stimulus-driven.

On the one hand, the concept of stimulus-driven behavior implies that habitual behavior is caused directly by stimuli in the environment. Although the causal impact of stimuli on behavior cannot be observed directly, it is relatively easy to infer the environmental causes of behavior by manipulating the presence of stimuli and examining how this influences the presence of the behavior. If the behavior is present when a certain stimulus is present in the environment but absent when that stimulus is absent, this provides strong grounds for arguing that the stimulus is causally related to the occurrence of the behavior.

On the other hand, in the context of habit research, “stimulus-driven” not only implies that a stimulus is causally related to the behavior but also that the behavior is not a function of its anticipated consequences. Put differently, stimulus-driven behavior is not directed at goals (Adams, 1982; Heyes and Dickinson, 1990; see Moors et al., 2017, for a detailed analysis of what it means to say that behavior is goal-directed). Hence, establishing that a behavior is habitual requires arguments for the conclusion that the behavior is not directed at goals². There are, however, several reasons why it is not easy to convincingly demonstrate that behavior is not directed at goals. First, goal representations are mental entities that cannot be observed directly by researchers (and, in the case of unconscious goals, also not by the person who possesses the goal). Second, whether these entities have a causal impact can also not be observed directly because causality always needs to be inferred from observations. Third, verifying the absence of causal impact of mental entities is even more difficult to achieve than verifying the presence of these entities and their causal impact on behavior.

Habit researchers have tried to circumvent the first two problems by using behavioral proxies of the causal impact of goal representations on behavior. They reasoned that if a goal causally mediates a behavior, then changing the goal or its relation to the behavior should also change the behavior. For instance, in order to establish that lever pressing is mediated by the goal to eat a specific food, one could reduce the goal to eat that food by making it aversive (i.e., devaluation test) or by no longer delivering the food after a lever press (i.e., contingency degradation test). From a cognitive point of view, it is indeed relatively safe to conclude that the behavior is mediated by a particular goal representation if those interventions change behavior (e.g., Adams, 1982; Heyes and Dickinson, 1990).

Whereas this strategy might circumvent the first two problems that were noted above, it does not solve the third problem.

²If one interprets “stimulus-driven” in a strict manner as indicating that the behavior is a function solely of the stimulus, then demonstrating the stimulus-driven nature of behavior would also require evidence that the behavior does not depend on any enabling conditions, such as the availability of sufficient attentional resources (Bargh, 1989). In this paper, however, I focus only on the assumption that goals are not causally involved in stimulus-driven behavior.

More specifically, behavior may be mediated by goals even if an effect of devaluation and contingency degradation is not found (Heyes and Dickinson, 1990; Thraill and Bouton, 2015; Moors et al., 2017). It is indeed possible that the intervention was not strong enough (e.g., it did not fully eliminate the palatability of the food), that statistical power was insufficient for establishing the presence of goals and their impact on behavior (see Vadillo et al., 2019), or that the intervention targeted another goal than the one that actually mediates behavior (De Houwer et al., 2018).

With regard to the latter point, De Houwer et al. (2018) examined one of the most widely used paradigms in research on habits in humans, namely the fabulous fruit game (e.g., de Wit et al., 2007). Without going into detail, in this task, participants repeatedly press keys in order to generate images of fruits, some of which are worth points. During an outcome devaluation phase, the value of some of the fruits is reduced (i.e., they are no longer worth points). Habits are typically inferred from the lack of impact of fruit-devaluation on key presses. However, the data reported by De Houwer et al. support the idea that these seemingly habitual key presses are still directed at the goal of obtaining points. For instance, changing the value of points did influence responding even when changing the value of fruits did not.

As another example, consider the well-known study of Neal et al. (2011). These authors observed that people who often eat popcorn when watching a movie in a cinema theater (“habit” group) will continue to eat popcorn even when it is stale (i.e., devalued) whereas people who do not often eat popcorn in cinemas (“nonhabit” group) stop eating stale popcorn. Although this suggests that eating popcorn in the “habit” group is not mediated by the goal to have tasty food whereas that goal does mediate popcorn eating in the “nonhabit” group, it does not necessarily imply that popcorn eating in the “habit” group was stimulus-driven. For instance, it is possible that eating popcorn in the “habit” group was mediated by the goal to have a more complete cinematic experience. Let us assume that for people who often eat popcorn in a cinema theater, the cinematic experience is not complete without eating popcorn whereas for controls, the richness of the cinematic experience does not depend on eating popcorn. If this assumption is correct, then eating stale popcorn will be goal-conductive for members of the “habit” group but not for controls. In other words, people in the “habit” group might be more willing to tolerate the bad taste of the stale popcorn because for them, eating popcorn while watching a movie has merit as such, even when it does not taste good. Of course, it remains to be seen whether these auxiliary assumptions about the differences in the goal-conduciveness of eating popcorn in the “habit” and “nonhabit” group are valid. If additional studies do not provide support for the alternative goal-directed account, one should be willing to accept the conclusion that the behavior is habitual rather than adhere to the irrefutable claim that the behavior must be mediated by some type of goal. Nevertheless, researchers should consider the possibility that devaluation and contingency degradation

tests lack sensitivity or fail to target the goal that is actually driving behavior (De Houwer et al., 2018)³.

These problems cannot be sidestepped by inferring the lack of goal-directedness from the automatic nature of behavior. Because stimulus-driven behavior is assumed to be automatic, one might see evidence for automaticity as an indication of the fact that behavior is stimulus-driven. However, it is now generally accepted that automatic behavior is not necessarily stimulus-driven (e.g., Bargh, 1989, 1990; Aarts and Dijksterhuis, 2000; for a recent discussion, see Huang and Bargh, 2014). Even addictive behaviors, which are often seen as prototypical examples of automatic behavior because they are emitted despite their obvious negative consequences, are now considered by some to be directed at realizing goals (e.g., Baumeister, 2017; Hogarth, 2018; Kopetz et al., 2018). Moreover, if one would decide to infer the stimulus-driven nature of behavior from its automaticity, there remains the problem of establishing whether behavior qualifies as automatic. Just like there are many definitions of the concept “habit,” there are many definitions of the concept “automatic.” Most of these definitions refer to one or more automaticity features, such as unintentional, involuntary, fast, efficient, or unconscious (Bargh, 1989, 1994; Moors and De Houwer, 2006). Because different automaticity features do not necessarily co-occur, establishing automaticity thus requires one to specify the automaticity features one has in mind and to test the presence of each feature individually. This opens up debates about which features are crucial for determining whether a behavior is automatic and whether the term “automaticity” is still useful as a unifying concept (Fiedler and Hütter, 2014). Moreover, Moors (2016) convincingly argued that the extent to which a type of behavior or process displays a certain feature of automaticity (e.g., the extent to which semantic processing depends on conscious input) can vary across contexts. Hence, there is little merit in saying that a process or behavior has a certain feature of automaticity in an absolute sense (e.g., that semantic processing is an unconscious process)⁴. For all these reasons, there is little

³Note that this problem in part arises because in this and many other studies with humans, researchers did not have full experimental control over the outcomes that at which behavior is directed. Instead, researchers often look at behaviors that were acquired before participants took part in the study (e.g., popcorn eating in cinema visitors). In most animal studies, on the other hand, the potentially habitual behavior has been established experimentally by linking it with a particular outcome (e.g., food). In these cases, there is more certainty about the outcome that is actually controlling the behavior during its initial stages. Hence, it is unlikely that the behavior is controlled by a different outcome when devaluation or contingency degradation tests suggest that it is no longer controlled by the original outcome of the behavior. Nevertheless, even in fully experimental research, one should take care that statistical power is sufficient to establish the absence of an effect (Vadillo et al., 2019) and that the devaluation and contingency degradation tests are sensitive enough.

⁴One could argue that the time needed to initiate or complete a behavior is related to whether the behavior can be considered a skill rather than to whether a behavior is considered to be a habit. Nevertheless, speed of performance has explicitly been put forward by some as a characteristic of habitual behavior, next to other automaticity features (e.g., Wood and Rünger, 2016, p. 292). Because of the difficulty in distinguishing conceptually and empirically between skills and habits, I will sidestep this issue in the current paper.

merit in establishing the stimulus-driven nature of behavior on the basis of its automaticity.

Many definitions of habits do not only incorporate the assumption that habits are stimulus-driven but also assumptions about the factors that are responsible for the stimulus-driven nature of habits (see Gardner, 2015). First, it is often assumed that behavior becomes stimulus-driven if it has been frequently emitted in the context of a certain stimulus. Second, many researchers assume that stimulus-driven behavior is instigated *via* the activation of S-R associations that have been formed gradually as the result of the frequent co-occurrence of a stimulus and a behavior. Although both proposals certainly have merits, from the current perspective, they add additional causal elements to the definition of habits and habitual behavior which result in additional difficulties in establishing whether behavior qualifies as habitual.

Let us first consider frequency as a cause of stimulus-driven behavior. Many researchers assume that behavior should become more habit-like (i.e., stimulus-driven) the more frequently it is emitted in a certain context, that is, the more it is overtrained (e.g., de Wit et al., 2018). The causal role of frequency can be examined by manipulating the frequency of a behavior and observing indices of the stimulus-driven nature of the behavior. Note, however, that experimental designs allow for causal conclusions only if confounding variables are controlled for. For instance, the frequency of behavior (i.e., how often it occurs) can be confounded with the recency of behavior (i.e., the time elapsed between the test phase and the most recent occurrence of a behavior). Assuming that the impact of past events decreases with time and/or the number of intervening events (Ebbinghaus, 1913), it is possible that differences at test between frequent and infrequent behavior reflect recency rather than frequency or a combination of both. A confound between frequency and recency is typically avoided by varying frequency while keeping recency constant across conditions (e.g., de Wit et al., 2018). Another approach which has been implemented less frequently in the literature on habits is to manipulate frequency and recency orthogonally so that the relative contribution of and interaction between both factors can be examined (e.g., Schmidt et al., 2019, submitted).

Researchers often also refer to S-R associations as the mental cause of stimulus-driven behavior. More specifically, they assume that a stimulus can initiate a behavior by activating its representation in memory, activation which can then spread *via* the S-R association to the response representation and thereby bring about the response without the involvement of the representations of goals. Traditionally, S-R associations are assumed to form gradually as the result of the frequent co-occurrence of a stimulus and a behavior, as well as the presence of rewards that follow the behavior when it is emitted in the presence of the stimulus (e.g., de Wit et al., 2007; Wood and R nger, 2016). Establishing that stimulus-driven behavior is mediated by S-R associations is faced with the same problems as establishing stimulus-driven behavior (see above) but also with the additional problem of demonstrating the mediating role of S-R associations. The fact that neither S-R associations themselves nor their causal impact can

be observed directly complicates efforts to verify the involvement of S-R associations in behavior. Procedures have been developed to assess the strength of S-R associations indirectly (e.g., *via* their impact on lexical decision times; e.g., Neal et al., 2012) but the usefulness of these procedures depends on how valid they are (see De Houwer, 2011).

One could argue that stimulus-driven behavior is by definition mediated by S-R associations and that evidence for the stimulus-driven nature of behavior thus constitutes evidence for the mediating role of S-R associations. However, in that case, it is not clear what the idea of S-R associations adds to the notion of stimulus-driven behavior. Such added value can come only from specific theoretical ideas about what those S-R associations are (e.g., abstractive links between mental representations of stimuli and responses), how they are formed (e.g., gradually as the result of repetition or rewards), and how they influence behavior (e.g., *via* the automatic spreading of activation from the stimulus to the associated response representation; e.g., de Wit et al., 2007). Hence, incorporating the notion of S-R associations into the definition of habits comes with specific theoretical commitments, which requires the specification of additional criteria to distinguish “real” habitual behavior (i.e., stimulus-driven behavior that is mediated by a specific type of S-R associations that develops under specific conditions) from other stimulus-driven behavior (i.e., behavior that is stimulus-driven but mediated by another type of representation). In sum, defining habits in terms of S-R associations only aggravates the problem of empirically verifying whether behavior is “truly” habitual.

THE THEORETICAL LEVEL: EXPLAINING HABITUAL BEHAVIOR

Introducing causal elements within the definition of habits and habitual behavior not only results in challenges at the empirical level (i.e., the possibility of verifying that behavior is habitual) but can also limit theoretical innovation. In this section, I focus primarily on definitions of habits that refer to S-R associations. After discussing their dominance, I sketch two alternative theories of habitual (in the sense of stimulus-driven) behavior. In that way, I hope to clarify that it is not only possible to consider other models when trying to explain habitual behavior but also that it can be beneficial to do so. Considering these other models is, however, only possible if one removes the notion of S-R associations from the definition of habits.

The current theoretical literature on habits is dominated by S-R association formation models. Even researchers who do not explicitly define habits in terms S-R associations often consider only S-R associations when trying to explain habitual behavior (e.g., Wood and R nger, 2016). This dominance of S-R association models in the literature on habits is probably based on the fact that these models are compatible with the widespread definition of habitual behavior as behavior that is automatic and stimulus-driven as the result of frequent stimulus–response co-occurrences. Behavior that is driven by S-R

associations (1) must have been emitted frequently enough to allow for the gradual formation of an S-R association, (2) has features of automaticity because activation can spread across associations automatically, and (3) is stimulus-driven in that the activation of S-R associations is instigated only by a stimulus and does not involve goal representations. In fact, the match between the mechanism of activating gradually acquired S-R associations and the phenomenon of habitual (i.e., frequency-induced automatic stimulus-driven) behavior is so good that one might wonder whether any other mechanism could account for behavior that is stimulus-driven and automatic. In this section, I briefly discuss two of these alternatives just to illustrate that (1) other mechanisms are possible and (2) there is merit in at least allowing for theoretical diversity.

First, Logan (1988) pointed out that behavior could be mediated by the similarity-based automatic retrieval of episodic representations from memory. Episodic memory traces differ from S-R associations as they are typically conceived of as being non-abstractive: whereas different experiences all contribute to the strength of a single association, episodic memory models assume that each individual experience is stored as a separate memory trace (e.g., Medin and Schaffer, 1978). Moreover, whereas simple associations do not specify the way in which events are related (e.g., whether A causes, predicts, or merely co-occurs with B; see Lagnado et al., 2007), episodic memory traces encode the way in which an event is constructed by an individual, including assumptions that are made about the relation between events. According to episodic models, stimuli in the current environment automatically activate episodic memory traces that contain information about similar stimuli. If those activated memory traces also contain information about a particular response, then this can lead to the automatic execution of that response. The likelihood that responses are automatically activated depends on the number of episodes that encompass both the stimulus and the response, as well as time that has elapsed since the episode was constructed. Hence, episodic models differ in important ways from S-R association formation models as they are typically conceived of (i.e., concrete vs. abstract; relational vs. associative; similarity-based retrieval vs. spreading of activation). Nevertheless, according to episodic models, a stimulus in the current environment can result in behavior that (1) has frequently been emitted in the context of that stimulus, (2) has features of automaticity, and (3) is merely stimulus-driven⁵. As such, episodic memory models such as

those proposed by Logan (1988) provide an interesting alternative for S-R association formation models of habitual behavior.

Considering also episodic models of habitual behavior will increase theoretical diversity within the literature on habits, which is bound to enrich theoretical discussions and empirical research. For instance, unlike typical S-R association formation models, episodic memory models assign an important role to the recency of events and can thus inspire research that examines the relative contribution of frequency and recency in habitual behavior (see Schmidt et al., 2019, submitted, for an example). Moreover, because episodes can encode also instructions, episodic models might provide a new perspective on the finding that automatic behavior can result from simple instructions and implementation intentions (e.g., Martiny-Huenger et al., 2017; Meiran et al., 2017). Finally, because episodic models assign an important role to factors at retrieval, they can also inspire research on the context dependency of habitual behavior. With regard to the latter point, it remains to be seen whether the emphasis on retrieval factors in episodic models fits with what is known about the functioning of habit memory.

Second, habit researchers have recently benefited from another alternative for the traditional S-R association formation model, namely predictive coding models. These models have been highly influential in neurocognitive research on a variety of topics such as perception, memory, and attention (e.g., Friston, 2010, 2018; Clark, 2013). The core assumption is that organisms constantly build a mental model of the world which allows them to behave in ways that minimize energy expenditure. Both model construction and behavior selection are assumed to be based on inferential processes that can operate under conditions of automaticity. As such, predictive coding models provide a natural account of automatic behavior (Van Dessel et al., 2019). Stimulus-driven behavior, on the other hand, could be conceptualized as behavior that is guided by simple (i.e., hierarchically shallow) models that do not include information about higher order goals of the organism (see FitzGerald et al., 2014; Friston et al., 2016, for more details). Although predictive coding theories are not incompatible with the idea of gradually acquired S-R associations, they do provide a new perspective on how those S-R associations are formed and influence behavior. Moreover, considering predictive coding models could offer highly formalized theories of habits that allow for new predictions, for instance, with regard to what happens when habitual behavior does not lead to predicted outcomes (FitzGerald et al., 2014).

Note that considering alternative theories about habitual behavior does not change the fact that it is advisable to ban assumptions of specific mental representations and processes from the definition of habits. Regardless of the nature of the mental representations or processes that are assumed to be crucial for habits (S-R associations, episodes, predictive coding), it will always be difficult to verify empirically the involvement of a specific type of mental representation or process. Acknowledging multiple mental process explanations of habitual (in the sense of stimulus-driven) behavior also does not solve

⁵Note, however, that episodes can also contain information about the goals that someone has when performing an action. Hence, automatic retrieval of episodes (and thus automatically activated behavior) could also depend on goals at the time of retrieval (Logan, 1988). In those cases, the behavior at the time of retrieval would not qualify as stimulus-driven. It would be interesting to run simulations to see whether there are circumstances in which goals at retrieval do not influence the automatic retrieval of episodes (e.g., when stimulus-response relations remain constant while goals vary). Such simulations could provide new insights into whether and when behavior is stimulus-driven.

the problem that it is difficult to demonstrate with certainty that behavior is not directed at (hidden) goals. However, considering multiple theories about the mental representations and processes that mediate habitual behavior does enrich the theoretical debate and can thus lead to new discoveries.

FINAL THOUGHTS ON OVERCOMING THE CHALLENGES OF HABIT RESEARCH

Habit research is faced with many challenges (see de Wit et al., 2018; Watson and de Wit, 2018, for a discussion). The central message of this paper is that many of these challenges result from the inclusion of causal elements within the definition of habits. This not only makes it difficult to establish and thus study habits and habitual behavior (see section “The Empirical Level: Establishing the Presence of Habitual Behavior”) but can also constrain thinking about the mechanisms mediating habitual behavior (see section “The Theoretical Level: Explaining Habitual Behavior”). Hence, from this perspective, a possible solution for the challenges of habit research is to reduce the number of causal elements from the definition of habits and habitual behavior.

Based on the arguments presented above, it can be strongly recommended to remove any assumptions about S-R associations from the definition of habits and habitual behavior. Such assumptions aggravate the problems of empirically verifying the presence of habitual behavior and entail the risk of impoverishing theoretical debate by (implicitly or explicitly) committing researchers to *a priori* assumptions about the nature of the representations and processes that mediate stimulus-driven behavior. Theories about S-R associations can still be an important part of habit research but rather than being a part of the explanandum (i.e., that which needs to be explained) the role of these theories would be firmly restricted to that of one possible explanans (i.e., that which explains; Hempel, 1970).

What about the widespread idea that habits are stimulus-driven? As noted above, this idea introduces a number of causal elements within the definition of habits, not only about what is a cause of behavior (i.e., the stimulus) but also about what is not a cause of behavior (i.e., goals). Especially the latter element hampers the capacity to determine whether a behavior qualifies as habitual. However, as Gardner (2015) correctly pointed out, scientists are engaged with the causes of behavior rather than with simply describing behavior. As far as psychological explanations go, the distinction between explanations that do and do not involve goals is a fundamental one, not least because it has important implications for how to influence behavior (i.e., manipulating goals will only affect behavior that is mediated by those goals; De Houwer, 2019). Hence, it is understandable that cognitive scientists are interested in studying stimulus-driven behavior, that is, behavior that is not mediated by goals.

Nevertheless, researchers who wish to study habits in the sense of stimulus-driven behavior are well advised to proceed

cautiously. The stimulus-driven nature of behavior cannot be observed directly, nor are there perfect proxies for establishing stimulus-driven behavior. As noted above, devaluation and contingency degradation test are strong indicators of the involvement of goals but not of the non-involvement of goals. Establishing the automaticity of behavior is not only difficult but also does not guarantee that the behavior is stimulus-driven. Although these difficulties should not stop researchers from examining stimulus-driven behavior, they need to be aware of these problems and take them into account when drawing conclusions.

Another option is to ban any reference to the stimulus-driven nature of habits from the definition of habits, which would leave only the notion that habits result from the frequent performance of a behavior in a certain context, as well as the notion that habits are automatic (Gardner, 2015). Choosing this option would imply that behavior is regarded as habitual if it can be established that (1) its presence is due to its past frequency and/or (2) it has features of automaticity. As noted above, both criteria are also not without problems. Establishing the role of frequency requires well-controlled experimental studies. Establishing the automaticity of behavior can entail many different, non-overlapping, and context-dependent automaticity features, some of which are difficult to verify because they refer to mental processes (Moors, 2016). Moreover, focusing exclusively on frequency would eliminate recency as a possible cause of habitual behavior. Hence, researchers who wish to define habit research as the study of frequent or automatic behavior should also be aware of the challenges entailed by this view on habits.

A possible way to reduce these challenges is to focus on features of automatic behavior that are relatively easy to verify. For instance, habit researchers could study behavior that is instigated quickly in certain contexts or that people subjectively experience as having little conscious control over. These criteria can be verified using experimental tasks or questionnaires. Once consensus over these criteria has been reached, researchers could document the moderators of those behaviors (i.e., the conditions under which behaviors with those automaticity features occur), which constrains theories about the mental mechanisms that produce those behavior. Such an approach would imply a clear separation between the explanandum of habit research (i.e., specific instances of automatic behavior) and the explanans of habit research (i.e., assumptions about the causal mechanisms that produce automatic behavior). It would also bring academic habit research closer to the notion that lay people have about habits.

Different researchers will probably choose different paths to overcome the challenges of habit research. Those whose primary interest lies in studying whether and when behavior is stimulus-driven (i.e., not mediated by goal representations) will probably continue to define habitual behavior as stimulus-driven behavior but, hopefully, ban any reference to specific S-R theories from their definitions, as well as use proxies of stimulus-driven behavior in a cautious manner. Those who

wish to understand why behavior can be initiated quickly and why people sometimes report to have little control over their behavior will be probably be happy with defining habits as automatic behavior. The aim of this paper is not to convince researchers to ban all causal elements from the definition of habits (or other concepts in psychology), nor to promote a particular definition of habits. Instead, the main aim is to highlight that choosing a definition of habits has important implications for both empirical research (i.e., how to establish whether behavior is habitual) as well as theory development (i.e., proposals about the mechanisms that underlie habitual behavior). Hence, it is important to make explicit the causal assumptions that researchers make when using a particular definition of habits, as well as to acknowledge the challenges that these assumptions imply.

REFERENCES

- Aarts, H., and Dijksterhuis, A. (2000). Habits as knowledge structures: automaticity in goal-directed behavior. *J. Pers. Soc. Psychol.* 78, 53–63. doi: 10.1037/0022-3514.78.1.53
- Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *The quarterly journal of experimental psychology section B—comparative and. Physiol. Psychol.* 34, 77–98.
- Bargh, J. A. (1989). “Conditional automaticity: varieties of automatic influence on social perception and cognition” in *Unintended thought*. eds. J. Uleman and J. A. Bargh (NY: Guilford), 3–51.
- Bargh, J. A. (1990). “Auto-motives: preconscious determinants of social interaction” in *Handbook of motivation and cognition*. eds. R. M. Sorrentino and E. T. Higgins (New York: Guilford Press), 93–130.
- Bargh, J. A. (1994). “The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition” in *Handbook of social cognition (2nd ed.)*. eds. R. S. Jr. Wyer and T. K. Srull (Hillsdale, NJ: Lawrence Erlbaum), 1–40.
- Barnes-Holmes, D., Barnes-Holmes, Y., Luciano, C., and McEnteggart, C. (2017). From the IRAP and REC model to a multi-dimensional multi-level framework for analyzing the dynamics of arbitrarily applicable relational responding. *J. Contextual Behav. Sci.* 6, 434–445. doi: 10.1016/j.jcbs.2017.08.001
- Baumeister, R. F. (2017). Addiction, cigarette smoking, and voluntary control of action: do cigarette smokers lose their free will? *Addict. Behav. Rep.* 5, 67–84. doi: 10.1016/j.abrep.2017.01.003
- Chiesa, M. (1992). Radical behaviorism and scientific frameworks: from mechanistic to relational accounts. *Am. Psychol.* 47, 1287–1299. doi: 10.1037/0003-066X.47.11.1287
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspect. Psychol. Sci.* 6, 202–209. doi: 10.1177/1745691611400238
- De Houwer, J. (2019). Moving beyond the distinction between system 1 and system 2: conditioning, implicit evaluation, and habitual responding might also be mediated by relational knowledge. *Exp. Psychol.* 66, 257–265. doi: 10.1027/1618-3169/a000450
- De Houwer, J., Tanaka, A., Moors, A., and Tibboel, H. (2018). Kicking the habit: why evidence for habits in humans might be overestimated. *Motiv. Sci.* 4, 50–59. doi: 10.1037/mot0000065
- de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., et al. (2018). Shifting the balance between goals and habits: five failures in experimental habit induction. *J. Exp. Psychol. Gen.* 147, 1043–1065. doi: 10.1037/xge0000402
- de Wit, S., Niry, D., Wariyar, R., Aitken, M. R. F., and Dickinson, A. (2007). Stimulus-outcome interactions during conditional discrimination learning

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The preparation of this paper was made possible by Ghent University Grant BOF16/MET_V/002 to JD.

ACKNOWLEDGMENTS

I thank Agnes Moors for comments on an earlier draft of this paper.

- by rats and humans. *J. Exp. Psychol. Anim. Behav. Process.* 33, 1–11. doi: 10.1037/0097-7403.33.1.1
- Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University.
- Fiedler, K., and Hütter, M. (2014). “The limits of automaticity” in *Dual-process theories of the social mind*. eds. J. W. Sherman, B. Gawronski, and Y. Trope (New York: Guilford Press), 497–513.
- FitzGerald, T. H. B., Dolan, R. J., and Friston, K. J. (2014). Model averaging, optimal inference, and habit formation. *Front. Hum. Neurosci.* 8, 1–11. doi: 10.3389/fnhum.2014.00457
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Friston, K. (2018). Does predictive coding have a future? *Nat. Neurosci.* 21, 277–295. doi: 10.1038/s41593-018-0200-7
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O’Doherty, J., and Pezzulo, G. (2016). Active inference and learning. *Neurosci. Biobehav. Rev.* 68, 862–879. doi: 10.1016/j.neubiorev.2016.06.022
- Gardner, B. (2015). A review and analysis of the use of ‘habit’ in understanding, predicting and influencing health-related behavior. *Health Psychol. Rev.* 9, 277–295. doi: 10.1080/17437199.2013.876238
- Gardner, B., de Bruijn, G. J., and Lally, P. (2011). A systematic review and meta-analysis of applications of the self-report habit index to nutrition and physical activity behaviours. *Ann. Behav. Med.* 42, 174–187. doi: 10.1007/s12160-011-9282-0
- Hempel, C. G. (1970). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.
- Heyes, C. M., and Dickinson, A. (1990). The intentionality of animal action. *Mind Lang.* 5, 87–104. doi: 10.1111/j.1468-0017.1990.tb00154.x
- Hogarth, L. (2018). “A critical review of habit theory of drug dependence” in *The psychology of habit*. ed. B. Verplanken (Berlin, Germany: Springer), 325–341.
- Huang, J. Y., and Bargh, J. A. (2014). The selfish goal: autonomously operating motivational structures as the proximate cause of human judgment and behavior. *Behav. Brain Sci.* 37, 121–135. doi: 10.1017/S0140525X13000290
- Hughes, S., De Houwer, J., and Perugini, M. (2016). The functional-cognitive framework for psychological research: controversies and resolutions. *Int. J. Psychol.* 51, 4–14. doi: 10.1002/ijop.12239
- Kopetz, C., Woerner, J. I., and Briskin, J. L. (2018). Another look at impulsivity: could impulsive behavior represent strategic goal pursuit? *Soc. Personal. Psychol. Compass* 12:e12385. doi: 10.1111/spc3.12385
- Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., and Sloman, S. A. (2007). “Beyond covariation: cues to causal structure” in *Causal learning: Psychology, philosophy, and computation*. eds. A. Gopnik and L. Schulz (New York: Oxford University Press), 154–172.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychol. Rev.* 95, 492–527. doi: 10.1037/0033-295X.95.4.492
- Machado, A., and Silva, F. (2007). Toward a richer view of the scientific method: the role of conceptual analysis. *Am. Psychol.* 62, 671–681. doi: 10.1037/0003-066X.62.7.671

- Martiny-Huenger, T., Martiny, S. E., Parks-Stamm, E. J., Pfeiffer, E., and Gollwitzer, P. M. (2017). From conscious thought to automatic action: a simulation account of action planning. *J. Exp. Psychol. Gen.* 146, 1513–1525. doi: 10.1037/xge0000344
- Medin, D. L., and Schaffer, M. M. (1978). Context theory of classification learning. *Psychol. Rev.* 85, 207–238. doi: 10.1037/0033-295X.85.3.207
- Meiran, N., Liefoghe, B., and De Houwer, J. (2017). Powerful instructions: automaticity without practice. *Curr. Dir. Psychol. Sci.* 26, 509–514. doi: 10.1177/0963721417711638
- Moors, A. (2016). Automaticity: componential, causal, and mechanistic explanations. *Annu. Rev. Psychol.* 67, 263–287. doi: 10.1146/annurev-psych-122414-033550
- Moors, A., Boddez, Y., and De Houwer, J. (2017). The power of goal-directed processes in the causation of emotional and other actions. *Emot. Rev.* 9, 310–318. doi: 10.1177/1754073916669595
- Moors, A., and De Houwer, J. (2006). Automaticity: a conceptual and theoretical analysis. *Psychol. Bull.* 132, 297–326. doi: 10.1037/0033-2909.132.2.297
- Neal, D. T., Wood, W., Labrecque, J., and Lally, P. (2012). How do habits guide behavior? Perceived and actual triggers of habits in daily life. *J. Exp. Soc. Psychol.* 48, 492–498. doi: 10.1016/j.jesp.2011.10.011
- Neal, D. T., Wood, W., Wu, M., and Kurlander, D. (2011). The pull of the past: when do habits persist despite conflict with motives? *Personal. Soc. Psychol. Bull.* 37, 1428–1437. doi: 10.1177/0146167211419863
- Skinner, B. F. (1953). *Science and human behavior*. New York: MacMillan.
- Thraillkill, E. A., and Bouton, M. E. (2015). Contextual control of instrumental actions and habits. *J. Exp. Psychol.* 41, 69–80. doi: 10.1037/xan0000045
- Vadillo, M. A., Linssen, D., Orgaz, C., Parsons, S., and Shanks, D. R. (2019). Unconscious or underpowered? Probabilistic cuing of visual attention. *J. Exp. Psychol. Gen.* doi: 10.1037/xge0000632 [Epub ahead of print].
- Van Dessel, P., Hughes, S., and De Houwer, J. (2019). How do actions influence attitudes? An inferential account of the impact of action performance on stimulus evaluation. *Personal. Soc. Psychol. Rev.* 23, 267–284. doi: 10.1177/1088868318795730
- Watson, P., and de Wit, S. (2018). Current limits of experimental research into habits and future directions. *Curr. Opin. Behav. Sci.* 20, 33–39. doi: 10.1016/j.cobeha.2017.09.012
- Wood, W., and Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychol. Rev.* 114, 843–863. doi: 10.1037/0033-295X.114.4.843
- Wood, W., and Neal, D. T. (2009). The habitual consumer. *J. Consum. Psychol.* 19, 579–592. doi: 10.1016/j.jcps.2009.08.003
- Wood, W., and Rünger, D. (2016). Psychology of habit. *Annu. Rev. Psychol.* 67, 289–314. doi: 10.1146/annurev-psych-122414-033417

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 De Houwer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Hierarchical Action Control: Adaptive Collaboration Between Actions and Habits

Bernard W. Balleine^{1*} and Amir Dezfouli²

¹Decision Neuroscience Laboratory, School of Psychology, University of New South Wales Sydney, Sydney, NSW, Australia,

²Data 61, Commonwealth Scientific and Industrial Research Organisation, Sydney, NSW, Australia

OPEN ACCESS

Edited by:

John A. Bargh,
Yale University, United States

Reviewed by:

Emilio Cartoni,
Italian National Research
Council, Italy
Samuel Joseph Gershman,
Princeton University,
United States

*Correspondence:

Bernard W. Balleine
bernard.balleine@unsw.edu.au

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 31 May 2019

Accepted: 19 November 2019

Published: 11 December 2019

Citation:

Balleine BW and Dezfouli A (2019)
Hierarchical Action Control:
Adaptive Collaboration Between
Actions and Habits.
Front. Psychol. 10:2735.
doi: 10.3389/fpsyg.2019.02735

It is now commonly accepted that instrumental actions can reflect goal-directed control; i.e., they can show sensitivity to changes in the relationship to and the value of their consequences. With overtraining, stress, neurodegeneration, psychiatric conditions, or after exposure to various drugs of abuse, goal-directed control declines and instrumental actions are performed independently of their consequences. Although this latter insensitivity has been argued to reflect the development of habitual control, the lack of a positive definition of habits has rendered this conclusion controversial. Here we consider various alternative definitions of habit, including recent suggestions they reflect chunked action sequences, to derive criteria with which to categorize responses as habitual. We consider various theories regarding the interaction between goal-directed and habitual controllers and propose a collaborative model based on their hierarchical integration. We argue that this model is consistent with the available data, can be instantiated both at an associative level and computationally and generates interesting predictions regarding the influence of this collaborative integration on behavior.

Keywords: goal-directed action, habits, action sequences, chunking, model-based, model-free, reinforcement learning

INTRODUCTION

Although it has long been debated how precisely actions variously called volitional, voluntary or goal-directed should be defined, over the last 20 years or so it has proven fruitful to define as goal-directed those actions demonstrably sensitive to changes in: (1) the causal relationship to their consequences and (2) the value of those consequences (Balleine and Dickinson, 1998). When the performance of an action demonstrates sensitivity to both of these changes, it is defined as goal-directed; when its performance is insensitive to these changes, it is not. By taking this approach, considerable progress has been made not only in providing evidence for goal-directed action in a variety of species (including humans!) but also for the neural bases of these kinds of action. In addition, the usefulness of these tests to delineate goal-directed from non-goal-directed actions has inspired various investigators to apply them as a means of establishing whether the performance of an action reflects the operation of a second form of action control, usually referred to as habits.

Despite their apparent simplicity, habits are actually quite complicated. Although most theories of habit are very clear about what they are – referring to their non-cognitive, repetitive regularity,

their stimulus control, and so on – demonstrating that an action is a habit is not straightforward. For example, numerous papers have advanced the idea that a habit is an action that is *insensitive* to changes in the action-outcome relationship and in outcome value (reviewed in Balleine and O'Doherty, 2010). However, in practice, where the effects of such changes are evaluated against some control group, this has meant asserting the null hypothesis. Thus, for example, when the experimental group differs from, say, a non-devalued or a non-degraded control, then performance of the former is regarded as goal-directed. However, when these experimental and control groups do not differ, then performance of the former is regarded as habitual. Furthermore, these criteria fail to differentiate habits from other forms of reflex; for example, although habits are insensitive to changes in the action-outcome relationship, so are Pavlovian conditioned reflexes (although, whereas habits are insensitive to devaluation, Pavlovian CRs often are not; cf., Dickinson and Balleine, 1994, 2002). The general problem, however, is asserting an action is a habit when it fails to satisfy the tests for goal-directed action, because this does not discriminate that action from performance when the actor is simply confused, forgetful, or having trouble integrating beliefs regarding action outcomes with their desire for a particular outcome. In such cases, behavior may appear habitual when it is in fact controlled by a faulty goal-directed controller. This could be the case in people suffering psychiatric conditions, addictions, or brain damage of various kinds and in such cases although the evidence might confirm their behavior is not normatively goal-directed, it may not be entirely habitual either.

COMPETITION BETWEEN GOAL-DIRECTED AND HABITUAL CONTROL

This latter criticism has obvious implications for how habits should be defined but also affects how we should think about the way that habitual and goal-directed actions interact. Generally speaking, the consensus supposes these forms of action control as competing, at least as far as instrumental performance is concerned (**Figure 1A**). At a behavioral level, for example, it is usual to point, first, to the relatively clear evidence that distinct associative processes underlie the two forms of action control; whereas goal-directed actions depend on the action-outcome association, habits are commonly thought to involve a process of stimulus-response association (Dickinson, 1994). Based on this distinction, various dual process accounts of the way these distinct learning processes influence instrumental performance have been developed, perhaps the most influential of which suggests that, whereas an action, such as lever pressing in rodents, begins under goal-directed control, the net influence of the action-outcome association declines as the strength of the S-R association increases until the influence of the latter exceeds the former and so takes over motor control (Dickinson et al., 1983; Dickinson, 1985, 1994). And, indeed, a number of studies have reported behavioral evidence consistent with the dual process perspective (reviewed in Dickinson and Balleine, 2002).

Neural Evidence

Considerable evidence for competition has also come from studies assessing the neural bases of these two forms of control. Thus, for example, sometime ago we reported evidence that lesions of the prelimbic prefrontal cortex, the dorsomedial striatum, and the mediodorsal thalamus had in common the effect of reducing the sensitivity of instrumental performance to changes in both the action-outcome relationship and outcome value; i.e., compromising goal-directed control appeared to cause a reversion to habit, consistent with the idea that these control processes compete (Balleine, 2005; Balleine et al., 2007; Balleine and O'Doherty, 2010). Following the criticism above, however, loss of sensitivity to tests of goal-directed action may not necessarily mean the action has become a habit and may instead reflect a loss in the accuracy of retrieval or in translating learning to performance. Again, what is required to support this claim is positive evidence that a loss of goal-directed control increases habitual control.

There are two other sources of positive evidence from studies assessing the neural bases of action control consistent with a competitive process. The first comes from findings suggesting that one effect of goal-directed control is to inhibit the performance of habits (Norman and Shallice, 1986). For example, although extensively trained instrumental actions are insensitive to outcome devaluation, this insensitivity is only observed in tests conducted in extinction; i.e., in a situation in which outcome delivery is withheld and so does not provide direct and immediate negative feedback. When feedback is provided, by delivering the devalued outcome contingent on the action, then the performance of even extensively trained actions rapidly adjusts; punishment appears to result in response suppression which is as rapid for an extensively trained action as a relatively modestly trained one (see, for example: Adams, 1982; Dickinson et al., 1983, 1995). Importantly, damage to, or inactivation of, the neural network mediating goal-directed control attenuates this effect of punishment and results in the persistence of an action even when it delivers a demonstrably devalued outcome (one, for example, that the animal will not consume; e.g., Balleine et al., 2003; Yin et al., 2005). Second, a number of studies have found that damage to, or inactivation of, the dorsolateral striatum (Yin et al., 2004), or structures interacting with dorsolateral striatum, such as the central nucleus of the amygdala (Lingawi and Balleine, 2012), can block habitual control resulting in even extensively trained actions remaining goal-directed. Although also consistent with other accounts (see below), this effect is nevertheless consistent with a competitive interaction between habitual and goal-directed control processes and the view that, at least in some situations, when habits are inhibited goal-directed action control is liberated from its competing influence.

Nevertheless, other features of habitual performance tend, on their face, to reduce the importance of much of this evidence for competition between control processes. One factor is the increase in the speed of performance commonly observed to accompany habits. For example, a number of studies have found evidence that the speed of both response initiation (reaction time) and motor movement is increased with experience. Thus, biases in reaction time appear to depend on parameters experienced during prior training rather than new computations (Wong et al., 2017)

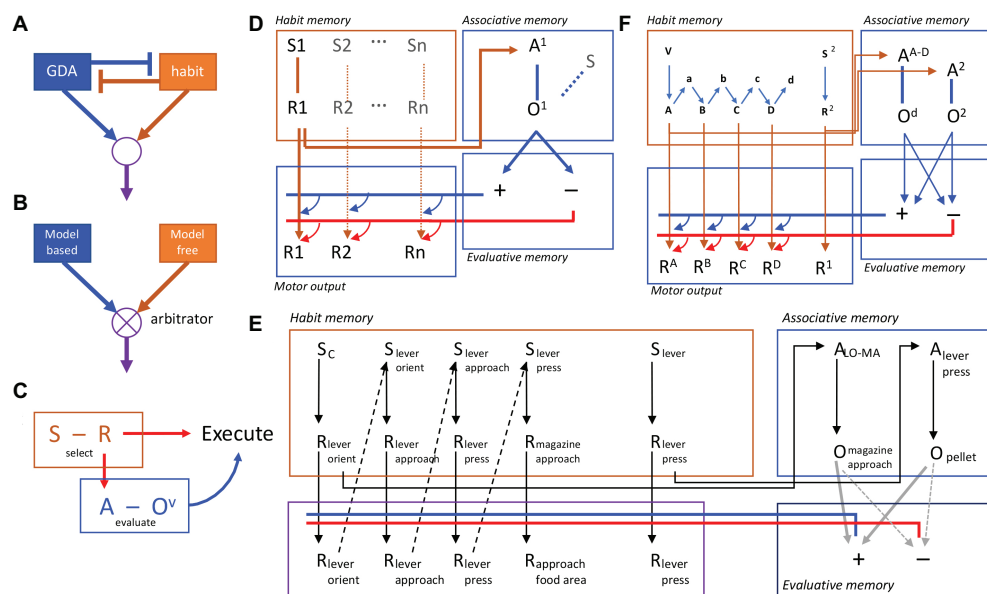


FIGURE 1 | Competition and collaboration in goal-directed and habitual action control. **(A)** Simple model of competition for performance with goal-directed and habitual controllers mutually inhibiting one another. **(B)** More sophisticated approach to competition, with goal-directed and habitual controllers competing through arbitration. **(C)** Behavioral evidence suggests, in contrast to competition, that habit and goal-directed processes are intimately connected and collaborate in action selection, evaluation, and execution. **(D)** A formal associative architecture that instantiates the collaboration between habit and goal-directed controllers through the interaction of habit memory and associative memory systems, the latter feeding back to control performance. Action selection in the habit memory is mediated by the association of $S1$ and $R1$ that feeds forward to provide both subthreshold activation of the motor output and activation of the action representation, $A1$, in the associative memory provoking retrieve of the action outcome ($O1$) and its evaluation through the interaction of the associative and evaluative memory systems. The latter provides a promiscuous, feedback (cybernetic) signal that sums with the forward excitation from the habit memory. If positively evaluated (blue lines/arrows), it provokes action execution; if negatively evaluated (red lines/arrows), it blocks performance. **(E)** An example of the representation of a complex habit sequence in the habit memory incorporating lever press and magazine approach responses together with a simple lever press action. Both are represented in the habit memory (the expanded sequence, the acquisition of which is supported by proprioceptive feedback from motor output) and its chunked representation in the associative memory (e.g., A^{LO-MA}). **(F)** The formal associative-cybernetic model incorporating chunked action sequences and simple actions in both the habit memory and the associative memory.

and are manifest as costs when conflicting response strategies, involving the repetition of movements at a particular speed or toward a particular direction, influence the kinematics of subsequently performance (Huang et al., 2011; Verstynen and Sabes, 2011; Hammerbeck et al., 2014). This is true of studies of human action but has also long been claimed in studies of habit in animals, especially rodents working in runways of various kinds (reviewed in Bolles, 1967, ch. 8). The suggestion that performance speed increases as actions become habitual in some ways trivializes competition between goal-directed and habitual controllers because, if habits are faster, they could potentially be completed before the goal-directed system is engaged and so will not directly compete with goal-directed control. Indeed, it is this speed of action that allows us to make sense of the errors that habits bring; e.g., the planning errors and slips of action apparent in selecting or completing an action that will otherwise result in an unwanted, devalued, or aversive outcome. It can also explain the reversion to goal-directed action induced by inactivation of the dorsolateral striatum; when habitual control is offline, there is simply more time to implement goal-directed control.

Computational Evidence

But what of evidence that goal-directed control inhibits habits? Another class of account consistent with a competitive view

has been driven by the computational descriptions of goal-directed and habitual action control derived from distinct forms of reinforcement learning (RL): model-based RL in the case of goal-directed action and model-free RL in the case of habit (see Dolan and Dayan, 2013 for review). The former views goal-directed control as a planning process; the actor foresees the future actions and the transitions between future states necessary to maximize reward *via* a form of tree search and integrates these into an internal model of the environment. In contrast, model-free RL supposes that action selection in a particular state is determined by the predicted long run future reward value of the action options in that state. Within this literature, whether an agent selects goal-directed or habitual control has been argued to be the outcome of a competitive arbitration process (Figure 1B); computationally, the actor selects the control process for which the state-action value is least uncertain (Daw et al., 2005). And this is true too of more recent accounts; whether framed in terms of reliability (Lee et al., 2014) or costs and benefits to determine the outcome of arbitration (Pezzulo et al., 2013; Shenhav et al., 2013; Keramati et al., 2016), they also contend that actions and habits compete for control. Treatments (whether behavioral or neural) that influence arbitration will be predicted to influence the balance between goal-directed and habitual control; viz., if reduced

reaction times bias arbitration toward a model-free process, then perhaps the delivery of an unexpected aversive or noxious outcome biases arbitration toward a model-based one.

Evidence for competitive model-based and model-free controllers has been most clearly derived from computational analyses of performance on a class of task that attempts to pit goal-directed and habitual choices against one another in a multistage discrimination situation (Daw et al., 2011). The aim of this task is essentially to set up a continuous revaluation procedure across trials. In one version, a first stage choice transitions probabilistically to one of two second stage states, one with a higher probability than the other. At the second stage, a choice results, again probabilistically, in reward or no reward, the probability of which changes slowly throughout the task to encourage the decision-maker to sample new options. The task is, therefore, structured on a RL view of the world with explicit states and action-related transitions between those states. Importantly, it is assumed that learning that a choice in the second stage state results in reward (or no reward) revalues (or devalues) that state as a goal. The question then becomes: does the decision-maker take advantage of that information or not? If so then their choice is *assumed* to reflect planning based on the interaction of the stage 1 and stage 2 states and so to be model-based (or goal-directed). If not, it is *assumed* that the choice merely recapitulates prior performance and is model-free (or habitual).

Based on these assumptions, the stage 1 choices of human subjects on the two-stage task show a mixture of model-based and model-free control (Daw et al., 2011) that can be biased toward one or other process by a variety of factors; e.g., amount of training (Gillan et al., 2015), cognitive load (Otto et al., 2013), altered activity in the dorsolateral frontal cortex (Smittenaar et al., 2013), and (likely relatedly) *via* the influence of various psychiatric conditions (Gillan et al., 2016). Although this intermixing of controllers is consistent with variations in the influence of competitive controllers, trial-to-trial variation presents something of a puzzle, the explanation of which – if we are to maintain this perspective – returns us to the issue of arbitration.

Importantly, while there are computational theories of arbitration (e.g., Griffiths et al., 2015), whether an arbitrator actually regulates the contribution of each system remains unknown. And, in fact, analyses that break the world into discrete states may not be the best way to assess this problem. Although such analyses may be helpful both when the experimenter is trying to tie neural events to behavioral responses or is hoping, computationally, to apply a reinforcement learning approach to these tasks, the original data that inspired our understanding of distinct goal-directed and habitual forms of action control came from continuous, self-paced, unsignaled situations in which humans and other animals explore the environment, discover its structure, learn new actions and their causal consequences, and then utilize that knowledge to maximize reward. Non-human animals in particular encode these relationships based on their own experience and not *via* the instructions of the experimenter. In contrast, what human participants learn on multistage discrimination tasks can

be difficult to discern and may not accord with the assumptions of model-based and model-free RL analyses as to the drivers of performance. There are issues in establishing whether the assumptions from model-based and model-free reinforcement learning are consistent with the subjects' behavior; how accurately they update common and rare transition probabilities; how large the state-space that subjects use to make choices actually is (see Akam et al., 2015 for discussion). Furthermore, other factors, such as performance rules or environmental cues, including the stimulus predictions embedded in the task, could also influence performance; indeed, it has never been clear why experimenters commonly use both actions and stimuli to predict the second stage states in two-stage tasks. Another factor recently suggested to influence arbitration between model-based and model-free control is the integration of the costs and benefits of each system; i.e., the rewards based on the average return of model-based and model-free control against which are contrasted the intrinsic cost of model-based control (Kool et al., 2016). Interestingly, evidence has been collected from novel versions of the two-stage task suggesting variations in reward value and costs based on planning complexity can alter the model-based and model-free trade-off (Kool et al., 2017, 2018). Importantly, however, these factors do not appear to influence arbitration on the original version of the two-stage task, likely due to its intransigence in the calculation of reward estimates due to a lack of access to the second stage reward outcomes (Kool et al., 2016). Indeed, whereas model-based and model-free RL provide reasonable simulations of the first stage choices of the two-stage task, experimenters investigating these positions have typically not generated predictions about what animals will do on the second stage choice (Dezfouli and Balleine, 2013, 2019). It is clear, therefore, that our understanding of what animals and humans are actually doing on these complex tasks is very far from settled.

Taken together, these issues concerning the behavioral, neural, and computational evidence for competition between action controllers raise significant questions regarding: (1) how habits are best characterized; (2) the kind of evidence that we should accept for their occurrence; and (3) whether explaining their interaction with goal-directed control requires the generation of a third kind of quasi-controller positioned to arbitrate between the other two. Fortunately, there are other accounts available that allow us to move beyond each of these issues.

COLLABORATION BETWEEN CONTROLLERS

Against the competition view, alternative positions have been developed proposing that goal-directed and habitual controllers collaborate to coordinate instrumental performance. In the past, we have described a number of sources of behavioral evidence for this perspective (Balleine and Ostlund, 2007), among the strongest of which comes from studies assessing the factors controlling the selective reinstatement of instrumental actions (Ostlund and Balleine, 2007). The basic phenomenon was established as an assessment of the effects of outcome delivery

on subsequent action selection. Rats trained on two actions for distinct outcomes were then given a period of extinction on both actions until performance was completely withheld. At that point, one or other of the two outcomes was delivered non-contingently. The question at issue was what the free outcome delivery would produce; if the outcome retrieved the action with which it was associated then we should expect that action to be selected and executed, and that is what we observed. Subsequently, we sought to assess whether the outcome selected the action that delivered the non-contingent outcome as a goal or whether that outcome served as a stimulus that retrieved the next performed action. To achieve this, rats were again trained on two actions for different outcomes; however, each action–outcome pair was trained in alternation; i.e., $A1 \rightarrow O1$ was always followed by $A2 \rightarrow O2$. Again, both actions were extinguished before we assessed the effects of non-contingent outcome delivery on the reinstatement of $A1$ and $A2$. If an outcome retrieves the action that delivered it as a goal, then delivering, say, $O1$ should retrieve $A1$. If, however, $O1$ acts as a stimulus that retrieves the next action, then $O1$ should retrieve $A2$. In fact, we found the latter result; outcome-specific reinstatement appears to reflect the effect of a forward outcome–response association on performance. Furthermore, this effect was not diminished by devaluing the reinstating outcome suggesting that outcome-mediated response retrieval is not dependent on the outcome's value but on its stimulus properties. This result suggests, therefore, that instrumental action *selection* is initiated by a form of S–R process in which the stimulus properties of the outcome are the proximate cause of action retrieval (Balleine and Ostlund, 2007).

Importantly, subsequent studies found that, when retrieved in this way, it is the outcome that serves the selected action as a goal that mediates the *execution* of the action. To establish this, we used a similar training situation except that the outcomes were used as explicit discriminative cues for action selection, and found that these kinds of stimuli can, in fact, engage an evaluative process but of the action subsequently retrieved by those discriminanda (Ostlund and Balleine, 2007). Devaluing the outcome that served as a goal for the retrieved action reduced the vigor of performance but not the ability of the outcome to serve a discriminative cue, consistent with other reports using more traditional discriminative stimuli (Colwill and Rescorla, 1990; Rescorla, 1994). That is, performance, but not action selection, was attenuated if the outcome earned by the reinstated action was devalued. In the ordinary course of events, therefore, the outcome controls actions in two ways: (1) through a form of S–R, or ideomotor, association in which the stimulus properties of the outcome can select the action with which they are associated; and (2) through the standard R–O association in which a selected action retrieves its specific outcome as a goal. Clearly, the subsequent retrieval of the value of the outcome is a necessary step toward the actual performance of the action. Hence, this behavioral evidence suggests that a selection–evaluation–execution sequence lies at the heart of instrumental performance and that this control requires the collaborative integration of habitual S–R and goal-directed R–O control processes (Figure 1C).

Cybernetic Control

At least two kinds of account accord with this collaborative control process. The first, advanced some years ago, is what has become known as the associative-cybernetic model of instrumental performance (Dickinson and Balleine, 1993). This account has its origins in Thorndike's (1931) ideational theory of instrumental action proposing that a stimulus that evokes a response urge or tendency calls to mind the consequences of the action selected by that tendency and these two processes – driven essentially by stimulus–response and action–outcome associations – check or favor one another to release action execution. In addition to providing a clear basis for the collaborative integration of habitual and goal-directed controllers, this view also has the merit of providing an answer to one of the thornier questions; why do we do anything at all? Early cognitive theorists, concerned by the poverty of the stimulus–response approach, developed models of action based on more elaborate internal variables (e.g., Tolman, 1932). Nevertheless, how thought initiates action remained an ongoing issue; the concern being, as Guthrie put it, that such views left the actor buried in thought (Guthrie, 1935). When and why does thinking about actions and their consequences stop and acting begin? Thorndike's account suggests that it is external stimuli rather than thoughts that initiate this process by urging a response; that the action and its consequences are brought to mind only subsequently, at which point the value of the latter provides the basis for either checking the urge, when the consequences are punishing, or favoring it, when they are rewarding, thereby providing the necessary feedback to modulate action execution.

These ideas have been developed in a number of ways to capture both the behavioral data on instrumental performance and their neural bases (reviewed elsewhere; Dickinson and Balleine, 1993; Dickinson, 1994; Balleine and Ostlund, 2007). Generally, it has been suggested that a stimulus–response memory interacts with an associative memory to drive the retrieval of a specific action and its consequences, that the latter retrieves an incentive memory of the outcome that, by marshaling specific motivational and emotional processes, determines the value of the outcome, to potentiate or de-potentiate the motor signal associated with the response tendency of the S–R memory, thereby increasing the probability that the action will be executed. It is this latter process that constitutes the cybernetic or feedback component of the model (Figure 1D).

Hierarchical Control

Alternatively, we have recently argued that goal-directed and habitual control processes interact in a hierarchical manner; i.e., that habits are selected by a goal-directed control process as one means of achieving a specific goal (Dezfouli and Balleine, 2012). Within this account, although habits are often described as single-step actions, their tendency to combine or chunk with other actions and their insensitivity to changes in the value of, and causal relationship to, their consequences suggest that they are better viewed as forming the elements of chunked action sequences. In this context, chunking means that the decision-maker treats the whole sequence of actions as a single action unit and so the individual actions of which the sequence

is composed are represented independently of their individual outcomes. As a consequence, the value of an action sequence will be established independently of the individual action-outcome contingencies and the values of the outcomes of the action elements inside the sequence boundaries, which will be invisible to the decision-maker. Once selected, each action will then be executed in the order determined by the sequence in an open loop manner; i.e., without further feedback from their individual consequences.

Integrating Cybernetic and Hierarchical Control

In fact, hierarchical and cybernetic control are not mutually exclusive and, indeed, starting with James (1890), there has been a long tradition of associative accounts of action sequences, particularly from within the behaviorist tradition that used stimulus-response sequences to explain apparently cognitive control processes. A good example of this approach is Hull's explanation of latent learning. Tolman, for example, was able to demonstrate that changing the value of a specific goal box in a previously explored maze by giving a rat food in that box was sufficient immediately to alter the speed and accuracy with which the rat reached the goal subsequently without the need for additional training (Tolman and Honzik, 1930). The natural interpretation of this effect is that the rat had learned about the change in value of the goal and was able to incorporate that knowledge into what it knew about the structure of the maze to alter its choice performance, much as we have argued for goal-directed actions generally. In response to effects like this, however, behavioral theorists introduced the fractional goal-response, responses such as chewing or licking, that, when associated with other responses within the maze, could form a sequence able to explain choice performance without resorting to goal-directed control (Hull, 1952).

Although these kinds of explanation are no longer favored for goal-directed actions, they give a feeling for how an account of habits in terms of action sequences might be constructed and deployed. In the simplest case, it would apply to overtraining-induced habits by arguing that the target action, say lever pressing, is incorporated into a sequence with other common responses performed around the lever press response; e.g., lever orienting, lever approach, lever press, magazine approach, magazine entry, magazine exit, lever orient, and so on (see **Figure 1E**). Initially, these sequences of responses would be purely incidental; the simple component action of lever pressing is sufficient and any tendency to press the lever will call to mind the action-outcome relationship resulting in outcome evaluation and the execution or suppression of the action. With practice, however, chunking these component responses together would allow the whole sequence to run off rapidly and smoothly using minimal cognitive resources. There are, however, costs associated with this form of action control; chunking these component responses together may allow stimuli antecedent to the response tendency to set off the habitual chain without requiring the animal to monitor each component action, however it will also render the consequences of responses within the chain and the value

of those consequences invisible to the decision maker. If such sequences are structured and selected independently of their simpler component actions, such as lever pressing, and if the sequence's relationship to and the value of its outcome are not dependent on these component actions, then one can immediately see how, when chunked within a sequence, a target action can appear insensitive to changes in its relationship to and the value of its programmed consequences (cf. Dezfouli and Balleine, 2012; Dezfouli et al., 2014).

Within the associative-cybernetic model, habit sequences would form within the habit memory through the integration of responses, perhaps *via* their feedback; i.e., the proprioceptive stimuli they evoke. This response-response chaining is what is meant by the chunking of an action sequence and, as an action, it can be selected in the associative memory just as any other action is selected; i.e., a response tendency, initiated in habit memory, activates the action sequence representation and its outcome in associative memory. If positively evaluated, each subsequent response will be executed without evaluation until the sequence is terminated (see **Figure 1F**).

Although it was argued above that such an account can explain why habits are insensitive to degradation and devaluation treatments, it might be asked, if the outcome of the sequence needs to be evaluated positively for the sequence to be initiated, why devaluation does not result in a reduction in the production of the overall sequence. The answer to this is that it can do so if the outcome that is devalued is the outcome associated with the sequence (Ostlund et al., 2009). If, however, the outcome that is devalued is associated with a response *inside the sequence boundaries*, then the devalued outcome will be invisible to the associative memory and will not be evaluated. In this case, the sequence will persist despite devaluation. That something like this must be going on is suggested by the fact that, after overtraining, habitual lever presses in rats have been found to become more sensitive to devaluation over the course of extinction as, presumably, the press-approach sequence described above was broken down (Dezfouli et al., 2014).

More direct evidence for this account has recently been reported by Ostlund and colleagues (Halbout et al., 2019). In this study, rats were trained to lever press for a food pellet reward before the goal-directed nature of this response was assessed using an outcome devaluation assessment conducted in extinction. The investigators developed a novel microstructural analysis of the performance of the animals during training and test, investigating the tendency to press the lever but also the degree to which such presses were followed by approach responses to the food magazine and how the relative incidence of these responses changed after devaluation. Importantly, they found evidence that the rats used two different strategies when initiating the lever press response, performing it as part of an action chunk (press-approach) or as a discrete action (press only). Consistent with an account in terms of habitual sequences, these distinct strategies appeared to be differentially sensitive to reward devaluation; whereas the rats were generally less likely to lever press for the devalued than for the valued reward, the press-approach chunk was found to be less sensitive to reward devaluation than presses that were not followed by

approach. Furthermore, the proportion of chunked lever press-approach actions was actually greater for the devalued action than for the valued action. This suggests there was a change in the willingness to select the chunked sequence on the devalued relative to the non-devalued action, consistent with the claim that the sequence had a higher value than the individual lever press after devaluation.

Generally, therefore, we argue that hierarchical control can be accommodated within an associative-cybernetic account of instrumental conditioning. In fact, it appears to be well suited to this account with individual actions and chunked action sequences sitting at the same level in the associative memory and with simple or serially chained stimulus-response associations sitting at the same level in habit memory. This account is also consistent with several other features of habitual control. First it is consistent with the increased speed of habit execution: without having to evaluate the individual actions through the cybernetic feedback component of the model, the action sequence can run off more rapidly than if each response is evaluated. Second, this account addresses slips of actions by pointing to the chaining of responses at a mechanistic level. Appropriate response feedback will initiate the next action in a chain irrespective of the outcome of that response (Matsumoto et al., 1999). Furthermore, feedback relating to a response in the middle of a chain should be expected to result in a “capture error”; i.e., in the completion of that chain even when the animal is pursuing some other outcome (Norman and Shallice, 1986).

EVIDENCE FOR HIERARCHICALLY ORGANIZED COLLABORATION

Given that hierarchical control can be implemented within an associative-cybernetic architecture that requires the integration of goal-directed and habit controllers to explain instrumental performance, what evidence exists for this kind of collaboration? Here we describe two sources of evidence from human and rodent subjects consistent with this account, both taken from performance on the two-stage task described above.

Human

As mentioned, the two-stage task developed by Daw et al. (2011) essentially arranges for changes in value to occur while the decision-maker is faced with an ongoing series of binary choices. Repeating past choices is assumed to be driven by the habit controller; altering choices in accord with predictions of future outcomes is assumed to be driven by the goal-directed controller. Critically for this analysis, all previous assessments of these factors have focused purely on stage 1 choices largely because popular reinforcement learning descriptions of choice on this task, i.e., model-based and model-free RL, only make differential predictions regarding stage 1 choices. However, it should be clear that, because the hierarchical-cybernetic model described above views habits as sequences of responses nested within a goal-directed controller and treats all actions as requiring collaboration between habit and goal-directed control,

this approach is unique in making differential predictions not just for the first stage choices but also for second stage (and indeed for further) choices too.

We constructed a version of the two-stage task – see **Figure 2A** (cf. Dezfouli and Balleine, 2013 for details) – in which human subjects were instructed to make a binary choice at stage 1 (i.e., A1 or A2), the outcome of which was either O1 or O2, which were distinct two-armed slot machines. Subjects could then make a second binary choice in stage 2, choosing one or other arm (i.e., R1 or R2), and were then rewarded or not rewarded for their choice. We arranged the relationship between the stages as in previous reports of this task: i.e., A1 commonly led to O1 and A2 to O2; however, on rare trials, A1 led to O2, and A2 to O1. As a consequence of this arrangement, the role of stage 2 choices was, essentially, to manipulate the value of O1 and O2 and, in order to revalue the outcomes during the session, the probability of reward following each stage 2 choice increased or decreased randomly on each trial, causing frequent devaluation or revaluation of the O1 and O2 outcomes during the course of the task. Whereas changes in outcome value are usually accomplished by offline treatments, such as specific satiety and taste aversion learning, in this task values are changed through exposure to rare transitions inserted among the more common transitions.

Replicating previous reports, we found that stage 1 choices were sensitive to this form of revaluation, confirming that these actions were goal-directed – **Figure 2C** (human data). However, and more importantly, because two steps are required to reach reward it is possible for subjects to expand their choice options from A1 and A2 by combining stage 1 and stage 2 actions to construct action sequences; i.e., A1R1, A1R2, A2R1, A2R2 and to choose between these options based on their relationship to reward – see **Figure 2B**. Although the choice of stage 2 action (R1 vs. R2) should be based on the outcome of the stage 1 action, we found that, when the previous trial was rewarded and subjects repeated the same stage 1 action (A1 or A2), they also tended to repeat the same stage 2 action (R1 or R2), irrespective of the outcome of the stage 1 action. In these cases, the stage 2 action was determined at stage 1 when the sequence was executed. This observation of the open-loop execution of actions was not due to the generalization of action values from the common to the rare second stage outcome (e.g., using the example in **Figure 2B**; from O1 to O2). If this were the case, then subjects should have been more likely to repeat the same stage 2 action irrespective of the stage 1 action chosen. However, subjects had a higher tendency to repeat the same stage 2 action (e.g., O2) only when they executed the same stage 1 action (e.g., A1) – **Figure 2D**.

Recall that, according to the hierarchical approach, actions will be habitual if they fall within the boundaries of a sequence. And that is the case here; the outcomes of the stage 1 choices (i.e. O1 and O2) fall within the boundaries of the action sequences, consistent with the claim that these sequences were not always revalued during rare trials. This finding suggests that subjects should also make systematic errors after revaluation; i.e., if revaluation occurs on a rare transition then selecting the same action sequence performed on the previous trial

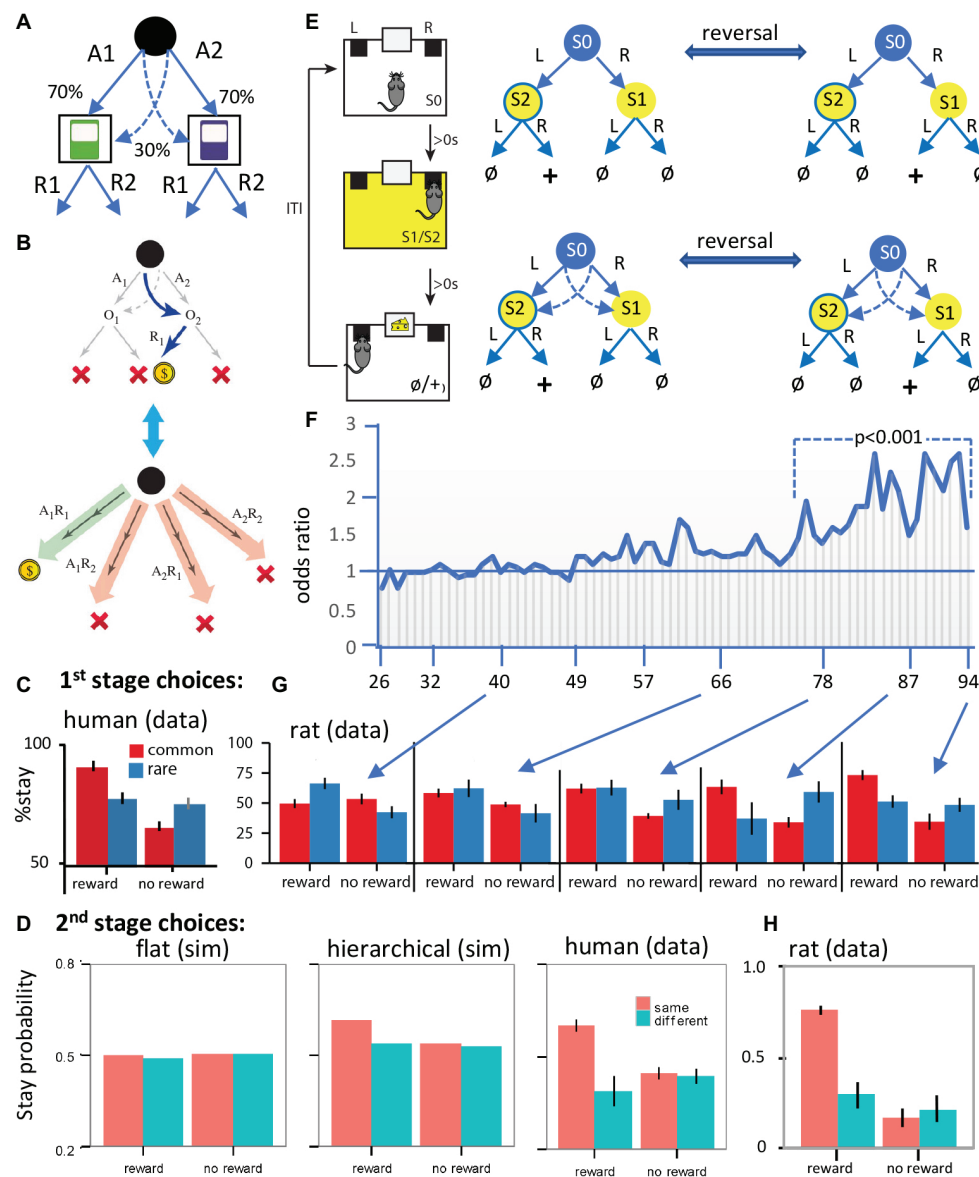


FIGURE 2 | Evidence for hierarchical collaboration in humans and rats. **(A)** Two-stage task in human subjects. **(B)** After a rare transition (example shown) and revaluation of O2 (upper panel), an expanded action repertoire using action sequences (e.g., A1R1) can induce insensitivity to revaluation of the second stage choice (e.g., R1). **(C)** The influence of reward and non-reward on the tendency to stay on the same first stage choice after a common and a rare transition in human subjects. **(D)** Simulated (sim) second stage choices from various flat model-based and/or model-free RL models (left panel), a hierarchical RL model (center), and the human data (right panel). **(E)** Design of a two-stage task in rats with training conducted on a two-stage discrimination that is reversed, initially, every four trials and subsequently every eight trials. At various points in training, we included rare transitions as probe tests (sessions 40, 66, 78, 87, and 94). **(F)** The odds ratio of staying on the same stage 1 action after reward on the previous trial over the odds ratio after no reward. The horizontal line represents the indifference point. Each vertical line is one session. **(G)** Results from the probe tests. Note the comparable performance of rats and humans when rats show evidence of having acquired an accurate representation of the multistage nature of the task. **(H)** Rat data from second stage choices using a comparable version of the task to that used in humans. Panels **(A–D,G,H)** are taken directly from Dezfouli and Balleine (2013, 2019). Panels **(E,F)** are redrawn from Dezfouli and Balleine (2019).

means the subject must have ignored the fact that it was the alternative stage one action that was revalued. And indeed, consistent with this, reward on the previous trial increased the likelihood of repeating the same stage 1 action, whatever outcome and stage 1 action was revalued. Importantly, as previously reported, we also found performance to be a mixture of responses apparently insensitive to outcome revaluation and

those sensitive to these manipulations. On previous accounts, such findings were argued to reflect competition between model-based and model-free controllers. On the hierarchical account, however, this merely reflects the difference between a model-based controller selecting simple actions (A1 and A2) on the one hand and habit sequences (A1R1, A1R2, A2R1, and A2R2) on the other. Importantly, we found evidence that,

whereas model-based and model-free RL were as successful as a hierarchical RL model in simulating the stage 1 choices, only the hierarchical RL model could capture the stage 1 and stage 2 choices, and this superiority was established using Bayesian comparison between these different model families – see simulations in **Figure 2D**.

Generally, therefore, this study found evidence of action sequences that were insensitive to a change in outcome value, a finding that is uniquely addressed by the collaborative hierarchical account. Another feature of this account is that it provides a straightforward reason why the chronometry of action and habits should differ. Any attempt to evaluate each simple action before execution will necessarily slow the temporal dynamics of choice between the two stages compared to habit sequences, which can run off continually in open loop fashion without intervening evaluation. As such, when the second action in the sequence is not taken at stage 2, then reaction times should increase. We found evidence for this prediction in the data: if the previous trial was rewarded, reaction times were significantly faster (<379 ms) when a subject completed an action sequence than when the second stage action was not executed as part of a sequence (>437 ms). Importantly, this effect was not significant when the previous trial was not rewarded, which rules out the possibility that the observed increase in the reaction times was because of the cost of switching to the other second stage action. Only when (1) the previous trial was rewarded, (2) the subject took the same first stage action, and (3) their reaction time was low did the subject repeat the second stage action, consistent with the prediction of the collaborative hierarchical account.

Rodent

A number of reports have now been published evaluating two-stage discrimination learning in rodents (Akam et al., 2015; Miller et al., 2017; Groman et al., 2019). In a recent study, using a task modeled on that described for use in humans above, we sought to investigate how the state-space and action representations adapt to the structure of the world during the course of learning without any explicit instructions about the structure of the task – which obviously cannot be provided to rats (Dezfouli and Balleine, 2019) – see **Figure 2E**. Briefly, we found evidence that, early in training, the rats made decisions based on the assumption that the state-space was simple and the environment composed of a single stage, whereas, later in training, they learned the true multistage structure of the environment and made decisions accordingly – **Figures 2E,G**. Importantly, we were also able to show that concurrently with the expansion of the state-space, the set of actions also expanded and action sequences were added to the set of actions that the rats executed in similar fashion to human subjects – **Figure 2D** vs. **Figure 2H**: human vs. rat data.

In more detail, the lack of instructions implies that the rats have first to establish the nature of what might be called the “task space,” in this case, the fact that the task has two stages. This means that the rats needed to use feedback from the previous trial to track which stage 2 state was rewarded so as to take the stage 1 action leading to that state. It was

clear that, early in training, the rats responded as if the first stage was not related to the second stage; as shown in **Figure 2F**, the rats failed to show a tendency to take the same stage 1 action after earning a reward on the previous trial and instead tended to repeat the action taken immediately prior to reward delivery; i.e., if they took “L” at stage 1, and “R” at stage 2 and earned reward, then they repeated action “R” at the beginning of the next trial. Therefore, actions were not based on a two-stage representation. Importantly, however, this pattern of choices reversed as the training progressed and the rats started to take the same stage 1 action that earned reward on the previous trial rather than repeating the action most proximal to reward – **Figure 2F**. Clearly, the rats had learned that the task has two stages and, at that point, acquired the correct state-space of the task. If this is true, however, then, during the course of training, the task space used by the animals expanded from a simple representation to a more complex representation consistent with its two-stage structure.

Importantly, learning the interaction of the two stages of the task is not the only way that the rats could have adapted to the two-stage structure of the environment; as mentioned above, in this task, reward can be earned either by executing simple actions in each stage or an action sequence; i.e., the rats could have learned to press the left or the right lever in series and/or to perform left → right or right → left as a chunked sequence of actions. Using these expanded actions, the rats could then repeat a rewarded sequence instead of merely repeating the action proximal to the reward. If this is true, however, then the transition in the pattern of stage 1 actions shown in **Figure 2F** could have been due to the development of action sequences rather than learning the task space. To establish whether the rats were using chunked sequences of actions, we examined their choices in probe test sessions in which the common (trained) transitions from stage 1 were interleaved with rare transitions; meaning that, after repeating the same stage 1 action, rats could end up in a different stage 2 state than on the previous trial – see **Figure 2G** for 1st stage choices and **Figure 2H** for 2nd stage choices. In this situation, we should expect them to take a different stage 2 action, if they were selecting actions singly, whereas, if they are repeating the previously rewarded sequence, they should take the same stage 2 action. In fact, the data revealed clear evidence for the latter and for the fact that the rats were using action sequences in this way – **Figure 2H**. Generally, if the previous trial was rewarded and the rats stayed on the same stage 1 action, then they also tended to repeat the same stage 2 action. Therefore, the pattern of choices at stage 2 we observed was consistent with the suggestion that the rats expanded the initial set of actions to a more complex set that included action sequences.

Hence, exactly as we found in human subjects, we found evidence that rats could incorporate both simple actions and complex action sequences into their repertoire and that, when responding on a sequence, the actions in the sequence were performed regardless of their specific consequences. We also sought to establish the computational model that best characterized the decision-making process used by the rats comparing non-hierarchical model-based RL, hierarchical model-based RL,

and a hybrid model-based RL and model-free RL and found, using Bayes model comparison, that hierarchical model-based RL provided the best explanation of the data.

Taken together, these experiments provide consistent evidence, across species in rats and humans, that a hierarchical collaborative process mediates instrumental performance in which simple actions and chunked sequences of actions are available for evaluation by the same goal-directed control process in associative memory and, when positively evaluated, add similarly to the impetus for those urges to be executed.

DISCUSSION

The issue of how to identify a habit is rapidly becoming an important one for neuroscience and behavioral analyses of decision-making and action control to resolve. The suggestion that habits are merely the obverse of goal-directed actions, i.e., are actions that can be shown to be insensitive to their causal consequences and to the value of those consequences, is simply too broad. Many actions will appear habitual by these criteria when they are not, and, as mentioned above, in practice, these criteria devolve to asserting the significance of the null hypothesis.

Defining Habits

In order to overcome this issue, positive qualities of habitual control need to be specified. Within the current framework, we advanced the claim that one way to identify habits is *via* their relationship to other actions within chunked action sequences. Habits, it was claimed, are not single solutions but sit within a flow of stimuli and responses with internal response-induced stimuli supporting the initiation of each subsequent action in a sequence of actions. This is not to say that sequences of this sort cannot be quite short, even though, with continuing practice, they are likely to become quite elaborate. Rather it is claimed that any action that is habitual will be performed in an open loop manner; that its antecedent causes are the effects of the immediately preceding action and its consequences relevant only for the next response in the chain. From this perspective flows other potential features of habits; for example, their chronometry: the reduced reaction time, and increased speed of movement that accompanies these kinds of action spring immediately from the nature of action sequences as open loop systems. The lack of dependency of each sequential movement on feedback from their external consequences ensures that each movement can be initiated quickly. Similarly, the refinement of each movement through repetition and its association with its specific eliciting conditions within the sequence ensures its topographical similarity across instances (meaning the invariance in the kinematics of the motor movement). Habits, then, are actions shown to accord with four distinct observations: (1) relatively rapidly deployed and executed, (2) relatively invariant in topography, (3) incorporated into chunked action sequences, and (4) insensitive to changes in their relationship to their individual consequences and the value of those consequences.

Actions and Habits Do Not Compete

The division of actions and habits into separate and competing control processes is difficult to sustain when their level of collaboration is fully recognized. As described here, the evidence points strongly to the integration of S-R and R-O selection processes through which the various options for action are evaluated. An urge can then be acted upon, whether through a single response or a sequence of responses, or it can be withheld. In some cases, the strength and speed of an urge can produce slips of action; i.e., actions that would otherwise have been withheld. In others, the selection of an action that is part of, or similar to an action that is part of, a sequence can result in “action capture” and the unintentional completion of a sequence of responses inappropriate to the situation. These errors are anticipated from a hierarchical control perspective, whereas from a competitive perspective they are not.

Although the behavioral, neural, and computational evidence for competition between controllers seems overwhelming, careful consideration of this evidence suggests that much of it is open to reinterpretation. From the current perspective, for example, the general claim is that factors argued to influence arbitration between goal-directed and habitual controls can be as readily argued to influence choice between simple actions and action sequences. Costs and benefits influencing this selection process will do so for much the same reason that has been suggested previously; except, of course, the emphasis will be largely on the reduced cost associated with selecting sequences and the potentially increased rewards associated with simple actions due to their more immediate adjustment to environmental constraints based on feedback. Similarly, to the extent that cognitive load and increased planning complexity favor habits (see, for example, Otto et al., 2013), a model-based controller should be expected to select action sequences more than simple actions. This is because the evaluation of action sequences is less cognitively demanding than a set of single actions as the former do not rely on calculating the value of middle states. Similarly, with changing planning complexity; in simple environments planning can be handled by individual actions, which have a higher accuracy, but as the environment becomes more complex the reliance on action sequences becomes more important because the cost of evaluating individual actions increases exponentially with the complexity of the environment. Nevertheless, although many of the interpretations of the behavioral and neural evidence have generated definitions of habit that are, ultimately, circular, the computational approach is different in this regard. The evidence from tasks and models is impressively closely related. Much of this evidence has, however, been driven by a number of simplifying assumptions that in many ways beg the question; such as equating habits with reward-related repetition and so with model-free control.

Computational Collaboration

We contend, therefore, that an architecture favoring the collaboration between controllers makes greater sense of the data, appears less subject to arbitrary assumption, and so more open to test. We advanced these ideas here by relating a

hierarchical reinforcement learning approach to the functions of the associative memory in an associative-cybernetic model of instrumental conditioning. The mechanics of the individual actions, or action sequences, we assume to be the province of the S-R memory, and the evaluation of these actions, including their costs, to be determined by an incentive memory. This provides a simple “algorithmic level” architecture within which collaboration is structurally determined through the selection-evaluation-execution of simple actions or action sequences and is amenable in computational terms to hierarchical reinforcement learning.

Perhaps for this reason, several computational accounts appear, superficially at least, to have similar features to the hierarchical account. For example, one collaborative view, Dyna (Sutton, 1991; see also Gershman et al., 2014; Momennejad et al., 2018), proposes that model-based replay can train the model-free system; a suggestion that devolves to something like rehearsal or perhaps consolidation. An animal simulating or thinking through previous choices through the steps of a decision tree could provide sufficient instances to enable a model-free system to learn more rapidly. This is, however, clearly *learning*-related collaboration; goal-directed and habitual controllers are collaborating in training habitual actions, not in the *performance* of instrumental actions generally. Although one could certainly imagine this kind of process contributing to the consolidation or chunking of habitual sequences of actions, it is not clear how it would function to select between the various options subsequently. It could, as has been argued (Momennejad et al., 2018), improve goal-directed planning, but in that case it remains unclear whether such improvement is due to better integration of performance factors or improved encoding of task structure.

Another interesting example is that of Cushman and Morris' (2015) habitual goal selection theory, which inverts the relationships described here, proposing model-free control over hierarchical goal selection. From this perspective, a habit controller provides the animal with goals toward which it can plan in a goal-directed manner. These ideas are interesting but require significant broadening of what is traditionally taken to be the subject matter of habitual control. More typically in the literature the goal of a habit is taken to be a specified motor movement; it is not a state of affairs in the world. An animal working to change the world to accord with its desires is usually taken to be working in a goal-directed manner; its aim is an external goal-state and the way in which its actions achieve that state is of only secondary importance (e.g., whether the rat presses the lever with its paw or its elbow is immaterial to ensuring delivery of a food pellet). In many ways, Cushman and Morris' claims have much in common with theories emphasizing the function of discriminative cues, such as occasion-setters in hierarchical S-(R-O) theories of instrumental action (Rescorla, 1991). On such views these associations are modulatory; the stimulus modulates the selection and performance of specific actions in a hierarchical fashion and not as a S-R habit. Within the hierarchical-cybernetic model described here, Cushman and Morris' habitual controller would not lie in the habit memory but would modulate action selection in the associative memory in line with associative accounts of modulation. Given the

division we have drawn between sequential and simple goal-directed actions, therefore, we suggest that habitual goal selection theory applies more directly to goal selection within the goal-directed system and is not related to habits.

An explicitly performance-based collaborative account has also been developed by Keramati et al. (2016) based on a “planning until habit” approach; i.e., a certain amount of goal-directed planning is undertaken until a habit is selected at which point the habit takes over the control of performance. This account has potentially a great deal more in common with the hierarchical approach because habits are nested within the goal-directed planner which selects habits at some point in the decision tree to complete the action; essentially a model-based process uses model-free values at the end of the decision tree to complete the action. In contrast, the hierarchical approach to habit described above can be implemented using hierarchical RL which eschews a description of this process as model-free. A similar approach is taken in a recent paper by Miller et al. (2019) who argue that habits are mediated by a value-free perseverative process that, following Thorndike's law of exercise and Guthrie's contiguity account, is determined by repetition alone. From this perspective, goal-directed actions are mediated by model-free and model-based processes, the former when outcomes are represented by their general affective qualities and the latter when they are characterized by their specific sensory properties. Nevertheless, these forms of action control do not collaborate and their interaction remains both competitive and mediated by an arbitrator, the latter sensitive to the strength of the action-outcome contingency.

It may be possible within a value-free model of habits to develop an account of chunked action sequences in which they are mediated by motor stimuli, much as we have argued for the integrated hierarchical-cybernetic model above. However, from the value-free perspective, if such sequences are habitual they will also be value-free and there is good evidence to suggest that this is not the case. For example, Ostlund et al. (2009) trained rats on two action sequences and found that, although the individual responses of which they were composed were insensitive to outcome devaluation and contingency degradation, these manipulations reduced the performance of the specific sequences that delivered the devalued or the non-contiguous outcome during these tests. Thus, although the individual actions in the sequences appeared habitual, the sequences themselves were clearly goal-directed.

AUTHOR CONTRIBUTIONS

BB wrote the paper. AD edited the paper.

FUNDING

The research reported in this paper was supported by a grant from the Australian Research Council, DP150104878, and a Senior Principal Research Fellowship from the NHMRC of Australia, GNT1079561, to BB.

REFERENCES

- Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q. J. Exp. Psychol. Sect. B* 34, 77–98. doi: 10.1080/14640748208400878
- Akam, T., Costa, R., and Dayan, P. (2015). Simple plans or sophisticated habits? State, transition and learning interactions in the two-step task. *PLoS Comput. Biol.* 11:e1004648. doi: 10.1371/journal.pcbi.1004648
- Balleine, B. W. (2005). Neural bases of food-seeking: affect, arousal and reward in corticostriatal limbic circuits. *Physiol. Behav.* 86, 717–730. doi: 10.1016/j.physbeh.2005.08.061
- Balleine, B. W., Delgado, M. R., and Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* 27, 8161–8165. doi: 10.1523/JNEUROSCI.1554-07.2007
- Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419. doi: 10.1016/S0028-3908(98)00033-1
- Balleine, B. W., Killcross, A. S., and Dickinson, A. (2003). The effect of lesions of the basolateral amygdala on instrumental conditioning. *J. Neurosci.* 23, 666–675. doi: 10.1523/JNEUROSCI.23-02-00666.2003
- Balleine, B. W., and O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35, 48–69. doi: 10.1038/npp.2009.131
- Balleine, B. W., and Ostlund, S. B. (2007). Still at the choice-point: action selection and initiation in instrumental conditioning. *Ann. N. Y. Acad. Sci.* 1104, 147–171. doi: 10.1196/annals.1390.006
- Bolles, R. C. (1967). *Theory of motivation*. New York: Harper & Row.
- Colwill, R. M., and Rescorla, R. A. (1990). Effect of reinforcer devaluation on discriminative control of instrumental behavior. *J. Exp. Psychol. Anim. Behav. Process.* 16, 40–47. doi: 10.1037/0097-7403.16.1.40
- Cushman, F., and Morris, A. (2015). Habitual control of goal selection in humans. *PNAS* 112, 13817–13822. doi: 10.1073/pnas.1506367112
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 1204–1215. doi: 10.1016/j.neuron.2011.02.027
- Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560
- Dezfouli, A., and Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* 35, 1036–1051. doi: 10.1111/j.1460-9568.2012.08050.x
- Dezfouli, A., and Balleine, B. W. (2013). Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Comput. Biol.* 9:e1003364. doi: 10.1371/journal.pcbi.1003364
- Dezfouli, A., and Balleine, B. W. (2019). Learning the structure of the world: the adaptive nature of state-space and action representations in multi-stage decision-making. *PLoS Comput. Biol.* 15:e1007334. doi: 10.1371/journal.pcbi.1007334
- Dezfouli, A., Lingawi, N. W., and Balleine, B. W. (2014). Habits as action sequences: hierarchical action control and changes in outcome value. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 369, pii: 20130482. doi: 10.1098/rstb.2013.0482
- Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philos. Trans. R. Soc., B* 308, 67–78. doi: 10.1098/rstb.1985.0010
- Dickinson, A. (1994). "Instrumental conditioning" in *Animal cognition and learning*. ed. N. J. Mackintosh (London: Academic Press), 4–79.
- Dickinson, A., and Balleine, B. W. (1993). "Actions and responses: the dual psychology of behaviour" in *Spatial representation*. eds. N. Eilan and R. A. McCarthy (Oxford: Basil Blackwell), 277–293.
- Dickinson, A., and Balleine, B. W. (1994). Motivational control of goal-directed action. *Anim. Behav.* 22, 1–18. doi: 10.3758/BF03199951
- Dickinson, A., and Balleine, B. W. (2002). "The role of learning in motivation" in *Learning, motivation & emotion, volume 3 of Steven's handbook of experimental psychology*. ed. C. R. Gallistel. Third ed (New York: John Wiley & Sons), 497–533.
- Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., and Boakes, R. A. (1995). Motivational control after extended instrumental training. *Anim. Learn. Behav.* 23, 197–206. doi: 10.3758/BF03199935
- Dickinson, A., Nicholas, D. J., and Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to Reinforcer devaluation. *Q. J. Exp. Psychol. Sect. B* 35, 35–51. doi: 10.1080/14640748308400912
- Dolan, R. J., and Dayan, P. (2013). Goals and habits in the brain. *Neuron* 80, 312–325. doi: 10.1016/j.neuron.2013.09.007
- Gershman, S. J., Markman, A. B., and Otto, A. R. (2014). Retrospective revaluation in sequential decision-making: a tale of two systems. *J. Exp. Psychol. Gen.* 143, 182–194. doi: 10.1037/a0030844
- Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., and Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *elife* 5, pii: e11305. doi: 10.7554/eLife.11305
- Gillan, C. M., Otto, A. R., Phelps, E. A., and Daw, N. D. (2015). Model-based learning protects against forming habits. *Cogn. Affect. Behav. Neurosci.* 15, 523–536. doi: 10.3758/s13415-015-0347-6
- Griffiths, T. L., Lieder, F., and Goodman, N. D. (2015). Rational use of cognitive resources: levels of analysis between the computational and the algorithmic. *Top. Cogn. Sci.* 7, 217–229. doi: 10.1111/tops.12142
- Groman, S. M., Massi, B., Mathias, S. R., Curry, D. W., Lee, D., and Taylor, J. R. (2019). Neurochemical and behavioral dissections of decision-making in a rodent multistage task. *J. Neurosci.* 39, 295–306. doi: 10.1523/JNEUROSCI.2219-18.2018
- Guthrie, E. R. (1935). *The psychology of learning*. New York: Harpers.
- Halbout, B., Marshall, A. T., Azimi, A., Liljeholm, M., Mahler, S. V., Wassum, K. M., et al. (2019). Mesolimbic dopamine projections mediate cue-motivated reward seeking but not reward retrieval in rats. *Elife* 8, pii: e43551. doi: 10.7554/eLife.43551
- Hammerbeck, U., Yousif, N., Greenwood, R., Rothwell, J. C., and Diedrichsen, J. (2014). Movement speed is biased by prior experience. *J. Neurophysiol.* 111, 128–134. doi: 10.1152/jn.00522.2013
- Huang, V. S., Haith, A., Mazzoni, P., and Krakauer, J. W. (2011). Rethinking motor learning and savings in adaptation paradigms: model-free memory for successful actions combines with internal models. *Neuron* 70, 787–801. doi: 10.1016/j.neuron.2011.04.012
- Hull, C. L. (1952). *A behavior system: An introduction to behavior theory concerning the individual organism*. New Haven CT: Yale University Press.
- James, W. (1890). *The principles of psychology*. New York: Henry Holt & Co.
- Keramati, M., Smittenaar, P., Dolan, R. J., and Dayan, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12868–12873. doi: 10.1073/pnas.1609094113
- Kool, W., Cushman, F. A., and Gershman, S. J. (2016). When does model-based control pay off? *PLoS Comput. Biol.* 12:e1005090. doi: 10.1371/journal.pcbi.1005090
- Kool, W., Gershman, S. J., and Cushman, F. A. (2017). Cost-benefit arbitration between multiple reinforcement learning systems. *Psychol. Sci.* 28, 1321–1333. doi: 10.1177/0956797617708288
- Kool, W., Gershman, S. J., and Cushman, F. A. (2018). Planning complexity registers as a cost in metacontrol. *J. Cogn. Neurosci.* 30, 1391–1404. doi: 10.1162/jocn_a_01263
- Lee, S. W., Shimojo, S., and O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81, 687–699. doi: 10.1016/j.neuron.2013.11.028
- Lingawi, N. W., and Balleine, B. W. (2012). Amygdala central nucleus interacts with dorsolateral striatum to regulate the acquisition of habits. *J. Neurosci.* 32, 1073–1081. doi: 10.1523/JNEUROSCI.4806-11.2012
- Matsumoto, N., Hanakawa, T., Maki, S., Graybiel, A. M., and Kimura, M. (1999). Role of [corrected] nigrostriatal dopamine system in learning to perform sequential motor tasks in a predictive manner. *J. Neurophysiol.* 82, 978–998. doi: 10.1152/jn.1999.82.2.978
- Miller, K. J., Botvinick, M. M., and Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. *Nat. Neurosci.* 20, 1269–1276. doi: 10.1038/nn.4613
- Miller, K. J., Shenhav, A., and Ludvig, E. A. (2019). Habits without values. *Psychol. Rev.* 126, 292–311. doi: 10.1037/rev0000120
- Momennejad, I., Otto, A. R., Daw, N. D., and Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *elife* 7:e32548. doi: 10.7554/eLife.32548
- Norman, D. A., and Shallice, T. (1986). "Attention to action" in *Consciousness and self-regulation*. eds. R. J. Davidson, G. E. Schwartz, and D. Shapiro (Boston, MA: Springer), 1–18.

- Ostlund, S. B., and Balleine, B. W. (2007). Selective reinstatement of instrumental performance depends on the discriminative stimulus properties of the mediating outcome. *Anim. Learn. Behav.* 35, 43–52. doi: 10.3758/BF03196073
- Ostlund, S. B., Winterbauer, N. E., and Balleine, B. W. (2009). Evidence of action sequence chunking in goal-directed instrumental conditioning and its dependence on the dorsomedial prefrontal cortex. *J. Neurosci.* 29, 8280–8287. doi: 10.1523/JNEUROSCI.1176-09.2009
- Otto, A. R., Gershman, S. J., Markman, A. B., and Daw, N. D. (2013). The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol. Sci.* 24, 751–761. doi: 10.1177/0956797612463080
- Pezzulo, G., Rigoli, F., and Chersi, F. (2013). The mixed instrumental controller: using value of information to combine habitual choice and mental simulation. *Front. Psychol.* 4:92. doi: 10.3389/fpsyg.2013.00092
- Rescorla, R. A. (1991). Associative relations in instrumental learning: the eighteenth Bartlett memorial lecture. *Q. J. Exp. Psychol. Sect. B* 43, 1–23.
- Rescorla, R. A. (1994). Transfer of instrumental control mediated by a devalued outcome. *Anim. Learn. Behav.* 22, 27–33. doi: 10.3758/BF03199953
- Shenhav, A., Botvinick, M. M., and Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* 79, 217–240. doi: 10.1016/j.neuron.2013.07.007
- Smittenaar, P., FitzGerald, T. H. B., Romei, V., Wright, N. D., and Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron* 80, 914–919. doi: 10.1016/j.neuron.2013.08.009
- Sutton, R. S. (1991). Dyna, an integrated architecture for learning, planning, and reacting. *ACM SIGART Bull.* 2, 160–163. doi: 10.1145/122344.122377
- Thorndike, E. L. (1931). *Human Learning*. New York: Century.
- Tolman, E. C. (1932). *Purposive behavior in animals*. New York: Century Books.
- Tolman, E. C., and Honzik, C. H. (1930). *Introduction and removal of reward, and maze performance in rats*, Vol. 4, University of California Publications in Psychology, 257–275.
- Verstynen, T., and Sabes, P. N. (2011). How each movement changes the next: an experimental and theoretical study of fast adaptive priors in reaching. *J. Neurosci.* 31, 10050–10059. doi: 10.1523/JNEUROSCI.6525-10.2011
- Wong, A. L., Goldsmith, J., Forrence, A. D., Haith, A. M., and Krakauer, J. W. (2017). Reaction times can reflect habits rather than computations. *elife* 6, pii: e28075. doi: 10.7554/eLife.28075
- Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19, 181–189. doi: 10.1111/j.1460-9568.2004.03095.x
- Yin, H. H., Ostlund, S. B., Knowlton, B. J., and Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.* 22, 513–523. doi: 10.1111/j.1460-9568.2005.04218.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Balleine and Dezfouli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Model-Free RL or Action Sequences?

Adam Morris* and Fiery Cushman

Department of Psychology, Harvard University, Cambridge, MA, United States

The alignment of habits with model-free reinforcement learning (MF RL) is a success story for computational models of decision making, and MF RL has been applied to explain phasic dopamine responses (Schultz et al., 1997), working memory gating (O'Reilly and Frank, 2006), drug addiction (Redish, 2004), moral intuitions (Crockett, 2013; Cushman, 2013), and more. Yet, the role of MF RL has recently been challenged by an alternate model—model-based selection of chained action sequences—that produces similar behavioral and neural patterns. Here, we present two experiments that dissociate MF RL from this prominent alternative, and present unconfounded empirical support for the role of MF RL in human decision making. Our results also demonstrate that people are simultaneously using model-based selection of action sequences, thus demonstrating two distinct mechanisms of habitual control in a common experimental paradigm. These findings clarify the nature of habits and help solidify MF RL's central position in models of human behavior.

OPEN ACCESS

Edited by:

John A. Bargh,
Yale University, United States

Reviewed by:

Richard P. Cooper,
Birkbeck, University of London,
United Kingdom
Dorsa Amir,
Boston College, United States

*Correspondence:

Adam Morris
adam.mtc.morris@gmail.com

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 05 April 2019

Accepted: 06 December 2019

Published: 20 December 2019

Citation:

Morris A and Cushman F (2019)
Model-Free RL or Action Sequences?
Front. Psychol. 10:2892.
doi: 10.3389/fpsyg.2019.02892

Keywords: reinforcement learning, action sequences, model-free control, habit, decision-making

1. INTRODUCTION

Sometimes people make decisions by carefully considering the likely outcomes of their various options, but often they just stick with whatever worked in the past. For instance, people sometimes flexibly plan a new route to work when their old route is under construction, but sometimes they follow the old route anyway. This fundamental distinction—often cast as “planned” vs. “habitual” behavior—animates a century of decision-making research and organizes a wide array of human and non-human behaviors (Dolan and Dayan, 2013).

This distinction is commonly formalized within the “reinforcement learning” (RL) framework (Sutton and Barto, 1998; Dolan and Dayan, 2013). In this framework, planning is a form of explicit expected value maximization, or “model-based” reinforcement learning (Daw et al., 2011; Doll et al., 2015). But what is the appropriate formal description of habitual action?

Currently, two basic accounts compete (Figure 1). The first posits that habits arise from a representation of historical value, averaging across similar past episodes—a form of *model-free reinforcement learning* (MF RL) (Schultz et al., 1997; Glascher et al., 2010; Dolan and Dayan, 2013). In other words, people repeat actions when they have been rewarded often in the past. For instance, a person might habitually pull their smart-phone out of their pocket when standing in line because they have often enjoyed using their phone in similar past circumstances.

In contrast, the second posits that habits arise from the “chunking” of actions into sequences that often co-occur (Dezfouli and Balleine, 2012, 2013; Dezfouli et al., 2014). For instance, the sequence of actions that a person uses when tying their shoes co-occurs commonly, and so this sequence has been “chunked.” Although the chunk itself may be assigned value and controlled by an instrumental system, the elements within the chunk are not assigned value; a person executing a chunked action sequence is simply on auto-pilot.

These models are regarded as competitors because they offer divergent accounts for many of the same empirical phenomena. Most pointedly, a recent influential critique from Dezfouli and Balleine (DB) (Dezfouli and Balleine, 2012, 2013; Dezfouli et al., 2014) seeks to explain current behavioral and neural evidence for model-free RL instead in terms of action sequences selected by a superordinate planning process. In other words, they posit that model-free RL is not employed by humans; value representations are employed exclusively during model-based planning, and habitual action exclusively reflects chunked action sequences.

In theory, however, these proposed mechanisms are not incompatible—they could operate side-by-side within a single cognitive architecture. Here, we show that both model-free RL and chunked action sequences simultaneously contribute to human decisions. To do this, we modify a popular set of “two-step” behavioral tasks to isolate unique behavioral signatures of each. Using the modified tasks, we demonstrate both (a) model-based control of action sequences (consistent with DB), but (b) model-free control of single, non-sequenced actions (inconsistent with DB). Thus, our results indicate two important and distinct forms of behavioral organization that contribute to “habitual” (i.e., non-planned) action.

We first review the reinforcement learning framework, and then present the standard two-step task designed to distinguish between MF and MB influence on choice. Then, following DB, we show how (for a particular representation of the task’s reward structure) model-based selection of chunked action sequences can produce seemingly MF-like behavior on this standard task. Finally, we demonstrate that an alternate variant of the task predicts separate behavioral signatures for model-free control and action sequences, and we present two experiments in which people simultaneously exhibit both signatures.

2. TWO MODELS OF HABITUAL ACTION

Reinforcement learning offers a powerful mathematical framework for characterizing different types of decision algorithms, and allows us to conceptually and empirically distinguish between two forms of habitual action: (1) model-free

RL, and (2) model-based RL with action sequences. In this section, we first introduce the classic distinction between model-based and model-free control, and describe an experimental paradigm, the “two-step task,” which was purported to provide evidence for model-free control in humans. We then introduce DB’s “action sequences” critique, and show how a model-based algorithm with action sequences could produce the patterns of habitual behavior in the original two-step task.

Before continuing, there are two theoretical issues worth clarifying. First, throughout this paper, we assume that the behavior produced by a model-free RL controller maps onto our intuitive notion of “habitual” behavior. This assumption, though common (Glascher et al., 2010; Dolan and Dayan, 2013), has been disputed (Miller et al., 2019). We do not engage with this important debate here. Our experiments dissociate model-free RL from model-based action sequences, and test whether humans actually employ model-free RL. If it turns out that model-free RL is not the right description of true habits, but instead represents a different type of unplanned behavior, then our results should be reinterpreted in that light.

Second, throughout this paper, we take “model-free” to mean a type of decision controller that does not store or use information about its environment’s “transition function”—i.e., what the consequences in the environment will be of taking an action from a particular state. It is sometimes difficult to draw a sharp line between model-free and model-based algorithms; there may be a spectrum between them (Miller et al., 2019). Nonetheless, there is a clear distinction between the two ends of the spectrum, with model-free algorithms relying primarily on caching from experience with minimal prospection at decision time, and model-based algorithms relying primarily on forward planning over a model of the environment’s transition function. For our simulations and model-fitting, we will rely on algorithms considered canonical examples of each type (Sutton and Barto, 1998).

2.1. Model-Based and Model-Free Reinforcement Learning

In the RL framework, an agent is in an environment characterized by the tuple (S, A, T, R) , where S is the set of states that the agent

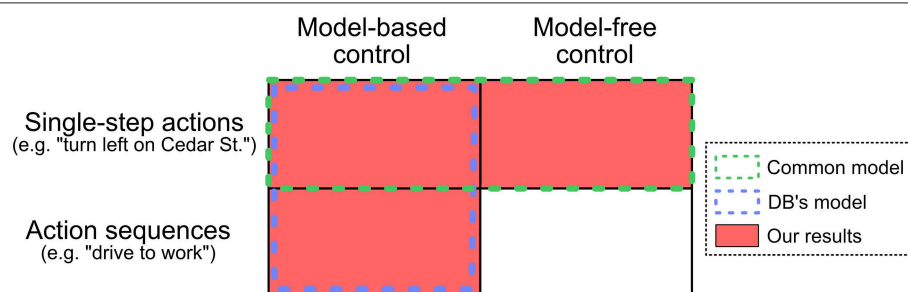


FIGURE 1 | Relationship between various models. Human behavior in sequential decision making tasks is often modeled as the red squares: a mixture of model-based and model-free control of single-step actions. Dezfouli and Balleine (2012, 2013) argue that previous empirical results can be explained by the blue outline: model-based control of single-step actions and action sequences. We report evidence that people simultaneously use both model-based control of action sequences and model-free control of single-step actions.

can be in, A is the set of actions available at each state, T is a function describing the new state to which an action transitions, and R is a function describing the reward attained after each transition (Sutton and Barto, 1998). (For simplicity, we assume there is no discounting). The agent's goal is to find a policy—a function that describes the probability of taking each action in each state—that maximizes the agent's long-term reward. To accomplish this, the agent estimates the sum of expected future rewards following each action, called the action's "value," and then simply chooses actions with high values. We will denote the value of an action a in state s as $Q(s, a)$.

In model-based RL, the agent learns a representation of the transition function T' and reward function R' .¹ For instance, the agent might represent that taking action 1 in state 3 has a 40% chance of leading to state 4—or formally, $T'(a = 1, s = 3, s' = 4) = 0.4$. (This is analogous to representing the consequences of one's actions—i.e., "turning left at this intersection will lead to Cedar Street"). Then, the agent might represent that transitioning to state 4 gives a reward of +10. (This is analogous to representing the desirability of those consequences—i.e., "Cedar Street is the fastest way to work"). Before making a choice, a model-based agent can recursively integrate over the decision tree implied by these two representations to compute the precise value of each available action:

$$Q_{MB}(s, a) = \sum_{s'} T'(s, a, s') * (R'(s, a, s') + \max_{a' \in A} Q_{MB}(s', a')) \quad (1)$$

where s is the agent's current state, a is the action under consideration, and s' are the possible subsequent states.

In contrast, model-free agents do not represent the transition or reward functions—i.e., they don't prospect about the consequences of their actions. Instead, model-free agents estimate action values directly from experience, and cache these value representations so they can be accessed quickly at decision time. (In other words, instead of learning that "turning left at this intersection will lead to Cedar street", a model-free agent will have simply learned that "turning left at this intersection is good"). In the popular model-free algorithm Q-learning (Watkins and Dayan, 1992), for instance, action values are updated *after* each choice according to the following formula:

$$Q_{MF}(s, a) \leftarrow Q_{MF}(s, a) + \alpha * (r + \max_{a' \in A} Q_{MF}(s', a') - Q_{MF}(s, a)) \quad (2)$$

where s is the agent's current state, a is the chosen action, s' is the subsequent state, r is the reward received, and α is a free parameter controlling the learning rate. By incorporating both the immediate reward and the next state's action values into the update rule, the $Q(s, a)$ value estimates converge to the long-term expected reward following each action, and Q-learning agents learn to maximize long-term reward accumulation without explicitly representing the consequences of their choices.

¹In many cases, the reward function is static and given to the RL agent ahead of time. But, in our experiments (and many others; e.g., Glascher et al., 2010; Kool et al., 2017), the reward function is constantly changing, and so the agent must continually learn it.

[Note that, although canonically considered a model-free algorithm (Sutton and Barto, 1998), Q-learning involves some minimal type of prospection: It uses value estimates of the actions a' in the subsequent state s' to update its value estimate for selecting a in s . As discussed above, the line between model-free and model-based is not always sharp (Miller et al., 2019). Nonetheless, like standard model-free algorithms, Q-learning does not use an explicit model of the transition function T . Moreover, its lack of forward planning at decision time means that it produces the standard signature of model-free control in the task used here, which we describe below. Hence, it is an appropriate formalization of model-free RL for our purposes].

Model-free RL is a particularly powerful model of habitual behavior. It captures human and animal behavior in a variety of paradigms (Glascher et al., 2010; Daw et al., 2011; Dolan and Dayan, 2013), as well as behavioral deficits in obsessive-compulsive disorder (Voon et al., 2015), Parkinson's (Frank et al., 2004), and drug addiction (Redish, 2004). It elegantly explains phasic dopamine responses in primate midbrain neurons (Schultz et al., 1997) and BOLD signal changes in the human striatum (Glascher et al., 2010). Finally, it forms the basis of models of other cognitive processes, such as moral judgment (Crockett, 2013; Cushman, 2013), working memory gating (O'Reilly and Frank, 2006), goal selection (Cushman and Morris, 2015), and norm compliance (Morris and Cushman, 2018).

2.2. The Two-Step Task

The difference between model-based and model-free control can be illustrated in a popular sequential decision paradigm called the "two-step task" (Daw et al., 2011). In the two-step task, participants go through a series of trials and make choices that sometimes lead to reward. On each trial, they make two choices (Figure 2A). The first choice ("Stage 1") presents two options ("Left" and "Right") that we label L1 and R1. These actions bias probabilistic transitions to two subsequent states ("Stage 2" states) which are yellow and green, and which are not rewarded. For instance, L1 might typically lead to a green screen, and R1 to a yellow screen. After transitioning to one of the Stage 2 states, people then make a second choice between two further options, L2 and R2. These each probabilistically transition to one of two terminal states: a state with reward, or a state without reward. These transition probabilities drift over the course of the experiment. Thus, to maximize earnings, participants must continually infer which Stage 2 state-action pair has the highest probability of reward, and make choices in both stages to attain that outcome.

The two-step task was initially designed to distinguish between model-based and model-free control of single-step, non-sequenced actions. The key logic of this experimental design depends on the probabilistic transitions between Stage 1 and Stage 2 (Figure 2A). 80% of the time, L1 leads to green and R1 to yellow. But, 20% of the time, the transitions are reversed. Participants' choices following rare transition trials reveal the distinction between model-free and model-based RL. Imagine an agent chooses L1, gets a rare transition to yellow, chooses

²We thank a reviewer for raising this point.

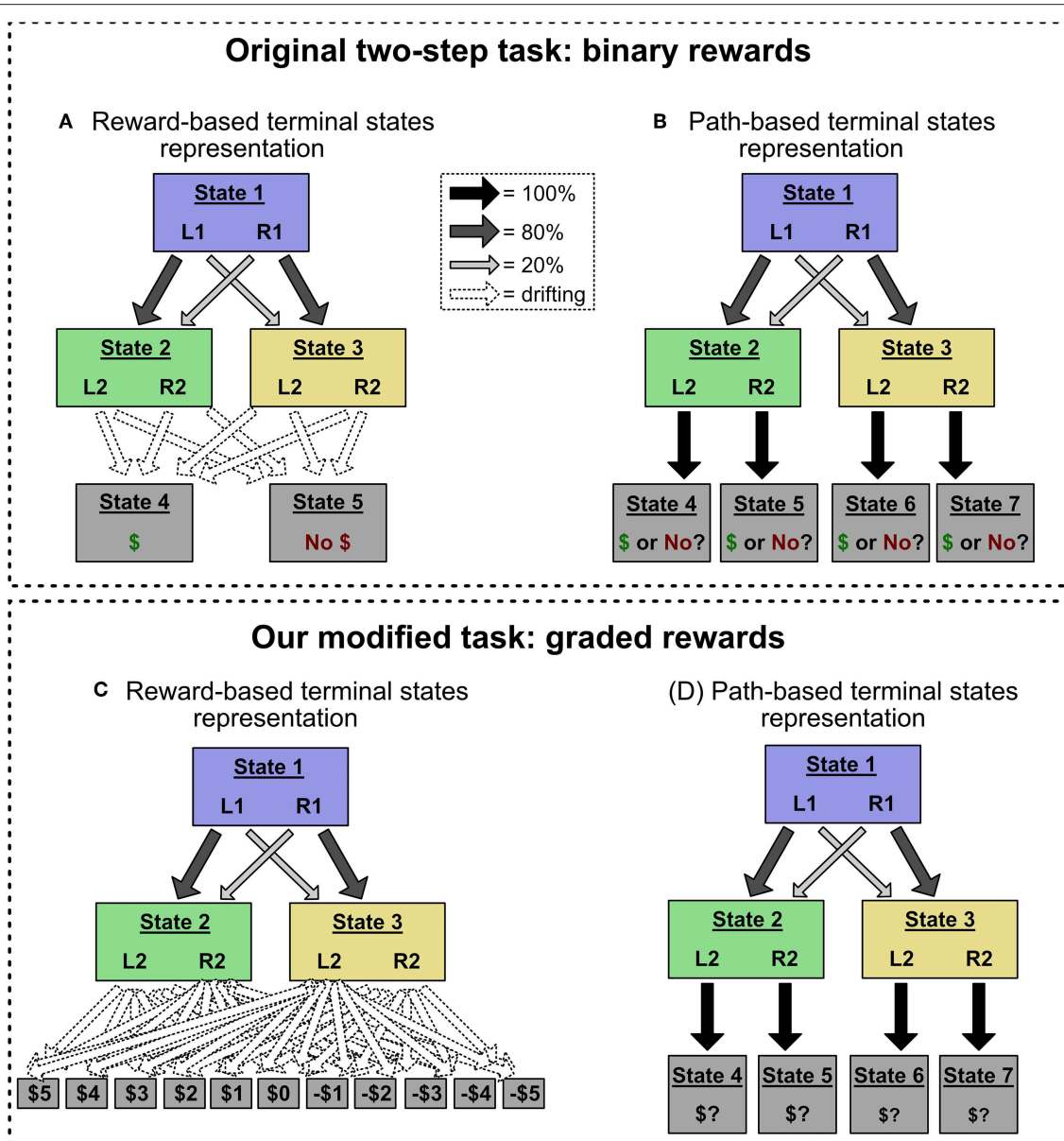


FIGURE 2 | (A,B) The original two-step task, which uses binary reward outcomes. This task is often represented with the structure in **(A)**, in which terminal states are defined by their reward values (e.g., State 4 gives a reward of 1, State 5 gives a reward of 0), and drifting reward probabilities are encoded as transition probabilities to those terminal states. However, the task can also be represented with the structure in **(B)**, in which each Stage 2 choice leads to a unique terminal state (choosing L2 in State 2 leads to State 4, and so on). In this alternate representation, drifting reward values are encoded as the value of those “path-based” terminal states. We show that action sequences can only mimic model-free choice patterns in the reward-based terminal state representation. **(C,D)** Our modified task, which uses graded reward outcomes (i.e., -5 through 5). Using graded reward outcomes precludes the reward-based terminal state representation, which would require eleven terminal states and forty-four transition probabilities (shown in **C**). Instead, this modified task induces the alternate, path-based terminal state representation (shown in **D**), allowing us to deconfound action sequences and model-free control.

R2, and receives a reward. How will that reward affect behavior on the next trial? A model-based agent will, using its internal model of the task, increase its value estimate of the Stage 1 action that typically leads to yellow: R1. [Formally, in Equation 1, the value of $Q(\text{State3}, R2)$ will get applied primarily to $Q(\text{State1}, R1)$, not $Q(\text{State1}, L1)$, because the former has a higher probability

of transitioning to State 3]. In contrast, a model-free agent, who has not represented the transition structure, will increase its value estimate of the Stage 1 action it chose: L1. In other words, a model-based agent's response to reward or no reward will depend on whether the preceding transition was rare or common; but a model-free agent will respond by becoming

more or less likely to repeat its last choice, no matter the transition type.

This logic leads to clear behavioral predictions. If an agent is model-based, the probability of repeating a choice will depend on the interaction between the reward type (reward vs. no reward) and the transition type (common vs. rare). In contrast, if an agent is model-free, the probability of repeating a choice will depend on the reward type only. When humans play the two-step task, they consistently show a mixture of both approaches (Glascher et al., 2010; Daw et al., 2011). They show both an interaction between reward and transition type (signature of MB RL), and a main effect of reward (signature of MF RL). The interpretation is that people are sometimes planners (captured by MB RL) and sometimes habitual (captured by MF RL). This finding is a pillar of support for the case that humans employ model-free RL in decision making.

2.3. The Action Sequences Critique

However, as DB show, the behavioral pattern in the original two-step task can be explained without invoking model-free RL. Instead, DB argue, people are employing model-based selection of chained action sequences (Dezfouli and Balleine, 2013). An action sequence is a series of actions that are precompiled into a single representation. For instance, a person tying her shoelaces does not consider each step in the sequence separately; rather, she simply chooses the abstract option "tie my shoes," and then executes the sequence of lower-level actions automatically. Similarly, a person driving to work may not consider each turn to be a new decision. Rather, she made only one decision, in which she chose the option "drive to work"; and the sequence of lower-level actions (e.g., start the car, turn left onto Cedar Street) followed automatically. Crucially, the action sequence model posits that the internal structure of the option is not guided by a value function; this is the key point of divergence with standard MF RL methods.

It is uncontroversial that people employ action sequences in some form (Dezfouli and Balleine, 2012, 2013; Dezfouli et al., 2014). We do not detail all the evidence DB marshal for the existence of action sequences. Rather, we focus on one key hypothesis: that action sequences can fully explain away any apparent role of MF RL in human behavior. Specifically, we focus on the claim that action sequences can produce the standard signature of MF RL in the two-step task.

2.3.1. Action Sequences in the Two-Step Task

On the action sequences model, when people make a Stage 1 choice, they employ model-based RL to choose between six possible options: the two single-step actions L1 and R1, and four action sequences L1-L2, L1-R2, R2-L1, and R2-L2 (**Figure 3A**). If a person chooses a single-step action like L1, she transitions to either the green screen or yellow screen and then uses that information to make her Stage 2 choice. But if a person chooses an action sequence like L1-L2, she selects L1 and then L2, no matter what screen she transitions to. In other words, she employs a form of "open-loop control" that is insensitive to information obtained during execution of the action sequence (Dezfouli and Balleine, 2012).

To see how the introduction of action sequences could explain seemingly model-free behavior in the two-step task, imagine that a participant chooses L1-L2, passes through yellow (rare transition), and receives a reward. Importantly, "receiving a reward" in the original paradigm is indicated by transitioning to a screen with a picture of money on it. This "rewarded" terminal state is not in any way specific to the path the person took to it—every unique sequence of actions terminates in one of two identical states: one that is rewarded, and another that is not. Thus, an agent who was insensitive to information obtained during the action sequence could learn from the reinforcement experience without ever referencing whether it had transitioned to the yellow or green state. All she would learn is that she had chosen the sequence L1-L2, and ended up at the screen with reward. (In figurative terms, when exiting "autopilot," she would know if she got money, but not where she had been). If rewarded, then, on the next trial, when consulting her internal model of the environment, she would become more likely to stay with L1-L2, not switch to R1-L2 (left-hand-side of **Figure 3B**). In this way, a purely model-based agent could mimic the signature of MF algorithms, and human behavior on the original two-step task can be explained without reference to MF RL.

In this paper, we demonstrate that action sequences can only produce MF-like behavior for this particular reward structure with binary outcomes. Then, in two experiments, we modify the two-step task to induce an alternate reward structure in which action sequences cannot produce MF-like behavior, and show that people still exhibit the behavioral signatures of MF RL. At the same time, our paradigm also produces unambiguous evidence that people do employ model-based control of action sequences. We conclude that people's habit-like behavior can be produced by both model-free RL and action sequences.

3. SIMULATIONS: ACTION SEQUENCES CAN ONLY MIMIC MF-LIKE BEHAVIOR FOR A PARTICULAR REWARD STRUCTURE

Although not previously emphasized, the action sequences model can only produce MF-like behavior in the original two-step task because the task has a peculiar property: The terminal reward conditions can plausibly be represented as two unified reward states (one for a reward, one for no reward), subject to drifting transition probabilities from each Stage 2 state-action pair (e.g., green-L2, yellow-R2). In other words, for any given action sequence that is selected at the beginning of the task, "reward probability" and "state transition probability" coincide perfectly—the relevant states are simply *defined* in terms of reward. For example, suppose an agent selects and executes the action sequence L1-L2, and that she then receives a reward. The result is encoded as an increased probability of L1-L2 transitioning to the "reward" state (i.e., State 4 in **Figure 3A**). Or, if she instead chooses R1-L2 and receives a reward, the result is again encoded as an increased probability of R1-L2 transitioning to the "reward" state. This representational scheme

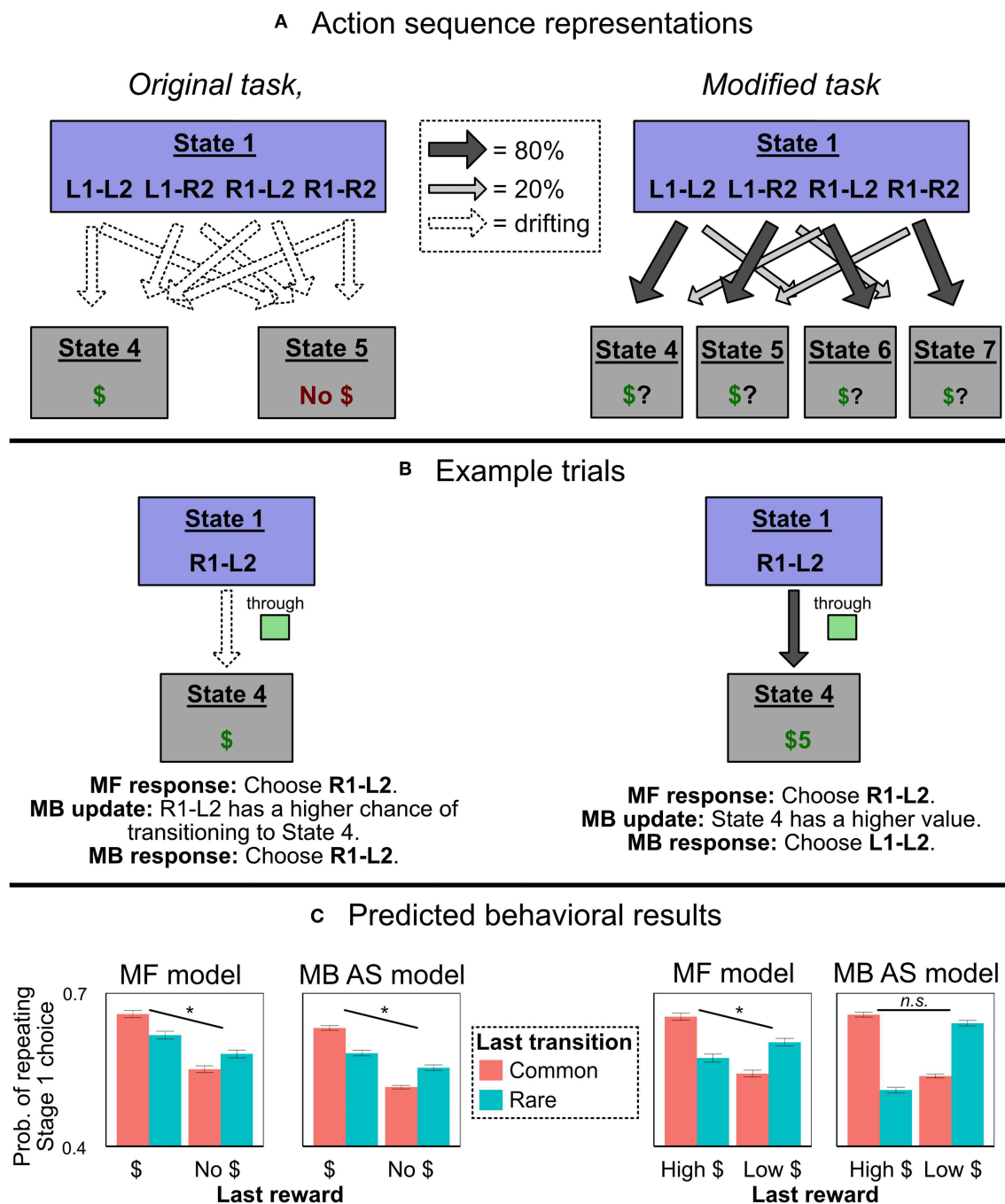


FIGURE 3 | Experiment 1 predictions. **(A)** How an agent using action sequences would represent the two task variants. In the original task with binary outcomes, they are represented as actions which lead directly, with drifting probabilities, to one of two reward states. In the modified task, they are represented as actions which lead, with fixed probabilities, to different terminal states with drifting values. **(B)** Example trials in which an agent chooses the action sequence R1-L2, receives a reward, and updates its beliefs. In the original task representation, both MF and MB controllers have the same response, allowing the model-based action sequences model to produce MF-like behavior; in the alternate representation, the MF and MB responses diverge. **(C)** Simulated probability of Stage 1 choice in the two representations, as a function of last trial's reward and transition type. We compared a traditional, flat model with partial MF control ("MF model") to an action sequences model with only MB control ("MB AS model"). The action sequence model produces MF-like behavior in the original representation, but not the alternate one. (Asterisks and "n.s." refer to the significance of the main effect of reward in each simulation. Error bars are ± 1 SEM).

has an important consequence: Model-based selection of action sequences is insensitive to the distinction between common and rare transitions.

Consider, however, an alternative representation of the reward structure (**Figure 3B**). Here, the current expected reward from each Stage 2 state-action pair is incorporated into the value of a

separate terminal state. For example, if the participant chooses R1-L2, passes through green, and receives a reward, she increases the value of the terminal state associated with green-L2 (State 4). Crucially, under this alternate task representation, model-based selection of action sequences cannot produce MF-like behavior. The critical test is: After choosing a sequence like R1-L2, passing through green, and receiving a reward, will she increase the probability of choosing R1-L2 (the MF-like option) or L1-L2 (the MB-like option)? Under the alternate representational scheme, a model-based planner will recognize that L1-L2 is more likely than R1-L2 to lead to the high-reward terminal state green-L2 (right-hand-side of **Figure 3C**). Thus, a model-based planner will not show the signature of model-free control, and cannot explain MF-like behavior in this version of the two-step task.

Put differently, in order for model-free and model-based controllers to make different behavioral predictions after a rare transition, the model-based controller needs to incorporate the fact that it was a rare transition into its post-trial update. When it chooses an action sequence, remains on autopilot through Stage 2, and arrives at an undifferentiated terminal state (the original task representation), the fact that it experienced a rare transition is not represented (explicitly or implicitly). But in the alternate task representation, the fact of the rare transition is encoded into the terminal state itself and, thus, it is naturally encoded in the MB controller's post-trial update.

In sum: The two-step task was designed to produce divergent behavior for model-free and model-based controllers after a rare transition. DB showed that, in the original task, a model-based controller with action sequences predicts the MF-like behavioral response (repeating the same Stage 1 choice after a rewarded rare transition). We show that this is only true for a "reward-based terminal state" representation of the task; in a "path-based terminal state" representation of the task, a MB controller with action sequences returns to predicting the MB-like, not MF-like, response³. We now report simulations confirming this theoretical analysis.

3.1. Methods

We simulated two algorithms: one that employed a weighted mixture of model-based and model-free control (the "MF

model"), and one that employed only model-based control but included action sequences (the "MB AS model"). In both algorithms, model-based and model-free Q-values were computed as described in section II; model-based Q-values⁴ were computed by recursively applying Equation (1), while model-free Q-values were computed via Q-learning (Equation 2). For the model-free Q-values, we included eligibility traces, with decay parameter λ . This means that, after participants chose an action in Stage 2, the reward prediction error was immediately "passed back" to update the Stage 1 action (discounted by λ ; see Sutton and Barto, 1998). (The presence of eligibility traces are critical for the analysis of the two-step task described above. Without eligibility traces, a reward on trial t would not immediately influence Stage 1 choice on trial $t + 1$; see Daw et al., 2011).

3.1.1. MF Model

In the MF model, agents estimate both model-based and model-free Q-values for single-step actions; these estimates must be integrated to ultimately produce a choice. How RL agents should, and how people do, arbitrate between model-based and model-free systems is a complex and important topic (Daw et al., 2005; Kool et al., 2017; Miller et al., 2019). Here, following past work (e.g., Daw et al., 2011; Cushman and Morris, 2015), we sidestep this question and assume that the model-based and model-free Q-values are ultimately combined with a mixture weight ω :

$$Q_{combined}(s, a) = \omega * Q_{MB}(s, a) + (1 - \omega) * Q_{MF}(s, a)$$

$\omega = 1$ leads to pure model-based control, and $\omega = 0$ leads to pure model-free control. This formalization is agnostic between different interpretations of the actual integration process, such as agents alternating between model-based and model-free systems on different trials, or agents estimating both types of Q-values on each trial and weighting them together. For a discussion of the distribution of ω values observed in our experiments, see the trial-level model fitting sections below. (For an in-depth analysis of the arbitration problem, see Kool et al., 2017).

After combining the model-based and model-free Q-values, agents chose actions with probability proportional to the exponent of the combined Q-values (plus a "stay bonus" capturing the tendency to repeat previous actions⁵). Formally, the probability of choosing action a in state s was given by a softmax function with inverse temperature parameter β , with a stay bonus v :

$$Prob(s, a) = \frac{e^{\beta * Q_{combined}(s, a) + v * 1_{a=a_{prev}}}}{\sum_{a' \in A} e^{\beta * Q_{combined}(s, a') + v * 1_{a'=a_{prev}}}}$$

⁴To compute the model-based Q-values, agents need an estimate of the transition function. Since participants in the experiments were explicitly told the transition probabilities and given practice with them, we assumed that participants would begin the task with an accurate estimate of the transition function. Thus, we gave agents an accurate model of T . The results do not change if we model agents as learning the transition probabilities dynamically.

⁵Note that, although we don't give it much attention here, some recent work theorizes that the stay bonus is actually a formalization of habits that is closer to our intuitive notion of what it means to be "habitual" (Miller et al., 2019).

³How does a path-based terminal state representation relate to the "Markov" assumption in reinforcement learning? Informally, the Markov assumption is that, after conditioning on the current state, the future is independent of the path taken to reach the current state. It is a key assumption in RL (Sutton and Barto, 1998). Path-based representations will still have the Markov property; if necessary, they can just build the path taken to reach a state into the representation of the state itself (e.g., if I sprint to my friend's house, the resulting state representation might include, not just "at my friend's house," but also "exhausted from the sprint"). This kind of augmented state representation is often necessary for complex applications of reinforcement learning (Sutton and Barto, 1998). However, in the tasks we use to induce the path-based representation, that kind of augmented state representation is actually not needed, because the different terminal states (States 4–7, **Figure 2**) are clearly differentiated from each other. Hence, the representation for State 4 does not need to explicitly include the information "I chose L1 and L2 to get here," because State 4 already clearly differs from the other terminal states. (Keeping with our analogy, the terminal states are more akin to different friends' houses). We still refer to this as a "path-based" terminal state representation only to emphasize that, unlike in the reward-based representation, the terminal states resulting from each Stage 2 choice path are different.

We used separate inverse temperature parameters for Stage 1 and Stage 2 choices. The MF model did not include action sequences.

3.1.2. MB AS Model

The MB AS model differed from the MF model in two ways. First, it employed only model-based Q-values to select actions (i.e., $\omega = 1$). Second, it included action sequences. In the MB AS model, in addition to being able to choose the two single-step actions in Stage 1, agents could also choose four additional action sequences: L1-L2, L1-R2, R1-L2, and R1-R2. Agents chose between all these options via a softmax function over the model-based Q-values (with a stay bonus). If the agent chose an action sequence in Stage 1, it executed the Stage 2 action automatically; if it chose a single-step action, then, at Stage 2, it made a second choice between the two single-step actions L2 and R2.

When using action sequences, there is a question of when to update their value estimates: Should an agent update its value estimate of an action sequence only after having selected it as an action sequence, or additionally after having chosen the single-step actions that happen to correspond to the sequence? Concretely, after choosing the single-step actions L1 and R2 (without invoking action-sequence control), should the agent then update the value representation associated with the action sequence L1-R2? Following Dezfouli and Balleine (2013), we present results assuming that the agent does update sequences after choosing their component actions; this assumption probably better captures how sequences are “crystallized” in real life (Dezfouli and Balleine, 2012; Dezfouli et al., 2014). However, all our results are similar if we assume the agent does not.

3.1.3. Parameter Values

Both models had a learning rate, two inverse temperatures, and a stay bonus. We used the same parameter distributions as Dezfouli and Balleine (2013). For each agent, the learning rate was randomly sampled from $Beta(1.1, 1.1)$; the inverse temperatures from $Gamma(1.2, 5)$; and the stay bonus from $Normal(0, 1)$. The MF model had two additional parameters: the mixture weight ω , which was sampled from $Uniform(0, 1)$, and the eligibility trace decay parameter λ , which was also sampled from $Uniform(0, 1)$.

We simulated 1,000 agents of each type playing each task variant (one with a reward-based terminal state representation, and one with a path-based terminal state representation). All agents performed 125 trials.

3.1.4. Analysis

Following the logic in section 2, we tested whether each model produced the signature of model-free control by estimating a logistic mixed effects models, regressing a dummy variable of whether they repeated their Stage 1 choice on (a) the last trial's transition type (common vs. rare), (b) the last trial's reward, and (c) their interaction. The classic signature of model-free control in this setting is a main effect of last trial's reward on Stage 1 choice.

3.2. Results

In the original task representation, both algorithms showed a main effect of reward on Stage 1 choice (left-hand-side of **Figure 3C**; for MF model, $p < 0.0001$; for MB AS model, $p < 0.0001$). But in the alternate representation, only the algorithm with model-free control showed a main effect of reward (right-hand-side of **Figure 3C**; for MF model, $p < 0.0001$; for MB AS model, $p = 0.57$).

3.3. Discussion

We simulated two algorithms—one that included model-free RL, and one that was purely model-based but included action sequences—and found the result predicted by our analysis. In the original task with a reward-based terminal state representation, model-based control of action sequences can mimic the signature of model-free control; but with a path-based terminal state representation, model-based control of action sequences cannot mimic model-free control. Thus, if people continue to show MF-like behavior in a version of the two-step task that induces the alternative representation, it would demonstrate that action sequences cannot account for all MF-like behavior.

In this simulation, we only reported the patterns of Stage 1 choices. Following past work (Daw et al., 2011; Cushman and Morris, 2015; Gillan et al., 2016), Stage 1 choice is the key variable we use to test for an effect of model-free control, and hence was the focus of this simulation. However, testing for an effect of model-free control is not our only goal; we also hope to show that people are simultaneously using action sequences (in Experiment 1), and to test whether those action sequences are themselves under model-free or model-based control (in Experiment 2). For those purposes, we will end up relying on two other outcome variables: Stage 2 choices, and the reaction times of Stage 2 choices. If people are employing action sequences, then their Stage 2 choices and reaction times will exhibit a unique pattern noted by Dezfouli and Balleine (2013) and described below in Experiment 1. Hence, Stage 2 choice and RT will be used in Experiment 1 to test for the presence of action sequences. Moreover, in the simulations for Experiment 2, we will show that Stage 2 choice and reaction time can also be used to distinguish between model-free and model-based control of action sequences. This logic will be described in section 5.

4. EXPERIMENT 1: DISAMBIGUATING CONTROL WITH A GRADED REWARD STRUCTURE

In our first experiment, we adopt a modified version of the two-step task that induces the “path-based terminal state” representation. In the original version, the amount of reward present in each terminal state was constant (e.g., 1 bonus point), and what drifted throughout the task was each Stage 2 state-action pair's probability of transitioning to the reward state vs. the non-rewarded state. For example, green-L2 might initially have a 75% chance of giving 1 bonus point, but later it might only have a 25% chance. This configuration supported the representation in **Figures 2A, 3A**, where drifting rewards are

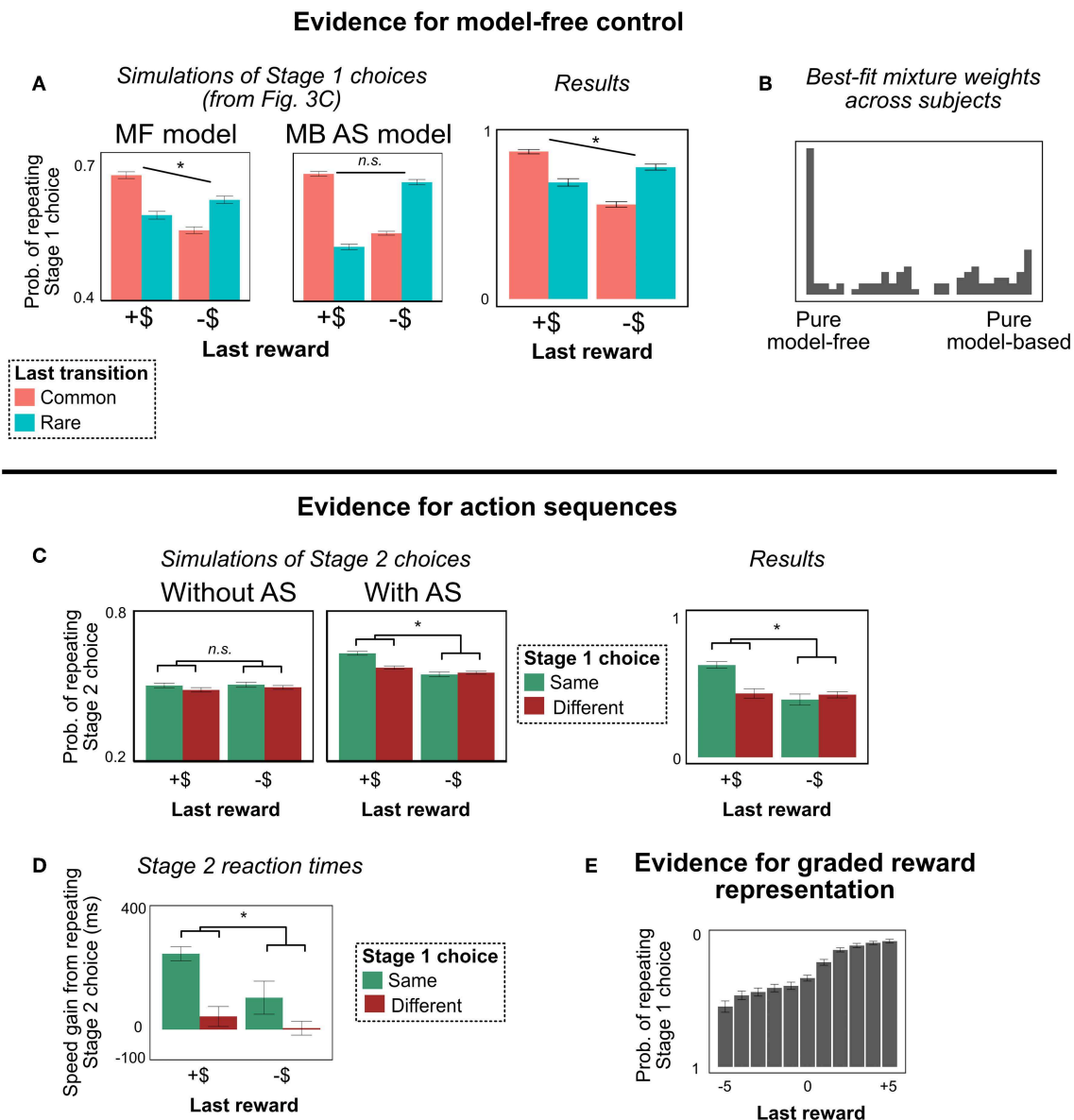


FIGURE 4 | Experiment 1 results. (For graphical convenience, we bin “last reward” into two categories: positive (“+\$”) and negative (“-\$”); we do not mean to imply that people actually represent the rewards this way. All statistical analyses are computed with unbinning reward). **(A)** Stage 1 choices as a function of last trial’s reward and transition type. People showed MF-like behavior (main effect of reward), even under the alternate task representation. **(B)** Histogram of best-fit mixture weights ω across subjects. People showed substantial model-free control. **(C)** Stage 2 choices as a function of last trial’s reward and this trial’s Stage 1 choice (on trials with a rare transition, following trials with a common transition). People’s decisions to repeat their Stage 2 choices were more correlated with their decisions to repeat their Stage 1 choices following a reward—a unique behavioral signature of action sequences (Dezfouli and Balleine, 2013). **(D)** Stage 2 reaction times. People are faster to repeat their Stage 2 action, and this effect is strongest for trials following a reward where they repeated their Stage 1 action. This pattern is another signature of action sequences. **(E)** Probability of repeating Stage 1 choice as a function of last trial’s unbinning reward. People are sensitive to the graded nature of the rewards, suggesting that they are not binning them into “positive”/“negative” categories. (This result is important for ensuring that people are employing a path-based, not reward-based, representation of the terminal states). All error bars are ± 1 SEM; asterisks indicate the significance of the main effect of last trial’s reward (in **A**), or the interaction between last trial’s reward and this trial’s Stage 1 choice (in **C, D**).

encoded as transitions probabilities to terminal states associated with “reward” or “no reward.”

In our version, rewards could take on a range of point values, and what drifted was the number of points associated with each Stage 2 state-action pair (Kool et al., 2016). For example, green-L2 might initially be worth 3 points, but later it might

be worth -4 points. (Point values were restricted to $[-5, 5]$ and drifted via a reflecting normal random walk with $\mu = 0, \sigma = 1.75$). This configuration induces the “path-based terminal state” representation in **Figure 2D**. To see why, imagine a person trying to use the original “reward-based terminal state” representation in our modified task. The person would have to represent eleven

separate terminal states (one for each possible point value), and forty-four terminal transition probabilities (**Figure 2C**). This would be a very inefficient representation and so we consider it unlikely. We further encouraged the path-based terminal state representation by reformatting the reward screen to clearly indicate which Stage 2 state-action pair had been chosen.

Thus, in this modified task, we assume that participants represent the task with path-based terminal states, and thus this task deconfounds the signatures of action sequences and model-free control. If, in this task, people still exhibit the behavioral signature of model-free control—a main effect of reward on subsequent Stage 1 choice—then it cannot be explained by model-based selection of action sequences.

Of course, graded rewards are not a new innovation, and have been used in several past studies (Cushman and Morris, 2015; Kool et al., 2016). Our contribution is to leverage graded rewards to deconfound the behavioral signatures of action sequences and MF RL. We collected new data, rather than reanalyze past studies, to ensure that the details of the task design were appropriate for the present question.

4.1. Methods

One hundred and one participants were recruited on Amazon Mechanical Turk. (We blocked duplicate IP addresses, only allowed IP addresses from the United States, and only used workers who had done over 100 previous studies on Turk with an overall approval rating of at least 95%). All participants gave informed consent, and the study was approved by Harvard's Committee on the Use of Human Subjects.

We used the version of the two-step task described in Kool et al. (2017), which has a cover story about spaceships to make the task more understandable. We explained the task in detail to participants, including explicitly telling them the transition structure. After being explained the task, participants completed 25 untimed practice trials which did not count toward their bonus payment. After the practice trials, participants were given a review of the task. Finally, they completed 125 real trials, in which each choice had a 2 s time limit.

Following Dezfouli and Balleine (2013), we did not counterbalance which side of the screen the actions appeared on; L1 was always on the left, R1 on the right, and so on. This feature maximizes the potential for participants to employ action sequences.

Participants were excluded if they completed the instructions in less than 1 min (suggesting that they did not read carefully); although the experiment was not pre-registered, this exclusion criterion was chosen in advance. Five participants were excluded, leaving 96 for the analyses.

We analyzed people's Stage 1 choices using logistic mixed effects models, regressing a dummy variable of whether they repeated their Stage 1 choice on (a) the last trial's transition type (common vs. rare), (b) the last trial's reward, and (c) their interaction. We included all random intercepts and slopes, and computed p -values with Wald z -tests. We estimated correlations between random effects, except in models with three-way interaction terms (where we disallowed random effect correlations to support model convergence). We report

unstandardized regression coefficients as b . (We analyzed people's Stage 2 choices similarly).

4.2. Results

4.2.1. Signature of Model-Free RL

The results of Experiment 1 are shown in **Figure 4**. People continue to show the signature pattern of MF-like behavior (**Figure 4A**). For Stage 1 choice, in addition to the interaction between last reward and transition type (signature of MB RL; $b = 0.29$, $z = 12.4$, $p < 0.0001$), there was a main effect of last reward (signature of MF RL; $b = 0.16$, $z = 8.4$, $p < 0.0001$). This result provides an example of MF-like behavior that cannot be explained by action sequences, and is the key finding of Experiment 1.

4.2.2. Concurrent Evidence for Action Sequences

Although action sequences cannot explain MF-like behavior in our task, we did find concurrent evidence that people are employing action sequences in this paradigm. This evidence is important, because it suggests that our task alteration did not discourage people from using action sequences; rather, people seem to employ MF RL and action sequences simultaneously. (We will also exploit these behavioral signatures in Experiment 2 in order to test for model-based vs. model-free control of action sequences).

The first piece of evidence for action sequences derives from logic originally presented by DB (Dezfouli and Balleine, 2013). Consider a trial in which a person chooses an action sequence, experiences a common transition, and receives a reward or punishment. She should be more likely to repeat the sequence on the following trial if she receives a reward, as opposed to a punishment⁶. Moreover, a consequence of her tendency to repeat the action sequence is that her decisions to repeat Stage 1 and Stage 2 actions will be correlated: If she repeats the same Stage 1 action, she will be more likely to repeat the same Stage 2 action. Putting these ideas together, a signature of action sequences is that repetition of Stage 1 and repetition of Stage 2 actions will be more correlated following a reward than following a punishment (simulations in **Figure 4C**).

This signature, however, is insufficient. There is an alternate explanation for it: Following a reward, a person using single-step actions should be more likely to repeat the same actions on the next trial. Thus, any factors which make her more likely to repeat her Stage 1 action—e.g., she was paying more attention on that trial—would make her more likely to repeat her Stage 2 action also, inducing a correlation.

To remedy this issue, we follow DB and restrict our analysis to trials in which the current Stage 2 state is different from the previous Stage 2 state (Dezfouli and Balleine, 2013). For instance, on the last trial, a person may have chosen R1-L2 and gone through the yellow state to State 6; but on this trial, the person may have chosen R1-L2 and, due to a rare transition, gone through the green state to State 4. If the correlation between

⁶This analysis is restricted to trials following a common transition, to avoid questions about model-based vs. model-free control of action sequences; that question is addressed in Experiment 2. Here, the goal is to demonstrate that people are using action sequences in some form.

Stage 1 and Stage 2 actions is due to action sequences, this restriction won't matter; people executing an action sequence are on "autopilot," and won't alter their behavior based on the Stage 2 state. But if the correlation is due to confounding factors like attention, then this restriction should eliminate the effect: A reward on the last trial would not influence a single-step agent's choice in a different Stage 2 state⁷.

As in prior work (Dezfouli and Balleine, 2013), people in our task showed precisely this pattern (interaction $b = 0.11$, $z = 7.4$, $p < 0.0001$), suggesting that they are indeed using action sequences (Figure 4C). However, the results in section 4.2.1 indicate that they are not relying on pure model-based control; there is a model-free influence on their choice. Hence, it appears that people are using both types of unplanned choice mechanisms: model-free control and action sequences. (For a discussion of how often people use each mechanism, see the section on trial-level model fitting below).

An additional signature of action sequences appears in participant's Stage 2 reaction times. While executing a sequence, people don't have to make any further decisions (e.g., to compare the values of alternative actions), and hence should be faster at selecting actions. This fact, combined with the effect above, leads to the following prediction. As described above, people tend to repeat the same sequence on consecutive trials following a reward. This implies that, on trials where they repeat their Stage 1 action, people should be faster to select a response in Stage 2 if they are repeating the same Stage 2 choice—i.e., if they are following the prescription of the action sequence. (We again restrict this analysis to trials following a common transition, with a different Stage 2 state than the previous trial). To test this prediction, we computed the difference in reaction times between trials when people repeated their Stage 2 action and trials when they didn't, conditioning on (a) whether they received a reward or punishment last trial, and (b) whether they repeated their Stage 1 choice. Replicating Dezfouli and Balleine (2013), we find the predicted interaction: People are faster when repeating their Stage 2 action, and this effect is strongest on trials following a reward where people chose the same Stage 1 action ($b = 14.9$, $t = 5.2$, $p < 0.0001$)⁸. The interaction is key: The fact that the boost in reaction time is stronger when participants chose the same Stage 1 action, and when they received a reward on the last trial, suggests that the effect is not due to an inherent time cost of switching Stage 2 actions (which would produce a main effect where people are always faster to choose the same Stage 2 action). The interaction is a unique signature of action sequences (Dezfouli and Balleine, 2013). (The raw reaction time data are presented in Figure A1 in Appendix).

⁷Moreover, if people are generalizing across Stage 2 states (i.e., blending together their value estimates for L2 in State 2 and in State 3), then they should exhibit a main effect of reward on Stage 2 choice—not the predicted interaction between reward and Stage 1 choice. See Dezfouli and Balleine (2013) for a more thorough justification for this test.

⁸For simplicity, we graph these results as a two-way interaction on the difference in reaction time. But, for the analysis, we properly test for a three-way interaction on the raw reaction times between last trial's reward, this trial's Stage 1 choice, and this trial's Stage 2 choice. Following DB, we did not apply a log transformation to RTs.

4.2.3. Are People Representing the Rewards as Graded?

As discussed above, the logic of Experiment 1 depends on people using the path-based, not reward-based, terminal state representation (Figure 2D, not Figure 2C). The reward-based terminal state approach is an implausible representation of a task with graded rewards. However, it is possible that, even though the task *has* graded rewards, people are not representing it that way; they could be representing the rewards as binary (e.g., either positive or negative). This possibility is problematic for our analysis, because it means that people could still be using the reward-based terminal state representation.

To demonstrate that people are treating the rewards as graded, we examined Stage 1 choices after rewards of each point value. (In this analysis, we focus exclusively on trials following common transitions). The results are shown in Figure 4E. People are clearly representing the full range of rewards, and not just binning them into positive or negative—every increase in point value is associated with an increase in stay probability. We tested this statistically by comparing two logistic mixed effects models: one that predicted Stage 1 choice from last trial's graded reward (i.e., the actual point value), and one that predicted choice from last trial's binned reward (i.e., either positive or negative). The former was heavily preferred (AIC of the former model was over 458 less than the AIC of the latter). Thus, the reward-based terminal state approach remains an implausible representation of our task⁹.

4.3. Trial-Level Model Fitting

As an additional analysis, we fit several variants of the model-free and action sequence models to participant choices at a trial level, and used Bayesian model selection to adjudicate between them. We fit five models: one model that did not use sequences, and four models that used sequences and employed different elements of MF and MB control (Table 1). For each model, we first estimated each subject's maximum *a posteriori* parameters, using the same priors as in the simulations and the *fmincon* function in MATLAB. To get each subject's best-fit parameters, we reran the optimization procedure ten times with randomly chosen parameter start values and selected the overall best-fitting values. We then used the Laplace approximation to compute the marginal likelihood for each subject for each model (Daw, 2011), and used the random-effects procedure of Rigoux et al. (2014) to estimate protected exceedance probabilities (PXPs)—i.e., the probability that each model is the most prevalent in the population.

The results are shown in Table 1. The preferred model used a mixture of model-free and model-based values to select both single-step actions and action sequences (PXP = 0.60), although it was closely followed by the model that did not use action sequences (but still used a mixture of MF and MB values to select single-step actions; PXP = 0.40). Analyzing the subject-level mixture weight ω , we find that subjects' behavior showed

⁹Of course, when graphing our results, we treat the rewards as binary (positive or negative). But this approach is just for graphical convenience, and does not imply that people are representing the rewards that way.

TABLE 1 | Models used in model comparison, and comparison results.

Model	Uses sequences	Control of single-step actions	Control of action sequences	PXP in Expt. 1	PXP in Expt. 2
No sequences	No	Mixture of MF and MB	N/A	0.40	0
Pure MB	Yes	MB	MB	0	0
Mixture-actions/MB-sequences	Yes	Mixture of MF and MB	MB	0	1
MB-actions/Mixture-sequences	Yes	MB	Mixture of MF and MB	0	0
Mixture-actions/Mixture-sequences	Yes	Mixture of MF and MB	Mixture of MF and MB	0.60	0

PXP stands for protected exceedance probabilities (Rigoux et al., 2014).

In Experiment 1, the preferred model used a mixture of model-free and model-based methods to evaluate both single-step actions and action sequences (although it was closely followed by a model that omitted action sequences entirely). In Experiment 2, the preferred model used a mixture of model-free and model-based methods to evaluate single-step actions, but only model-based methods to evaluate action sequences – consistent with the behavioral results of Experiment 2.

substantial model-free influence (**Figure 4B**); the mean weight was 0.44, and the distribution was peaked near 0 (pure model-free), with only 16% of subjects showing a weight greater than 0.9.

These results are consistent with our central claim that people are employing model-free control in this task. On the other hand, these results are mixed about whether people are using action sequences. Given the strong behavioral evidence in favor of action sequences, both in our experiments and past work (Dezfouli and Balleine, 2013), we think it likely that most subjects were using them; this inconsistency in the model-fitting suggests that the behavioral results may provide more reliable tests of our hypotheses (see Palminteri et al., 2017 for an in-depth argument in favor of this approach). Nonetheless, we include the model-fitting results here for completeness. We consider inconsistencies between the model-fitting and behavioral results in the section 6.

One question that model-fitting can help answer is how often people employ each choice mechanism. The mean mixture weight was 0.44. This number could mean different things depending on the interpretation of action selection. If people are employing model-free methods on some trials and model-based methods on others, then a mean ω of 0.44 indicates that people would on average be employing model-free RL on 56% of trials. If people are instead averaging model-free and model-based values together on each trial, then a mean ω of 0.44 indicates that model-free value makes up 56% of the final value estimate. We remain agnostic between these interpretations (Kool et al., 2017). Either way, there was substantial between-subject variation in mixture weights (**Figure 4B**).

A more difficult question to answer is on what percentage of trials people are using action sequences. We do not know when a person used an action sequence, nor it is a parameter explicitly

estimated in the model-fitting procedure (Dezfouli and Balleine, 2013). We leave this question to future research.

4.4. Discussion

We modified the two-step task to induce an alternate reward representation in which model-based selection of action sequences could not produce MF-like behavior. In this modified task, people still showed the same behavioral pattern, including the signature main effect of reward on Stage 1 choice. This analysis suggests that people are employing MF RL in some capacity—even in a task where they are also using action sequences. Here, action sequences and MF RL seem to be complements, not competitors.

One potential concern with this experiment is that people are using a different representation of the task than the one we assume (**Figure 2D**). We believe it is implausible that people are using the reward-based terminal state representation assumed by Dezfouli and Balleine (2013) (**Figure 2C**); however, there is another representation that could be problematic for our analysis. Specifically, people might be collapsing States 4–7 into one undifferentiated terminal state, with the rewards encoded into the preceding actions—e.g., the reward in State 4 might actually be encoded as the reward from choosing L2 in State 2, with the terminal state ignored. This representation would be problematic for our analysis because a model-based action sequence controller could plausibly, after exiting the action sequence, be aware of the reward it received without being aware of the path it took to get that reward. Hence, a model-based action sequence controller could ignore the transition type, and mimic model-free behavior¹⁰.

A similar worry goes as follows. Even if people are using our assumed path-based terminal state representation (**Figure 2D**), a MB controller could still in principle select some type of “extended” action sequence that ignores the identity of the terminal state. For instance, imagine a MB controller chooses L1–L2, gets a rare transition in the middle of the sequence, receives a reward, and ignores the associated terminal state (e.g., State 6) because it is still on “autopilot.” This model-based controller would credit the reward to the sequence L1–L2 itself, and not to the and would thus appear model-free. This concern makes it seem as if a model-based controller can still mimic a model-free one in our task.

A priori, there is some reason to doubt these concerns. We clearly differentiated the four terminal states with unique visual features, which included an image of the last action taken to reach that terminal state. Moreover, people could not quickly pass through the screen indicating the terminal state; they were required to remain on that screen for several seconds. If an action sequence controller is ignoring all this easily-accessible information about the transition structure and instead crediting the reward directly to the action sequence itself, then it is not obvious that the controller is still model-based. It is showing no sensitivity to the task’s transition structure, and instead caching value directly to actions themselves—the definition of a model-free controller.

¹⁰We thank a reviewer for raising this point.

Nonetheless, we seek direct evidence against this possibility. Experiment 1 tested for the presence of model-free control, but it was not designed to test which type of controller was being used to select action sequences specifically. In the next experiment, we modify the design to produce a unique behavioral signature of model-free and model-based control of action sequences. This design allows us to address the aforementioned concern in the following way. If people are using an “unresponsive” model-based controller to select action sequences in a way that ignores the identity of the terminal state and mimics model-free control (e.g., through an undifferentiated terminal state representation, or an “extended” sequence), then we should find evidence of apparent model-free control of action sequences. Conversely, if we find no evidence of apparent model-free control of action sequences, that result would suggest that people do not use an “unresponsive” model-based controller for this family of tasks—and hence that the MF-like behavior in Experiment 1 was not produced by such a controller, and was genuinely model-free¹¹.

To preview our results, we find that people are *not* exhibiting apparent model-free control of action sequences; they instead produce the behavioral signature of accurate model-based control of action sequences (with knowledge of the differentiated terminal states). Yet, they still exhibit a signature of some type of model-free control. Together, this pattern suggests that people are not exhibiting apparent model-free control via an unresponsive model-based action sequence controller; rather, they are exhibiting genuine model-free control of single-step actions.

5. EXPERIMENT 2: TESTING FOR MODEL-BASED VS. MODEL-FREE CONTROL OF ACTION SEQUENCES

Experiment 2 was designed to answer the question: Which type of controller is being used to select action sequences? In principle, MB and MF control can be applied to both single-step actions and action sequences (Figure 1). Models of choice in the two-step task commonly assume that single-step actions are controlled by a mixture of model-based and model-free control (green box in Figure 1; Glascher et al., 2010; Daw et al., 2011; Kool et al., 2017). But what about action sequences? DB posited that action sequences would be chosen exclusively by MB control, but their paradigm did not allow them to test this claim. By using graded rewards in a modified task structure, we can test for unique behavioral signatures of model-based and model-free control of action sequences. We find strong evidence that, as DB predicted, action sequences are under model-based control. In contrast, although we find clear evidence that people are employing *some* type of model-free control, we find no evidence that they are using model-free RL to select action sequences. This result helps address the concern from Experiment 1—that an unresponsive

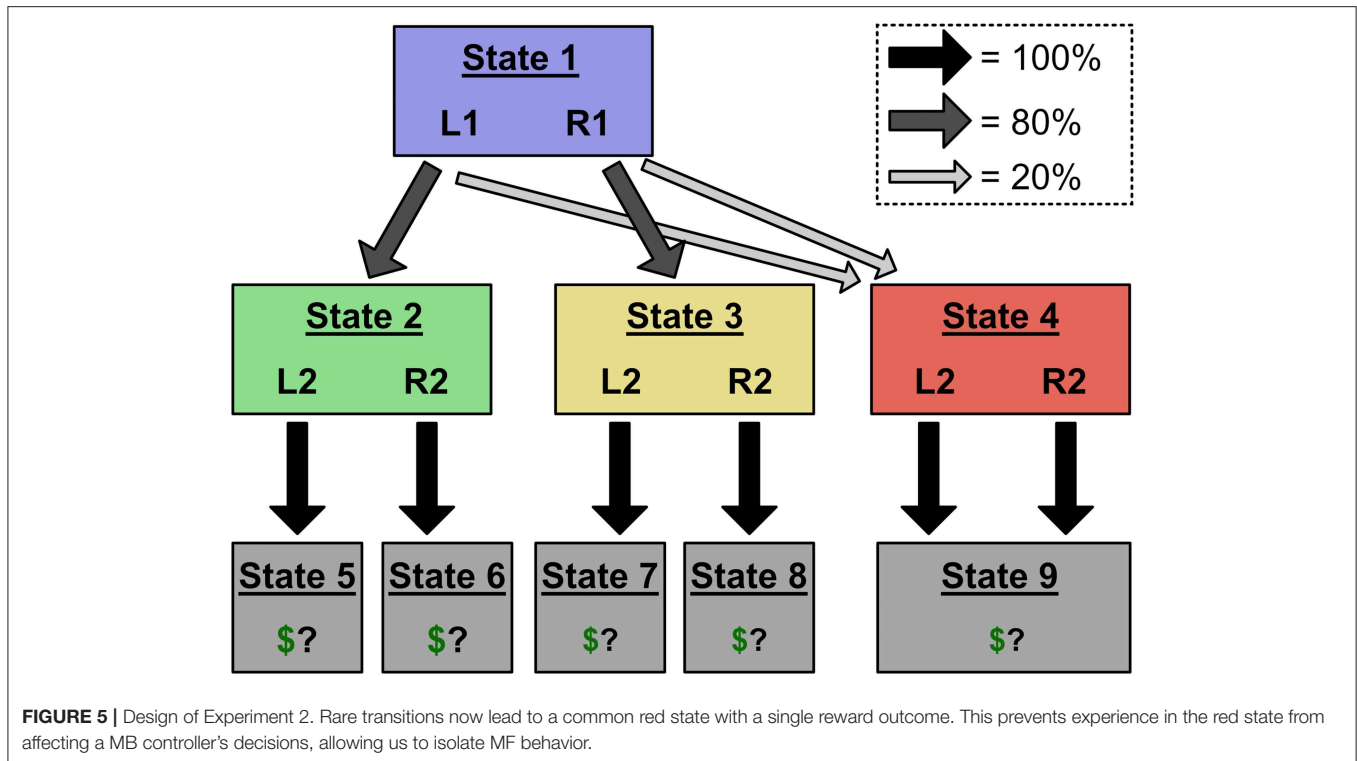
model-based action sequence controller was mimicking model-free control—by simultaneously demonstrating (a) model-free control of some type, but (b) no apparent model-free control of action sequences. More broadly, this result suggests that two types of habitual mechanisms coexist in this paradigm: (accurate) model-based selection of action sequences, and model-free control of single-step actions.

5.1. Logic of Experiment 2

The second experiment differed from the first only in the transition structure between Stages 1 and 2 (Figure 5). As before, L1 and R1 had an 80% chance of transitioning to the green and yellow states, respectively. But in Experiment 2, both actions have a 20% chance of transitioning to a novel red state (State 4). Since both Stage 1 actions have the same chance of transitioning to the red state, the value of State 9 should not influence a model-based controller's choices in Stage 1; a model-based controller will integrate out any experience it has in the red state, and be unaffected by feedback from State 9. This fact, combined with the effect of action sequences on Stage 2 choices and reaction times described in Expt. 1, elicits unique behavioral predictions for model-based and model-free selection of action sequences.

The key to Experiment 2 is that, following trials with transitions to the red state, a person using model-free control to select action sequences will show the Stage 2 action sequence effects, while a person using model-based control to select sequences will not (Figure 6A; right-hand-side of Figure 6B). Recall from Experiment 1 that, if a person is using action sequences, their Stage 2 choices will be predicted by a positive two-way interaction between last trial's reward and this trial's Stage 1 choice: They will be more likely to repeat their Stage 2 choice after being rewarded last trial and repeating their Stage 1 choice this trial (Figure 4C). [As described above, this interaction occurs because people will be most likely to be repeating an action sequence, and hence to repeat their Stage 2 choice, following reward and repeated Stage 1 choice. The same is true for their reaction times: They will be *fastest* to repeat their Stage 2 choice after being rewarded last trial and repeating their Stage 1 choice this trial. See Figure 4D. As in Experiment 1, we rule out confounds by restricting this analysis to trials in which the Stage 2 state differs from the previous trial, which does not matter for an action sequence controller because it is insensitive to transitions while executing the sequence (Dezfouli and Balleine, 2013)]. Experiment 2 combines this fact with a design ensuring that only a model-free controller will be affected by reinforcement after a rare transition; a model-based controller will ignore the reinforcement (Figure 5). In Experiment 2, if people are using model-free control of action sequences, they will show the signature of action sequences (the two-way interaction of last trial's reward and this trial's Stage 1 choice on Stage 2 choice/reaction time) after a rare transition; but if they are using model-based control of action sequences, they won't show this signature. (Both controllers will show the signature after common transitions; left-hand-side of Figure 6B). Hence, if people exhibit this two-way interaction after both common and rare transitions, we can infer that their action sequences are under some degree of model-free control. In contrast, if

¹¹Though the model-fitting results in Experiment 1 suggested that people were showing apparent model-free control of action sequences, Experiment 1 was not designed to address this question, and we prefer to address the question with clear behavioral predictions (Palminteri et al., 2017). We review and consider inconsistencies in the model-fitting results in section 6.



people exhibit this two-way interaction in common but not rare transitions, we can infer that their action sequences are under model-based control. (And if people exhibit the interaction after neither type of trial, we would infer that they are not using action sequences at all). These effects are summarized in **Table 2**.

Note that, if people are using model-based control of action sequences, we can go one step further in our analysis. As just discussed, we predict that in this case people would show the two-way interaction after common transitions but not rare transitions. Statistically, this means that they would show a significant two-way interaction after common transitions, a null effect for the interaction after rare transitions—and, critically, a significant *three-way* interaction when including common vs. rare transition as an additional regressor. In other words, they will show a significantly stronger interaction after common transitions than after rare transitions. This result would provide positive evidence for model-based control of action sequences that goes beyond a null effect after rare transitions.

To preview our results, we find precisely the patterns predicted by model-based control of action sequences: People show the signature two-way interaction (in both Stage 2 choices and reaction times) after common transitions but not rare transitions, and show a three-way interaction when including transition type as a regressor. This is strong evidence that, at the higher level of the action hierarchy (i.e., action sequences), people in this paradigm employ model-based control. At the same time, we find concurrent evidence that people are employing model-free control at some point in their decision making. Hence, our results again suggest that model-free control and action sequences coexist in people's decision making process, and that, at

least in this paradigm, model-free control may be more strongly applied at lower levels of the action hierarchy.

5.2. Simulations

We confirm this analysis by simulating agents performing the task in Experiment 2 (**Figure 6C**). We used the same methods as in the prior simulations, with one change. The two algorithms now both used a mixture of model-free and model-based Q-values to assign value to single-step actions (e.g., L1, R1), and both employed action sequences (e.g., L2-R2); they differed only in the type of value assignment to action sequences. One algorithm used model-free Q-values to assign value to action sequences ("MF AS"), while the other algorithm used model-based Q-values ("MB AS").

The results confirmed our theoretical analysis. After trials with a common transition, both MF AS and MB AS agents showed the predicted two-way interaction: Their Stage 2 choices were predicted by their Stage 1 choices times last trial's reward (p 's < 0.0001; left-hand-side of **Figure 6C**). In contrast, after trials with a rare transition, only MF AS agents showed the two-way interaction (p < 0.0001); MB AS agents showed no interaction (p = 0.71; Bayes factor in favor of null is 88; right-hand-side of **Figure 6C**). Moreover, when including last trial's transition type as a regressor, MB AS agents showed the predicted three-way interaction (p < 0.0001). These simulation results confirm the theoretical analysis above, and demonstrate that this paradigm can detect unique effects of model-free and model-based action sequence control. Next, we test for these effects empirically.

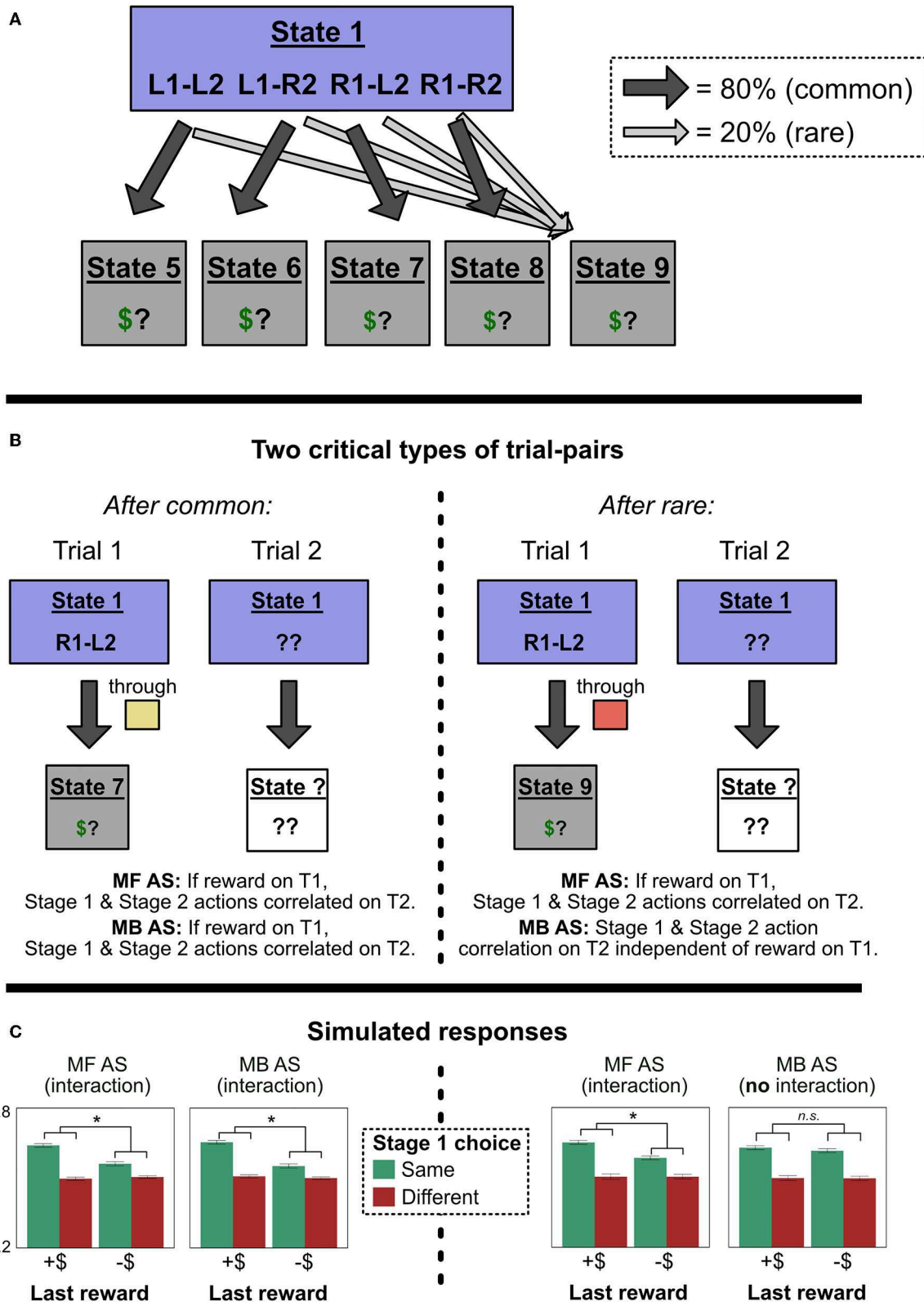


FIGURE 6 | Predictions in Experiment 2. **(A)** The representation of Experiment 2 by an agent using action sequences. All sequences have an 80% chance of leading to their respective terminal states, and a 20% chance of leading to State 9. This design ensures that a model-based controller's decisions will not be influenced by the value of State 9. **(B)** Two types of critical trials. On the left, we analyze trials following common transitions. Here, both model-free and model-based action sequence

(Continued)

FIGURE 6 | models (MF AS and MB AS) predict an interaction between Stage 1 choice and last trial's reward on Stage 2 choice. On the right, we analyze trials following rare transitions. Here, MF AS predicts the same interaction, but MB AS predicts that the interaction should disappear, because the value of State 9 will not matter for sequence selection. (This analysis is restricted to instances in which Trial 2 has a different Stage 2 state than Trial 1; this restriction rules out confounds described in Experiment 1). **(C)** Simulations to confirm the predictions in **(B)**. All error bars are ± 1 SEM; asterisks indicate significant interactions.

TABLE 2 | Key predictions in Experiment 2.

Sequence controller	Predicted pattern after common transitions	Predicted pattern after rare transitions
Pure MF, or mixture of MF/MB	2-way interactions of S2 choice \sim S1 choice * Last reward, S2 RT drop \sim S1 choice * Last reward	2-way interaction of S2 choice \sim S1 choice * Last reward, S2 RT drop \sim S1 choice * Last reward
Pure MB	2-way interaction of S2 choice \sim S1 choice * Last reward, S2 RT drop \sim S1 choice * Last reward	No 2-way interactions
No sequences	No 2-way interactions	No 2-way interactions

S1 stands for Stage 1; S2 stands for Stage 2; S2 RT drop indicates the predicted gain in speed (and hence drop in reaction time) from repeating a Stage 2 choice. If a person is using model-free control of action sequences, they will show the action sequences' signature two-way interactions (Stage 2 choice/RT drop \sim Stage 1 choice * Last trial's reward) after both a common and rare transition. But if a person is using model-based control of action sequences, they will show the interactions after common but not rare transitions, and hence will show three-way interactions of Stage 2 choice/RT drop \sim Stage 1 choice * Last reward * Last transition type.

5.3. Methods

Three hundred participants were recruited on Amazon Mechanical Turk, using the same filtering criteria as in Experiment 1. The task was identical to Experiment 1, except for the change in the state/transition structure. We excluded 18 participants who finished the instructions in less than 1 min, and 1 participant for whom the study severely glitched.

In the instructions, we emphasized to people that the transition probabilities to the rare state did not change over the course of the experiment, and that when a rare transition happened was completely random with no way to plan for it. To ensure that participants believed this key part of the experimental design, we added a question at the end of the experiment: "Did you believe that, on any given round, the two Stage 1 choices had the same probability of transitioning to the red state?" (We also added a second question: "Did you believe that, on any given round, the two actions in the red state always led to the same amount of bonus money?" The significance of this belief is discussed below). We excluded an additional 84 participants who answered "No" to either of these questions, leaving 197 participants for analysis. Also, at the end of the instructions, we included three comprehension check questions, asking, for each Stage 2 state, which of the Stage 1 actions was most likely to reach it (or whether both actions were equally likely). Participants generally understood the transition structure: The percentage of participants giving correct answers for the three Stage 2 states were (in order): 87, 94, and 95%. If participants got the comprehension check question wrong, they were told the

correct answer and reminded of the transition structure (but not excluded). Again, although these results were not pre-registered, all exclusion criteria were chosen in advance.

As can be seen in **Figure 5**, both actions in the red state lead to the same outcome; participants were told this fact explicitly. This design feature ensured that all action sequences had the same probability of transitioning to State 9, and that a model-based controller would not incorporate information from rare-transition trials into its subsequent Stage 1 choice.

All statistical methods are similar to those in Experiment 1. Bayes factors were computed with a BIC approximation (Wagenmakers, 2007).

5.4. Results

5.4.1. Evidence for Model-Free RL

First, we conceptually replicate the finding from Experiment 1 that model-free RL influences choice. In this paradigm, the signature of MF RL is simple (Cushman and Morris, 2015). If people are using MF RL, their Stage 1 choice should be influenced by the reward received on a rare transition; they should be more likely to repeat their Stage 1 choice after a reward in State 9, compared to a punishment. But if they are using only model-based RL (with or without action sequences), their Stage 1 choice should not be influenced by the value of State 9 (simulations in **Figure 7A**).

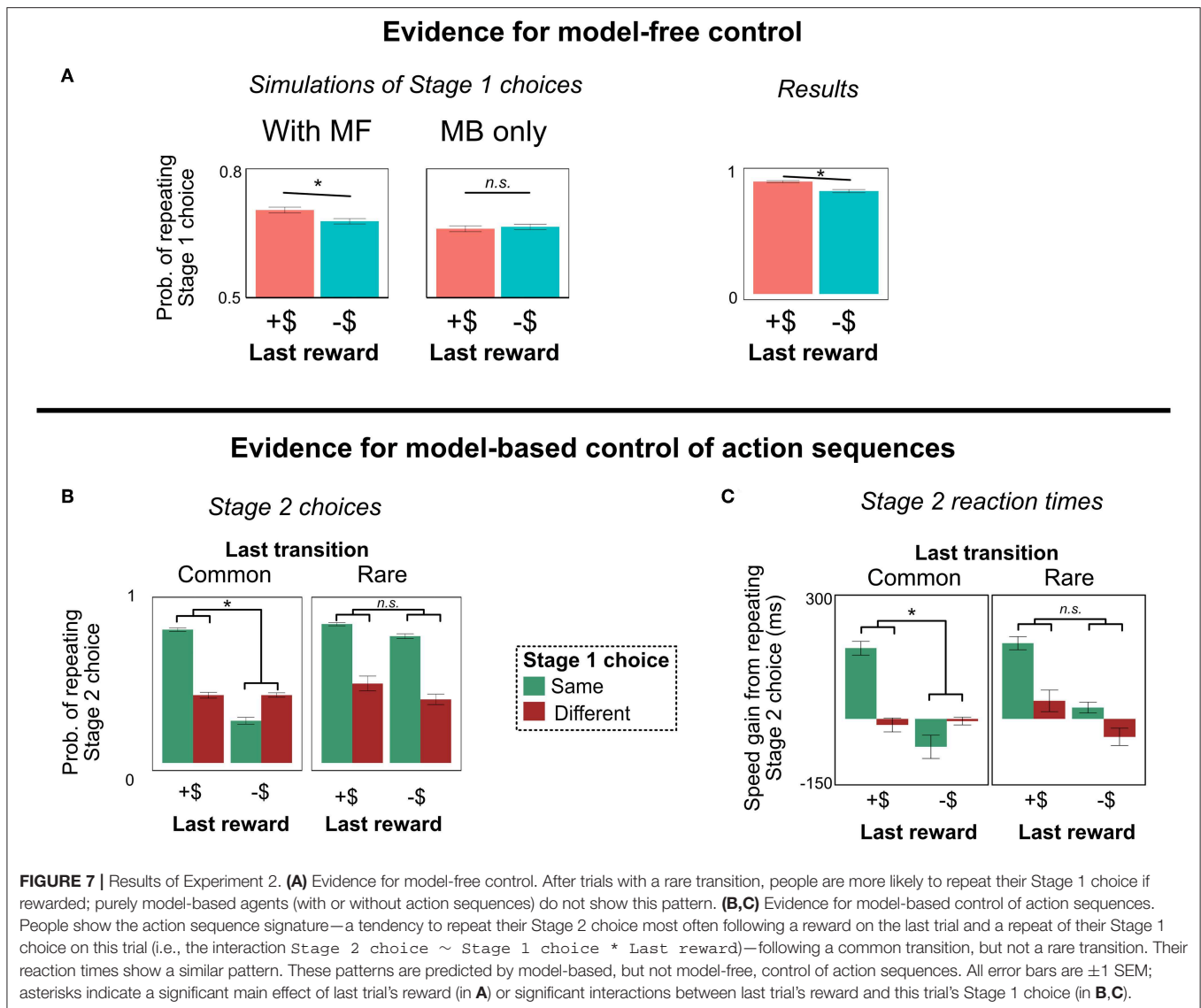
Indeed, people show the signature of MF RL (**Figure 7A**). They are more likely to repeat their Stage 1 choice following a more positive reinforcement in State 9 (main effect of last reward; $b = 0.20$, $z = 7.0$, $p < 0.0001$). This result demonstrates that MF RL influences people's choice in some way in this paradigm.

5.4.2. Evidence for Model-Based Selection of Action Sequences

Second, we turn to the main question of Experiment 2: Does model-free RL influence people's choices of action sequences, or are action sequences controlled primarily by model-based RL?

In this paradigm, people seem to choose action-sequences primarily through model-based RL (**Figure 7B**). As predicted by MB RL, in trials following a rare transition, there is no interaction on Stage 2 choice between Stage 1 choice and last reward (interaction term, $b = 0.022$, $z = 1.1$, $p = 0.27$, $BF_{null} = 41$). This result suggests that people are not using model-free RL to select action sequences.

Moreover, we find positive evidence for model-based control of sequences. Regressing people's Stage 2 choices on (a) their Stage 1 choice, (b) last trial's reward, and (c) last trial's transition type, we find the predicted three-way interaction: People show the signature of action sequences [an interaction between (a) and (b)] more in trials following a common transition, compared to trials following a rare transition (**Figure 7B**; interaction term, $b =$



0.44, $z = 10.7$, $p < 0.0001$). This is precisely the pattern predicted by model-based control of action sequences.

A similar signature of model-based control of action sequences comes from people's reaction times in Stage 2. As described above, if people are using model-based control of action sequences, then their gain in speed from repeating their Stage 2 choice should be predicted by the two-way interaction of their Stage 1 choice and last trial's reward—but only following common, not rare, transitions. And indeed, people exhibit precisely this pattern (**Figure 7C**): They showed the predicted interaction after common transitions ($b = 48.7$, $t = 10.8$, $p < 0.0001$), no interaction after rare transitions ($b = -7.1$, $t = -1.1$, $p = 0.26$ —although the Bayes factor was weak, $BF_{null} = 2.5$), and a significant interaction between those two effects when including transition type as a regressor ($b = 55.2$, $t = 6.8$, $p < 0.0001$)¹².

¹²As in Experiment 1, for readability, we describe and graph the RT effects with "Stage 2 RT drop from repeating Stage 2 choice" as the

5.5. Trial-Level Model-Fitting

As an additional analysis, we fit the same models from Experiment 1 to trial-level choices in Experiment 2 (using identical procedures as before). The preferred model used a mixture of model-free and model-based methods to evaluate single-step actions, but only model-based methods to evaluate action sequences ($PXP = 0.999$). This result is consistent with the behavioral results in Experiment 2. (On the other hand, it is

dependent variable. But in our actual analysis, we test the effects with raw Stage 2 RT as the dependent variable, and Stage 2 choice as an additional regressor interacting with the others. Hence, if people are using model-based control of action sequences, they will properly show a *three-way* interaction of Stage 2 RT ~ Stage 1 choice * Last reward * Stage 2 choice after common but not rare transitions, and hence a *four-way* interaction of Stage 2 RT ~ Stage 1 choice * Last reward * Stage 2 choice * Last transition type. These interactions are what we report here.

inconsistent with the model-fitting result in Experiment 1. We return to this issue in section 6).

5.6. Discussion

We replicated the finding from Experiment 1 that people are employing model-free control of some sort. Moreover, we found evidence that people's choice of action sequences was under model-based, and not model-free, control. People's patterns of Stage 2 choice qualitatively matched the simulated behavior of agents using model-based control of action sequences. Additionally, the best-fitting model used a mixture of model-free and model-based control to select single-step actions, but only model-based control to select action sequences.

These results validate the hypothesis of Dezfouli and Balleine (2013) that action sequences would be under model-based control. On the other hand, they further reinforce our primary claim that model-free RL is part of people's decision making repertoire, and not explained away by model-based control of action sequences. In particular, they provided evidence against the concern raised at the end of Experiment 1: that people could be using an unresponsive model-based action sequence controller which mimics model-free control. If that were the case, we would have seen evidence of apparent model-free control of action sequences in Experiment 2. Instead, we find that people select action sequences using an accurate model-based method, but select single-step actions with some degree of model-free control.

We do not make the strong claim that people never exhibit unresponsive model-based control of action sequences, or genuine model-free control of action sequences. It is difficult to draw strong conclusions from a null result. Nonetheless, in our paradigm, model-free control appears to be applied primarily to lower levels of the action hierarchy. We return to this question in section 6.

One worry with this experiment is that it depends on people believing that the two Stage 1 actions had the same probability of transitioning to the rare state on each trial. If people were committing a "hot hands" fallacy and believing that a Stage 1 action that produced a rare transition last trial was more likely to produce one this trial, that mistaken belief could potentially produce apparent model-free behavior (Gilovich et al., 1985). We mitigated this risk by repeatedly emphasizing to people that the transition probabilities did not change from trial to trial, and that each rare transition was unpredictable and independent of the others. Moreover, we excluded participants who reported not believing this fact. Nonetheless, it is possible that this belief persisted in polluting our data. Future work should rule out this potential confound more thoroughly.

6. GENERAL DISCUSSION

Our work aligns with many prior studies arguing that some form of model-free RL is implemented by humans. Model-free RL has proved a successful model of human and animal behavior in sequential decision tasks (Dolan and Dayan, 2013), phasic dopamine responses in primate basal ganglia (Schultz et al., 1997), fMRI patterns during decision making (Glascher

et al., 2010), and more. We defend this model against a recent critique (Dezfouli and Balleine, 2012, 2013; Dezfouli et al., 2014) by providing unconfounded evidence that, in a variant of the popular two-step task, people *do* employ model-free RL, and not just model-based control of chained action sequences.

At the same time, our work provides strong evidence that, in addition to model-free RL, people indeed employ model-based control over action sequences. This result suggests that the puzzle of habits will not be solved by one model; "habits" likely comprise multiple decision strategies, including both model-free RL and action sequences.

6.1. Relationship Between Behavioral and Model-Fitting Results

We presented two types of evidence: one-trial-back behavioral effects (e.g., the effect of last trial's reinforcement on this trial's choice), and model-fitting results. In general, these methods were in agreement. In Experiment 1, both methods indicated that people were employing model-free RL and (generally) action sequences. In Experiment 2, both methods indicated that people were using model-free RL to evaluate single-step actions, but only model-based RL to evaluate action sequences. This concordance reinforces those claims.

There were, however, two points on which the model-fitting results were inconsistent. First, in Experiment 1, the model-fitting suggested that many of the participants were not actually using action sequences. This is possible, but seems unlikely in light of our clear behavioral results and the results of past work (Dezfouli and Balleine, 2012, 2013; Dezfouli et al., 2014). Second, the preferred model differed between Experiments 1 and 2. In Experiment 2, the preferred model used only model-based RL to evaluate action sequences, but in Experiment 1 the preferred model used both model-based and model-free RL to evaluate them. It is possible that participants in Experiment 1 were actually using more model-free control of action sequences than in Experiment 2. On the other hand, since it was Experiment 2 that was designed to test for the type of action sequence controller, the preferred model in Experiment 2 is probably more informative on this point. In any case, the inconsistency casts doubt on the reliability of the model-fitting approach for answering these questions. We believe that our clear patterns of qualitative behavioral results are stronger evidence for our claims than the model-fitting results; for a detailed discussion of this point, see Palminteri et al. (2017).

6.2. Could MF-Like Behavior Be Produced by Model-Based Algorithms With Inaccurate Beliefs?

We presented evidence for model-free control in human behavior that is deconfounded from one potential alternative: model-based control of action sequences. There are, however, other model-based algorithms that could mimic model-free control by having an inaccurate model of the task. For instance, consider a person in Experiment 1 who believes that rare transitions lead to unique Stage 2 states—e.g., that a rare transition from L1 leads to a different state than a common

transition from R1 (**Figure A2A** in the Appendix). A model-based agent with this task model would produce MF-like behavior because it would be more likely to repeat its Stage 1 choice following both common and rare transitions (since a reward from a rare transition no longer suggests that the agent should switch its Stage 1 choice; **Figure A2B**). Other examples of inaccurate models that can produce MF-like behavior are given by da Silva and Hare (2019). In the most extreme case, MF-like behavior in this task can always be mimicked by an algorithm that ignores the task instructions and builds a transition model of the form “repeating behavior after being rewarded leads to more money at the end of the experiment.”

There is some reason to doubt that MF-like behavior can be explained this way, as an “inaccurate model-based” controller. A key feature of model-free RL is its computational simplicity (relative to model-based RL). This feature helps make sense of why people would exhibit MF-like behavior relatively more when under cognitive load (Otto et al., 2013), or when the financial stakes are lower (Kool et al., 2017). These results are more difficult to explain under an “inaccurate-model-based” account, since it is not clear that using an inaccurate model of the task is more computationally efficient. Moreover, there is strong neural evidence for model-free RL that is difficult to explain under an inaccurate-model-based account (Schultz et al., 1997; Dolan and Dayan, 2013).

However, this is an active area of debate (da Silva and Hare, 2019). Here, we do not rule out all the inaccurate-model-based alternative accounts of our behavioral results, or provide definitive evidence for model-free RL. We instead make the more modest claim that the signature of model-free RL observed here is not due to model-based control of action sequences.

6.3. At What Level of Abstraction Does Model-Free RL Operate?

Our results contribute to an ongoing investigation into the scope of model-free RL. Model-free RL—and habits in general—are often characterized as applying to relatively concrete actions (e.g., a rat pulling a lever, or a human pushing a button). But some research has suggested that MF RL can also apply to relatively abstract “actions”, like goal selection (Cushman and Morris, 2015) or working memory gating (O’Reilly and Frank, 2006).

Here, we tackled the question of whether MF RL also applies to the control of another type of abstract action: action sequences. We found no evidence that people used model-free RL for action sequences. Rather, in Experiment 2, we found strong evidence that people used model-based RL to evaluate sequences. This result aligns with the predictions of Dezfouli and Balleine (2013) that action sequences would be under model-based control.

There are two reasons, however, not to draw strong conclusions from this result. First, it is a null result; it is possible

that in other paradigms, or other experimental settings, people would have shown evidence of model-free sequence selection. Second, it is highly likely that *some* action sequences can be under model-free control. After all, the actions “pull a lever” or “push a button” actually comprise many motor subroutines—so if MF RL can apply to them, it must apply to sequences of some kind.

Nonetheless, our results raise important questions about when and how MF RL operates at higher levels of abstraction in the action hierarchy. This question is ripe for future research.

7. CONCLUSION

Humans exhibit many habit-like patterns of behavior. Our studies demonstrate one such pattern that is best explained by model-free RL, and another that is best explained by model-based selection of action sequences. This suggests that action sequences should be viewed as complements, not alternatives, to MF RL, and that combining MF RL with other approaches will give us a fuller understanding of habits.

DATA AVAILABILITY STATEMENT

All data and code used for this paper can be found at https://github.com/adammorris/action_sequences.

ETHICS STATEMENT

This study was approved by the Committee on the Use of Human Subjects at Harvard University under protocol IRB14-2016: A computational approach to human moral judgment. We obtained written informed consent from all participants by electronic approval at the outset of our online testing procedure.

AUTHOR CONTRIBUTIONS

This research was conceived and designed by AM and FC, implemented and analyzed by AM, and written by AM with assistance from FC.

FUNDING

This research was supported by Grant N00014-14-1-0800 from the Office of Naval Research to FC.

ACKNOWLEDGMENTS

We thank Sam Gershman and Josh Buckholz for their feedback on the manuscript, and Rani Moran for identifying an issue in the model-fitting code.

REFERENCES

Crockett, M. J. (2013). Models of morality. *Trends Cogn. Sci.* 17, 363–366. doi: 10.1016/j.tics.2013.06.005

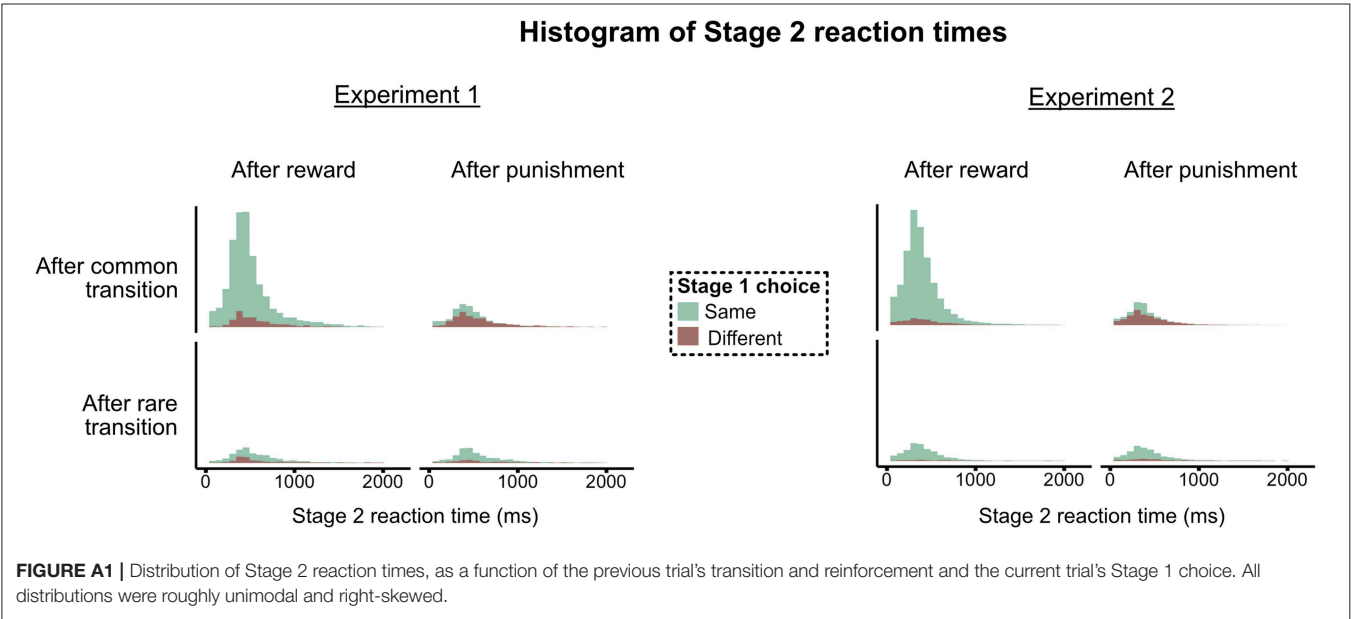
Cushman, F. (2013). Action, outcome, and value: a dual-system framework for morality. *Pers. Soc. Psychol. Rev.* 17, 273–292. doi: 10.1177/1088868313495594
Cushman, F., and Morris, A. (2015). Habitual control of goal selection in humans. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13817–13822. doi: 10.1073/pnas.1506367112

- da Silva, C. F., and Hare, T. (2019). Model-free or muddled models in the two-stage task? *bioRxiv* 682922. doi: 10.1101/682922
- Daw, N., Gershman, S., Seymour, B., Dayan, P., and Dolan, R. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 1204–1215. doi: 10.1016/j.neuron.2011.02.027
- Daw, N. D. (2011). "Trial-by-trial data analysis using computational models," in *Decision making, affect, and learning: Attention and performance XXIII*, Vol. 23, eds M. R. Delgado, E. A. Phelps, and T. W. Robbins (Oxford: Oxford University Press), 3–38.
- Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560
- Dezfouli, A., and Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* 35, 1036–1051. doi: 10.1111/j.1460-9568.2012.08050.x
- Dezfouli, A., and Balleine, B. W. (2013). Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Comput. Biol.* 9:e1003364. doi: 10.1371/journal.pcbi.1003364
- Dezfouli, A., Lingawi, N. W., and Balleine, B. W. (2014). Habits as action sequences: hierarchical action control and changes in outcome value. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130482. doi: 10.1098/rstb.2013.0482
- Dolan, R., and Dayan, P. (2013). Goals and habits in the brain. *Neuron* 80, 312–325. doi: 10.1016/j.neuron.2013.09.007
- Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., and Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nat. Neurosci.* 18, 767–772. doi: 10.1038/nn.3981
- Frank, M. J., Seeberger, L. C., and O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science* 306, 1940–1943. doi: 10.1126/science.1102941
- Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., and Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife* 5:e11305. doi: 10.7554/eLife.11305
- Gilovich, T., Vallone, R., and Tversky, A. (1985). The hot hand in basketball: on the misperception of random sequences. *Cogn. Psychol.* 17, 295–314.
- Glascher, J., Daw, N., Dayan, P., and O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66, 585–595. doi: 10.1016/j.neuron.2010.04.016
- Kool, W., Cushman, F. A., and Gershman, S. J. (2016). When does model-based control pay off? *PLoS Comput. Biol.* 12:e1005090. doi: 10.1371/journal.pcbi.1005090
- Kool, W., Gershman, S. J., and Cushman, F. A. (2017). Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychol. Sci.* 28, 1321–1333. doi: 10.1177/0956797617708288
- Miller, K. J., Shenhav, A., and Ludvig, E. A. (2019). Habits without values. *Psychol. Rev.* 126, 292–311. doi: 10.1037/rev0000120
- Morris, A., and Cushman, F. (2018). A common framework for theories of norm compliance. *Soc. Philos. Policy* 35, 101–127. doi: 10.1017/S0265052518000134
- O'Reilly, R., and Frank, M. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput.* 18, 283–328. doi: 10.1162/089976606775093909
- Otto, A. R., Gershman, S. J., Markman, A. B., and Daw, N. D. (2013). The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol. Sci.* 24, 751–761. doi: 10.1177/0956797612463080
- Palminteri, S., Wyart, V., and Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends Cogn. Sci.* 21, 425–433. doi: 10.1016/j.tics.2017.03.011
- Redish, A. D. (2004). Addiction as a computational process gone awry. *Science* 306, 1944–1947. doi: 10.1126/science.1102384
- Rigoux, L., Stephan, K. E., Friston, K. J., and Daunizeau, J. (2014). Bayesian model selection for group studies — Revisited. *Neuroimage* 84, 971–985. doi: 10.1016/j.neuroimage.2013.08.065
- Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599. doi: 10.1126/science.275.5306.1593
- Sutton, R. S., and Barto, A. G. (1998). *Introduction to Reinforcement Learning*, 1st Edn. Cambridge, MA: MIT Press.
- Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., et al. (2015). Disorders of compulsivity: a common bias towards learning habits. *Mol. Psychiatry* 20, 345–352. doi: 10.1038/mp.2014.44
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p-values. *Psychon. Bull. Rev.* 14, 779–804. doi: 10.3758/bf03194105
- Watkins, C. J. C. H., and Dayan, P. (1992). Q-learning. *Mach. Learn.* 8, 279–292. doi: 10.1007/BF00992698

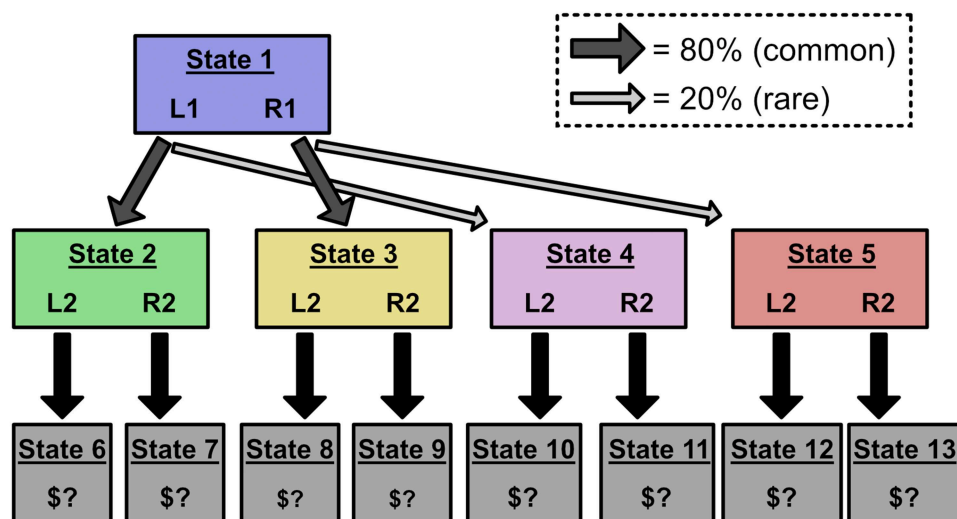
Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Morris and Cushman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A. APPENDIX



A Inaccurate task model



B Simulation results

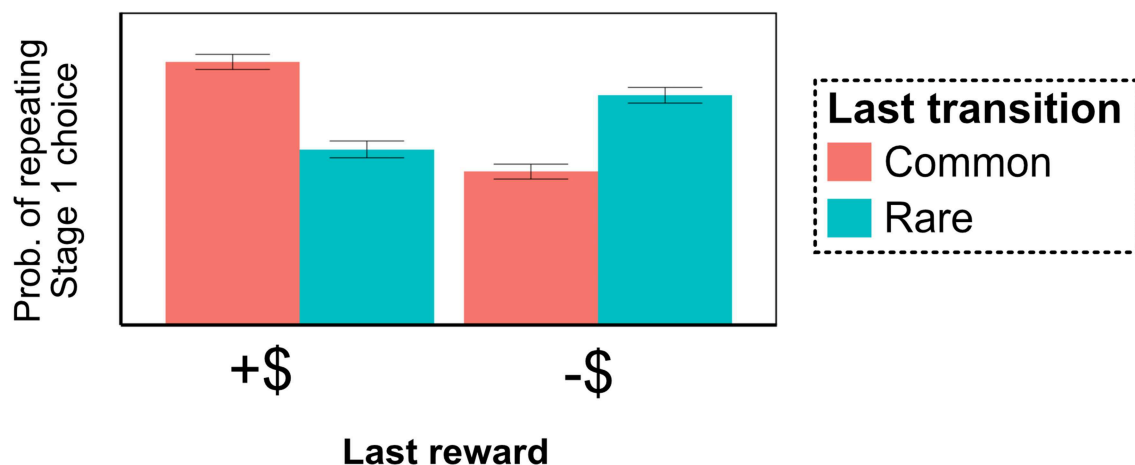


FIGURE A2 | Simulations illustrating how a model-based algorithm with an inaccurate task model could mimic model-free control. **(A)** The example inaccurate task model. The algorithm believes that rare transitions lead to unique Stage 2 states. **(B)** Simulation results for agents using a mixture of an accurate model-based algorithm and the inaccurate model-based algorithm. These agents produce both an interaction between last trial's reward and transition type (the classic signature of model-based control), and a main effect of last trial's reward (the classic signature of model-free control). Hence, these agents produce MF-like behavior without MF control. We consider possibilities like these in section 6. (We thank David Melnikoff for suggesting this example).



A Mobile Phone App for the Generation and Characterization of Motor Habits

Paula Banca^{1*}, Daniel McNamee^{2,3}, Thomas Piercy⁴, Qiang Luo^{5,6,7} and Trevor W. Robbins¹

¹Department of Psychology, Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, United Kingdom, ²Wellcome Centre for Human Neuroimaging, Institute of Neurology, University College London, London, United Kingdom, ³Max Planck UCL Centre for Computational Psychiatry, University College London, London, United Kingdom, ⁴Department of Psychiatry, Addenbrooke's Hospital, University of Cambridge, Cambridge, United Kingdom, ⁵Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China, ⁶Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Ministry of Education, Fudan University, Shanghai, China, ⁷State Key Laboratory of Medical Neurobiology, MOE Frontiers Center for Brain Science, Institute of Brain Science and Human Phenome Institute, Fudan University, Shanghai, China

OPEN ACCESS

Edited by:

Wendy Wood,
University of Southern California,
United States

Reviewed by:

Jan De Houwer,
Ghent University, Belgium
Blair T. Johnson,
University of Connecticut,
United States

*Correspondence:

Paula Banca
paula.banca@gmail.com

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 02 May 2019

Accepted: 02 December 2019

Published: 08 January 2020

Citation:

Banca P, McNamee D, Piercy T,
Luo Q and Robbins TW (2020)
A Mobile Phone App for the
Generation and Characterization
of Motor Habits.
Front. Psychol. 10:2850.
doi: 10.3389/fpsyg.2019.02850

Habits are a powerful route to efficiency; the ability to constantly shift between goal-directed and habitual strategies, as well as integrate them into behavioral output, is key to optimal performance in everyday life. When such ability is impaired, it may lead to loss of control and to compulsive behavior. Habits have successfully been induced and investigated in rats using methods such as overtraining stimulus-response associations and outcome devaluation, respectively. However, such methods have ineffectively measured habits in humans because (1) human habits usually involve more complex sequences of actions than in rats and (2) of pragmatic impediments posed by the extensive time (weeks or even months), it may take for routine habits to develop. We present here a novel behavioral paradigm—a mobile-phone app methodology—for inducing and measuring habits in humans during their everyday schedule and environment. It assumes that practice is key to achieve automaticity and proficiency and that the use of a hierarchical sequence of actions is the best strategy for capturing the cognitive mechanisms involved in habit formation (including “chunking”) and consolidation. The task is a gamified self-instructed and self-paced app on a mobile phone that enables subjects to learn and practice two sequences of finger movements, composed of chords and single presses. It involves a step-wise learning procedure in which subjects begin responding to a visual and auditory cued sequence by generating responses on the screen using four fingers. Such cues progressively disappear throughout 1 month of training, enabling the subject ultimately to master the motor skill involved. We present preliminary data for the acquisition of motor sequence learning in 29 healthy individuals, each trained over a month period. We demonstrate an asymptotic improvement in performance, as well as its automatic nature. We also report how people integrate the task into their daily routine, the development of motor precision throughout training, and the effect of intermittent reinforcement and reward extinction in habit preservation. The findings help to validate this “real world” app for measuring human habits.

Keywords: habit, skill, automaticity, motor sequence learning, extinction, sequence completion times, preparation time, routine

INTRODUCTION

The concept of habit learning has been extensively studied across distinct fields of research, using different methodologies (for a comprehensive review on habits, see Wood and R  nger, 2016). Habits are usually defined as automatic responses elicited by specific environmental stimuli (including contexts) performed autonomously of the goal (e.g., Lin et al., 2016; Robbins and Costa, 2017). Habits have been assumed to require practice or repeated training as demonstrated in experimental animals (e.g., Adams and Dickinson, 1981) and humans (Tricomi et al., 2009). However, it has proven to be surprisingly difficult to demonstrate robust habit learning in humans as a function of training (de Wit et al., 2018), possibly for reasons related to the time allowed for response preparation prior to execution (Hardwick et al., 2019), the need for much longer periods of training for humans than are possible in the laboratory and a focus on single actions rather than more complex sequences of behavior. Dezfouli and Balleine (2012) have argued that “habits are complex actions that reflect the association of a number of actions into rapidly executed action sequences.” Such action sequences have been understudied, especially given the evidence for the “chunking” together of elements of response sequences and their dependence on the striatum, a brain structure also associated with habit learning and performance (Graybiel, 1998; Sakai et al., 2003). Action sequences may also provide proprioceptive and kinesthetic sensory feedback that facilitates habit learning *via* the stimulus-response associations occurring as a consequence of the response chain and distal to the goal occurring at the end of the sequence. In this research, we aimed to develop a method for investigating habitual control of motor response sequences in the real world using a very familiar apparatus (the smartphone) over protracted training periods in human participants—and we report here a preliminary study aimed at validating a gamified application for this purpose.

Previously, self-reported questionnaires have been used to investigate aspects of habits distinguishing between routine and automatic tendencies in humans (Gardner et al., 2012; Ersche et al., 2017). These are useful but do depend on self-report rather than providing more objective measures of habits.

“Ecological” paradigms have also been used to track “real-world” habits (Lally et al., 2010; Fournier et al., 2017) assessing, among other elements, the “four horsemen of automaticity” as defined by Bargh and colleagues: awareness, intention, efficiency, and control (Bargh, 1994). Hence, we incorporated measures of automaticity of our response sequences, including speed, accuracy, and motor invariance.

Capitalizing on novel technology, we developed a smartphone motor sequence application to measure habit formation within a more naturalistic setting (at home). Habit strength is promoted here by the permanent accessibility of the app (given that most people carry their mobile phones everywhere), which facilitates training frequency and enables context stability since the tactile, visual, and auditory stimuli associated with the phone and its operation establishes a strong context for all participants regardless of their concurrent circumstances.

Thus phone-based tasks favor habit formation since as the frequency of the behavior increases in a stable context, so it increases the strength of the context-behavior association, an effect that is crucial for habit development (Verplanken and Wood, 2006). Indeed, mobile phones are notorious for their elicitation of absent-minded and unintentional use patterns which are suggested to be characteristic of automated behaviors (Bargh, 1994).

We continuously collected data online, in real time, thus enabling measures of progressive learning and of processes involved in habit formation such as “caching” (Haith and Krakauer, 2018) and “chunking” (Graybiel, 1998). Previous studies have shown that practice in itself is insufficient for habit development as it requires off-line consolidation computations, through longer periods of time (de Wit et al., 2018) and sleep (Walker et al., 2003; Nusbaum et al., 2018). This article presents the method in detail and preliminary data, acquired with 29 healthy human volunteers. Specifically, we report data on task engagement and how people integrated the task into their daily routine. We also report objective accuracy data and sequence completion times throughout a 30-day training period in order to measure task-related automaticity and motor precision.

The application incorporated attractive sensory features in a game-like setting, in which participants earned reward points according to their performance (see video for illustration in the “Methods” section). This app-based method for measuring habits in the real world is based on previous findings that have defined training frequency, context stability, and reward contingencies as important for increasing habit strength (Verplanken and Wood, 2006; Wood and R  nger, 2016). Previous work in experimental animals (Dickinson et al., 1983) has shown that the schedule of reinforcement (or reward) employed affects the speed of habit learning. Hence, we employed both continuous reinforcement (where each correct sequence received reward) and a more probabilistic schedule of rewards for correct sequences, with the hypothesis that the weaker correlation of correct sequences with reward would weaken goal-directed behavior in favor of habitual learning.

In order to assess the autonomy of habits from goal-directed actions behavioral neuroscientists employ goal devaluation or contingency degradation strategies as interventions to probe habitual control (Dickinson and Weiskrantz, 1985; Tricomi et al., 2009). Although such interventions may unmask habits only indirectly by removing goal-directed control (Gillan et al., 2015; Robbins and Costa, 2017) both rodent and human studies using them have successfully shown that well-learned action sequences can indeed become habitual and are hierarchically organized such that distinct decision-making processes may differentially control the initiation and execution of sequences (Dezfouli and Balleine, 2013; Garr and Delamater, 2019). We further report here the outcome of extinction (a form of contingency degradation; Balleine and Dickinson, 1998), by removing all the reward feedback stimuli and therefore determining how performance of the sequence was maintained.

MATERIALS AND METHODS

Participants

Twenty-nine volunteers, recruited from the community *via* advertisements (flyers), participated in the present study (11 males/18 females, mean age: 39.14 ± 11.79 years). They were all in good health, unmedicated, had no history of neurological or psychiatric conditions, and were also free from any substance dependence. Two participants who scored above 4 on the Beck Depression Scale (Beck et al., 1961) and higher than 6 on the Montgomery-Åsberg Depression Rating Scale (Montgomery and Åsberg, 1979) were excluded. Only one of our recruited participants used to play video games. All participants were given a letter of information, gave written informed consent prior to participation, in accordance with the Declaration of Helsinki, and were financially compensated for their participation (£20 in total: £5 incentive each week for keeping their motivation). They were told that this research aims at investigating how habits are formed, and therefore, we would need them to repeat the task for a longer period (1 month) than in usual studies. This study was approved by the East of England–Cambridge South Research Ethics Committee (16/EE/0465).

Habit Training Task Design

The task consisted of a motor practice program that participants committed to pursue daily, for a period of 1 month (see description of the task design in **Figure 1** and in the following video: <https://youtu.be/XSYrBzD7ZpI>).

Using a simple and self-instructed application downloaded to their mobile devices, participants learned and practiced two sequences of fingers movements, composed of chords (two or three simultaneous finger presses) and single presses (one finger only). Each sequence comprised six moves, performed using four fingers of the dominant hand (index, middle, ring, and little finger). Sequence generation was randomized so that each participant had their own pair of sequences to practice throughout the month. This randomization was conducted to rule out finger-specific effects at individual sequential positions, as each finger will contribute equally to the RTs at each sequential position. However, for each sequence, the order of finger movements was pseudo-randomly generated such that (1) all sequences had three single press moves, two two-finger chord moves, and one three-finger chord move and (2) difficult finger combinations were avoided, for example, a three-finger chord with simultaneous index, middle, and little fingers or index, ring, and little fingers. Therefore, despite being different, all sequences had a similar level of difficulty.

Participants were instructed to respond swiftly and accurately. They were required to keep their fingers very close to the keys to minimize amplitude variation and to enable them to play quickly. To enable sequence learning and memorization, three levels of increased difficulty guided practice. Initially, subjects responded to a visually and auditory cued sequence: they simply followed lighted keys, also associated with musical notes (level 1). These exteroceptive cues were slowly removed throughout the practice progression such that level 2 only included auditory cues and level 3 contained no cues. Successful

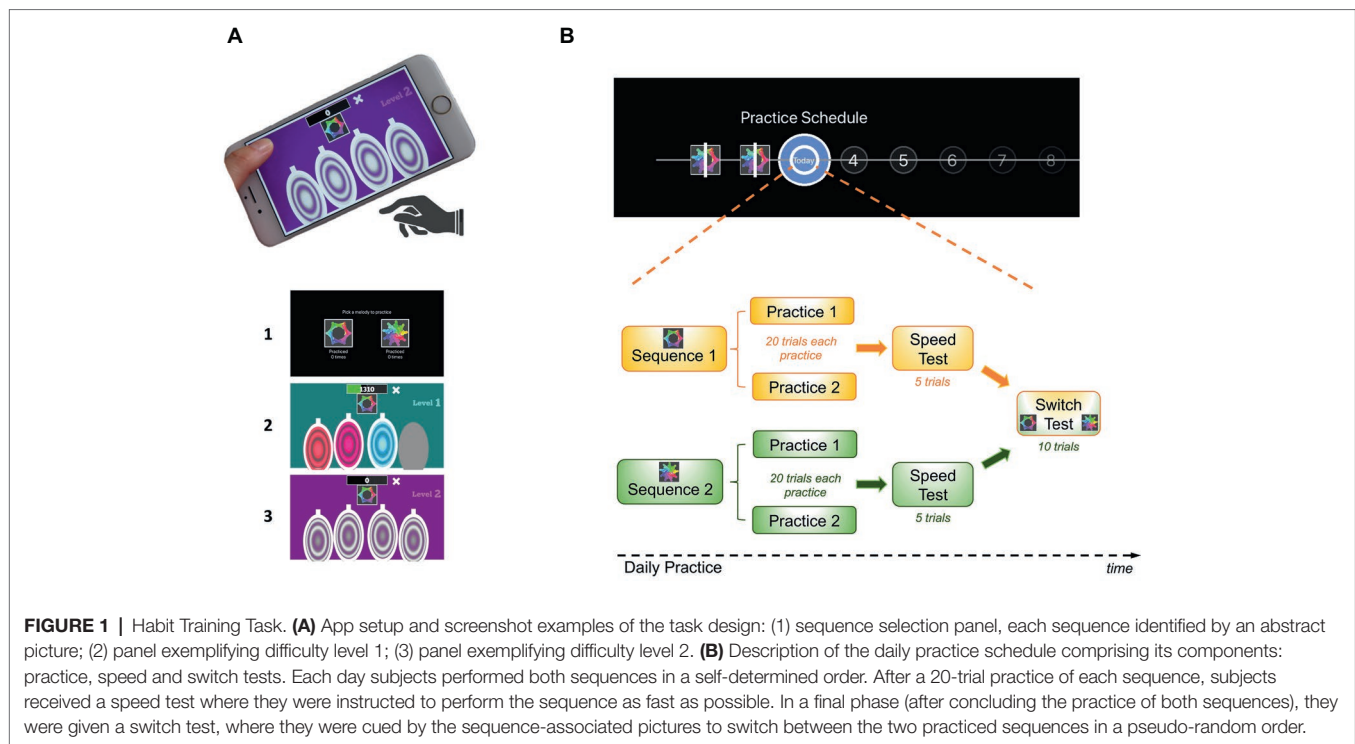
performance at each stage resulted in progression to the next. Unsuccessful performance resulted in titration to the immediately preceding stage.

Participants received continuous feedback on their performance. Successful trials were followed by a positive ring tone and mistakes by a negative ring tone. Every time a mistake occurred (irrespective of which move in the sequence they were), participants had to restart the sequence in order to perform it entirely correctly.

As previously mentioned, all participants had to practice two motor sequences, each identified by a specific abstract picture. Each sequence was associated with a specific reward schedule. In our design, one of the reward schedules was continuous reward (points were received for every successful trial, as a function of the speed of performance) and the other a variable reward schedule (points were randomly received on 37% of the trials). Calculation of the points was as follows: points decreased linearly from 100 to 0 over 1 second; the counting started as soon as the app became ready to receive the user's input. This counter reset and restarted counting after each move. As soon as the keys were pressed (for each move), the counter stopped and registered the points achieved for that move. The points received after each sequence was completed were the sum of all the points achieved on each move. All this within-move counting was done in the background so participants only saw the points gained for each sequence once they completed it. This system was implemented to promote speed: the faster participants played the sequence, the more points they gained. If they were too slow, that is, if the key press occurred after the counter had reached 0, then no points were gathered for that particular move. In the continuous reward sequence, participants received the total points acquired after each successful trial completed. In the variable reward sequence, there was a 63% chance that any points earned on a sequence would be set to zero. To compensate for the missing points, the earned points provided on this schedule were doubled. Therefore, both sequences resulted in similar scores by the end of practice. By this point, after 20 sequences had been completed (see “Practice Schedule” section below), subjects could see the total (cumulative) points achieved throughout the practice. While playing, they could also see their current total, gathered at a particular moment. To promote motivation, feedback was also given across daily practice sessions, so subjects could compare their performance across practices and see whether they were improving over days.

Practice Schedule

All participants were presented with a calendar schedule and were asked to practice both sequences daily (**Figure 1B**). They were instructed to practice as many times as they wish, whenever they wanted during the day and with the sequence order they would prefer. However, a minimum of two practice sessions per sequence was required every day; each practice comprised 20 sequences. The instruction was the following: “*You can practice as many times as you wish, whenever you want during your day and with the sequence order you want. Your minimum training required per day is 2 rounds of practice for each sequence*”



but since every person has different learning rates, you are responsible for assessing how much you need to practice in order to make sure you come back for a second session, in a month time, mastering the sequences. You need to know them by heart, automatically and quickly!”. Once the minimum practice sessions were completed, a short retention speed test of five trials followed, to assess that day’s performance. During this short session, participants were instructed to repeatedly tap a sequence as rapidly as possible while making as few errors as possible. After this, participants were asked to rate, on a percentage scale, the following two questions: (1) *How much did you enjoy playing this sequence?* and (2) *How confident are you that you know this sequence by heart?* Finally, participants were required to engage in a 10 trial-switch test, in which they would practice switching between the two sequences in a pseudo-random order. The sequence to be played was cued by the respective associated picture. Speed and switch tests never received reward feedback (only the practice sessions). This sequence of events (practice, speed, ratings and switch sessions) happened every day (**Figure 1B**). If subjects would miss a day of practice, they would need to catch up on the training the day after, that is, they would be required to do the minimum training for the current and previous day. To remove pauses in the training, a “dead man” switch procedure was implemented in the app.

Thirty days of practice were required, and all data were anonymously collected in real time, through an online server. At the 21st day of practice, the reward schedules were removed (extinction) to test how autonomous of external feedback the response sequence had become. This procedure (1) ensured that the response sequence was more dependent

on interoceptive (proprioceptive and kinesthetic) feedback and on the subjects’ internal motivation to continue the training and (2) ensured that we were able to measure and train response sequences triggered by their context, which persisted without explicit reinforcement.

An orientation session, lasting between 30 and 60 min depending on people’s dexterity, was conducted at the Herschel Smith Building, Addenbrooke’s Hospital, in Cambridge. During this session, the researcher helped the participant to download the app to their devices, reviewed the training instructions, and discussed how the task works. All participants were instructed to practice every day to make sure they could perform both sequences automatically and rapidly as they would be assessed in a second session taking place 1 month later. This cover story was introduced in preparation for a follow-up session including a devaluation strategy, which assessed participants’ preferences for habitual sequences over goal-seeking sequences. This task manipulation would test the hypothesis that the behavioral mechanism underlying the transition from a goal-directed to a habitual action is that the action, with repetition, acquires the rewarding properties of its outcome, which may simply be its own proprioceptive/kinesthetic feedback (data to be reported elsewhere).

Data Analyses

Behavioral output measures included sequence accuracy and sequence completion times (learning rates), temporal pattern of daily practice, days until habit acquisition, performance as function of different reward schedules, effect of reward extinction and finger position and timings.

For more detailed analyses, we broke down the sequence completion times into two components: (1) *move preparation time*: the time period between the last release of the previous move and the first press of the current move and (2) *move performance time*: the time period between the first press and the last release of each move, representing the duration of each move from the time participants press until they release the keys (i.e., muscle time).

App data were automatically uploaded to a Cloud-based database. Data analysis was performed using custom scripts in MATLAB and Python.

RESULTS

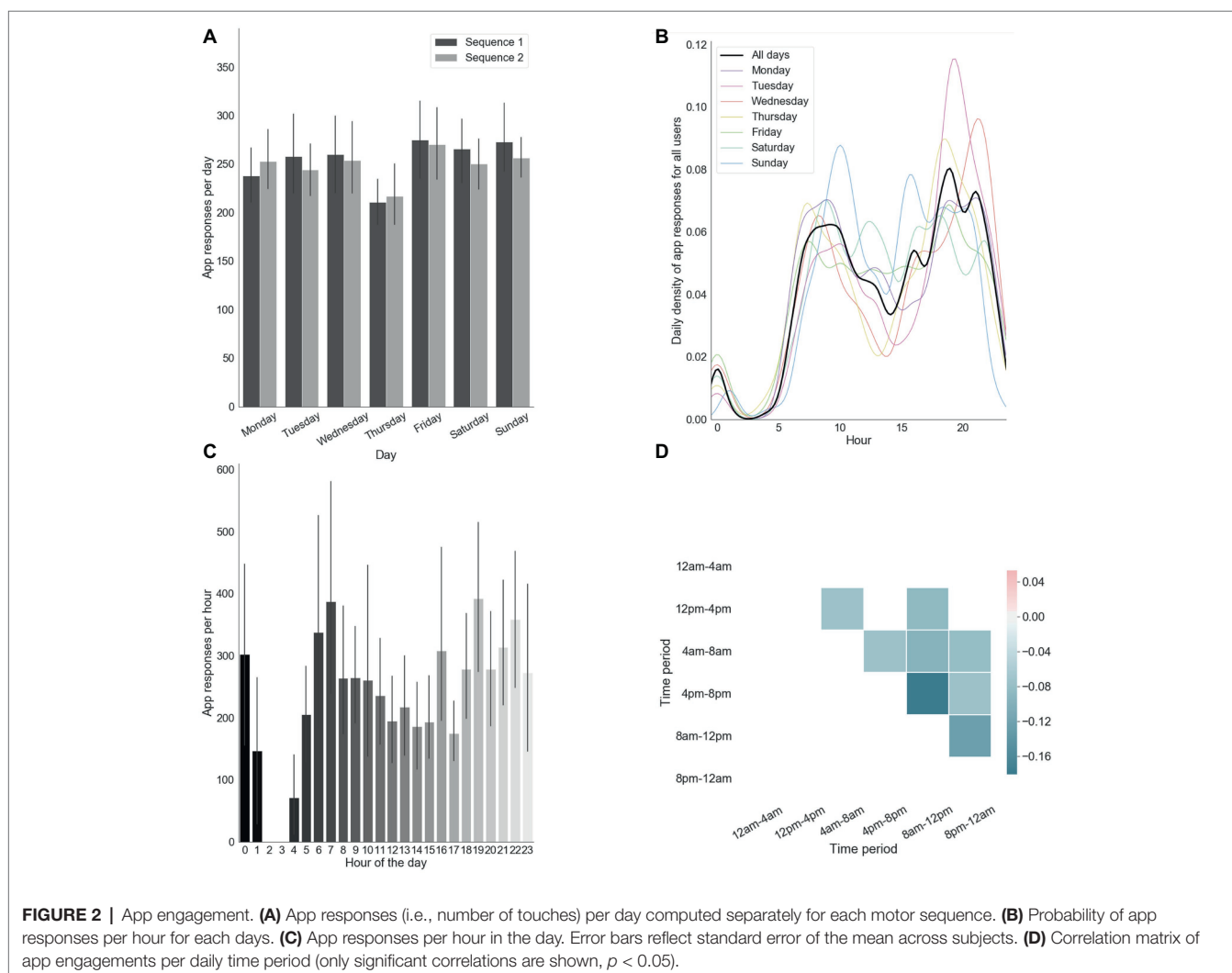
Validation of Training and Individual Routines

As shown in **Figures 2A–D**, our participants reliably committed to their regular training schedule. They generally fulfilled the requirement of practicing consistently both sequences every day

(**Figure 2A**). The approximately bimodal distribution observed in **Figure 2C** depicts our participant's tendency to practice mostly during early mornings (~7:00) and evenings (~19:00). This tendency was relatively consistent across days (**Figure 1B**). Moreover, on a daily basis, participants typically chose to practice at one time point of their day as shown by the anti-correlations in app engagement across different daily time periods (**Figure 2D**). In particular, those who chose to practice in the evening tended not do it in the morning and vice versa, as indicated by the strongest anti-correlation between 8 and 12 am and 4 and 8 pm.

Effects of Extinction

We analyzed the two blocks of practice pre- and post-removal of the external rewarding feedback occurring after 21 days of training. After *extinction*, during the practice session, there was a significant decrease in performance in terms of both increased errors ($p < 0.0001$) and longer sequence completion times ($p < 0.05$) (**Figures 3A,E**). This effect occurred irrespective of the reward feedback schedule (continuous versus variable, **Figure 3B**). Nevertheless, analyses



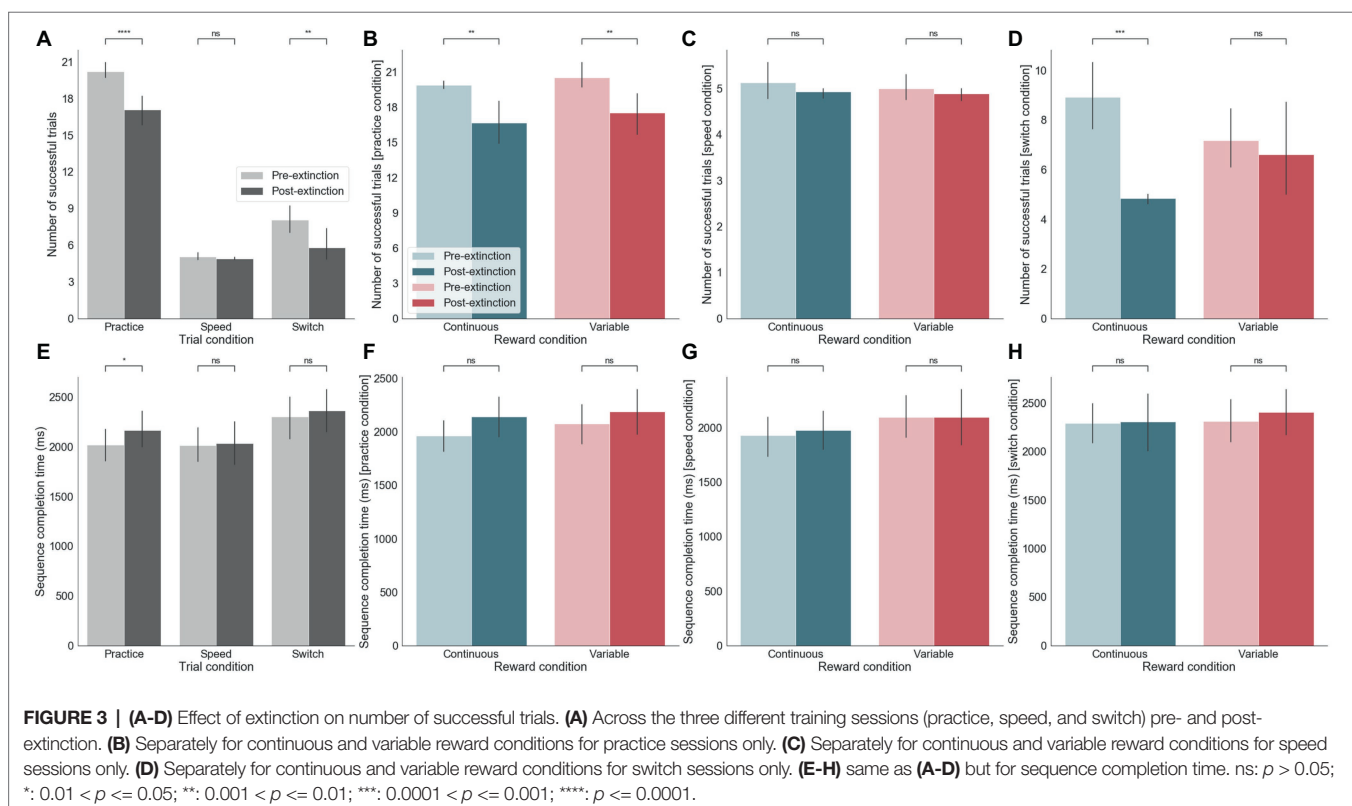
of subsequent effects of *extinction* on the switch and speed tests (although these had never previously received reward feedback) showed that there was a significant performance decrement post-*extinction* during the switch test, only following continuous reward feedback training ($p < 0.001$) (**Figure 3D**). There was however no effect on sequence completion time (**Figure 3H**). There was no effect on post-*extinction* performance during the speed test (**Figures 3C,G**). In summary, participants made significantly more errors after *extinction* in both sequences, irrespective of whether successful sequences were previously rewarding in a continuous or variable manner. During the switch test, this accuracy effect was only strongly observed for the continuous reward. Generally, accuracy seemed to be strongly affected by reward *extinction* (**Figure 3**, top row - number of successful trials) but sequence completion times were less sensitive to this manipulation (**Figure 3**, bottom row - sequence completion times).

Sequence Performance

Significant improvements in accuracy (**Figure 4A**) and normalized group-averaged decreases in sequence completion times (**Figure 4C**) throughout training indicate that learning occurred as expected. Participants started their training with a mean sequence completion of 3,719 ms in successful trials based on the first five blocks of practice (referred to as “early training”) and completed their training with a mean sequence completion time of 2,346 ms calculated using the last five blocks of practice (late training). A paired *t*-test between the mean sequence completion time per subject in the early versus late training

periods was significant at $p < 10^{-20}$ (**Figure 4D**). Accuracy also improved significantly ($p < 10^{-7}$) from early (mean success rate = 0.46) versus late (mean success rate = 0.75) training, with steep improvements occurring at the beginning of training and remaining stable to the end of app engagement (**Figure 4B**). There were no significant differences for either errors or sequence completion times as a function of the reward feedback schedule (i.e., continuous versus variable). For errors, performance appeared to reach an asymptote between blocks 15 and 20. In contrast, for sequence completion time, performance continued to improve throughout training suggesting that these behavioral measures are differentially sensitive to distinct learning processes. Throughout training, the sequence completion time of the first trial within each block appeared to be longer than subsequent trials within the same block (**Figure 4D**, dashed lines).

When decomposing the sequence completion time on successful trials into preparation (i.e., quantifying the time just before a move) and motor-related components (**Figure 5A**), there was an order effect by which the move number inversely correlated with the preparation time, consistent with a competitive queuing model of action sequence preparation (Averbeck et al., 2002; Rhodes et al., 2004). That is, as the sequence is performed successfully, fewer moves compete for motor output, thus resulting in shorter preparation times. There was a significantly larger preparation time for the first move, as compared with all the remaining moves of the sequence (**Figure 5A**). This time period before the first move also includes the time devoted to the sensory processing of the input stimuli from the app. The linear decrease in



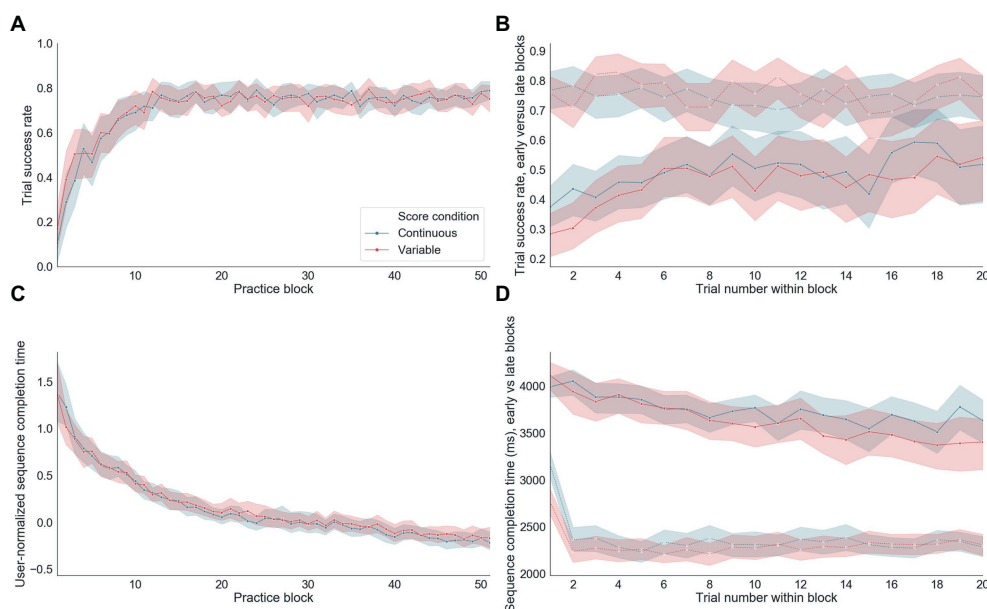


FIGURE 4 | Performance. **(A)** Trial success rate (i.e., rate of correct sequences) over practice training. **(B)** Trial success rate in early versus late blocks of practice. Dashed line: late blocks, Continuous line: early blocks. **(C)** Sequence completion time, normalized within each subject, on successful trials over practice. **(D)** Sequence completion time in successful trials in early versus late blocks of practice. Dashed line: late blocks, Continuous line: early blocks.

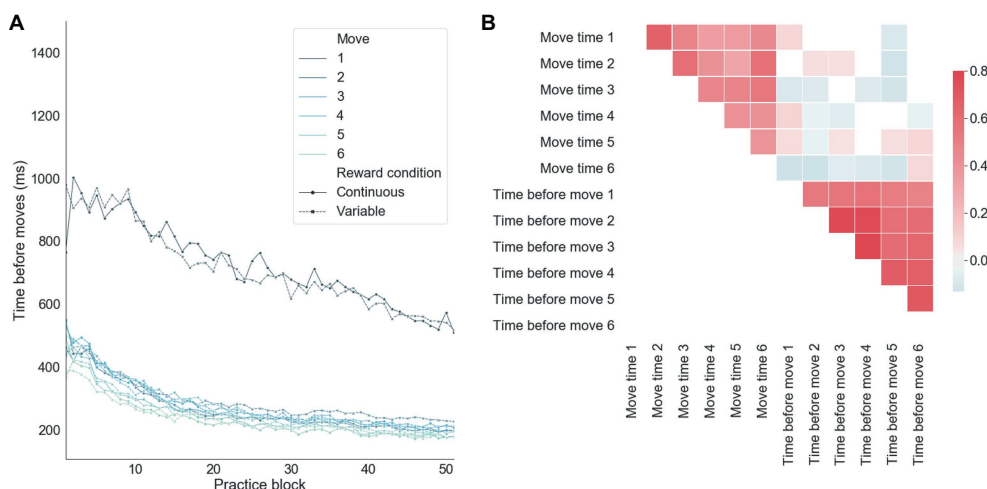
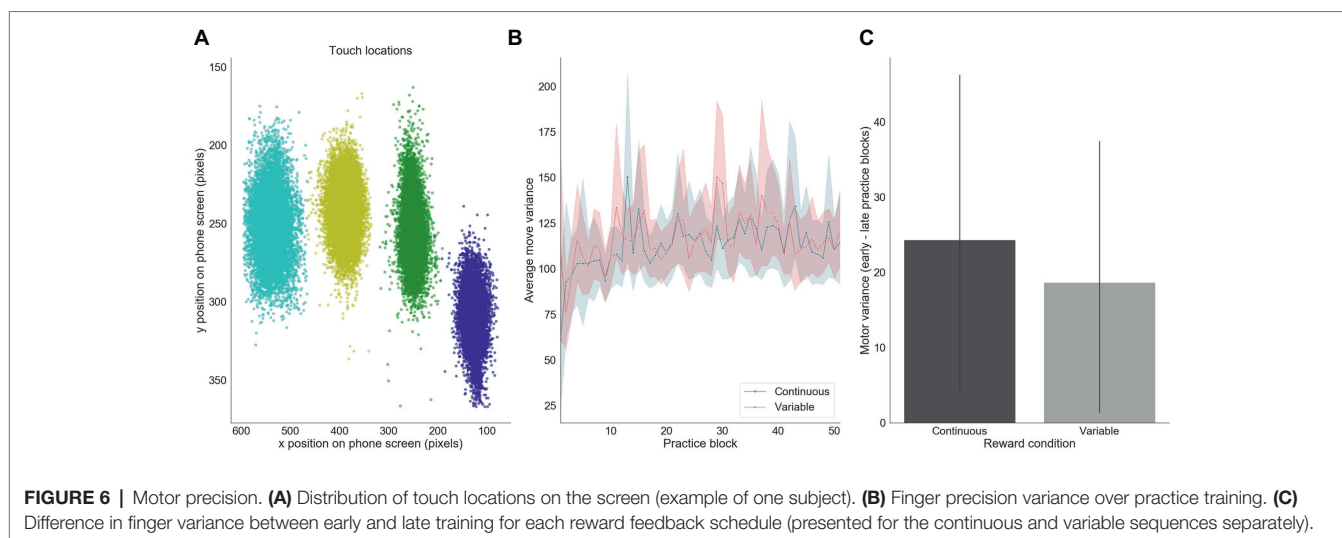


FIGURE 5 | Decomposing sequence completion times into preparation and motor related components. **(A)** Preparation time (time before moves) as a function of practice for each move in each sequence averaged across subjects. **(B)** Correlations between motor (move time) and preparation times.

sensorimotor processing before the first move over training was in contrast to the exponential decay toward baseline observed in the remaining moves. This suggests that qualitatively different learning processes are engaged by the brain in order to optimize sensory-to-motor and motor-to-motor mappings. In correlation analyses (**Figure 5B**), it was found that move preparation times and move motor times were (separately) strongly correlated, whereas preparation times and motor times were weakly anti-correlated. This suggests that, over learning, preparation times and motor times were improved in a consistent

manner across moves and that the brain may trade-off preparation and motor times in order to achieve an efficient balance between speed and accuracy. In particular, the anti-correlation between preparation and motor times on successful trials emerged due to trials with both fast preparation and fast motor times leading to errors. In summary, despite theoretical and empirical dependencies between these two components of the RT, there was some degree of independence between them as reflected in the relatively lower correlation cross-component correlation values.



Motor Precision

We also assessed how motor precision, as measured by finger position variance, varied throughout training. This measure was computed using the X and Y pixel coordinates of participants' screen touches (**Figure 6**). There was a decrease in average motor precision throughout training, mainly during the first 10 blocks of practice (**Figure 6B**). This decrease was slightly, but not significantly, more pronounced in the continuous reward condition ($p = 0.040$) than in the variable reward one ($p = 0.061$) (**Figure 6C**).

DISCUSSION

We have presented an experimental paradigm based on motor sequence learning which can be employed to study, in a systematic and controlled way, the building blocks of more complex behavioral sequences that make up our everyday real-world actions. Designed as a smartphone tool, and thus easily available to subjects, it enabled for the first time, the induction and measurement of habitual behavior in humans during their everyday schedule, routines, and environment (in the comfort of their homes), while collecting continuously 30 days of real-time data. Such a naturalistic experimental set-up may perhaps be useful for the future investigation of habits.

The test paradigm is assumed to encompass multiple and continuous cycles of model-free and model-based learning processes thought to be required for habit development, which include processes of instrumental or operant reinforcement, adaptation, plasticity, and other explicit cognitive processes (Krakauer and Mazzoni, 2011; Haith and Krakauer, 2013). It also assumes that practice is key to achieve automaticity and proficiency and that the use of a hierarchical sequence of actions is the best strategy for capturing the cognitive mechanisms involved in habit formation and consolidation.

This app-based method to measure habits in the real world is based on previous literature which has isolated frequency, context stability, rewards, and simplicity as important factors

that promote habit strength (Verplanken and Wood, 2006; Wood and R nger, 2016). Participants perform the task on a frequent basis in a similar context (i.e., the phone and app), supported by game-related rewards. Our purpose here is to present the method in detail and validate it based on data in healthy volunteers. Our preliminary analyses attest to its successful design and good tolerability. All subjects completed the training. After 1 month of training, their speed and accuracy greatly improved. They were also capable of learning the task and performed it with a pronounced degree of automaticity. Participants reported that the task became simpler and easier to perform throughout the training, corroborating the assumption that perceived complexity of a behavior is also an element that influences the extent to which automaticity is attained (McCloskey and Johnson, 2019). In agreement with recent questionnaire methods for parsing components of habits (e.g., Ersche et al., 2017), we observed both routine (evidenced by the anti-correlation in app engagement across different daily time periods) and automaticity (evidenced by a combination of an asymptotic performance and responsiveness in the absence of cues).

Automaticity was measured in terms of three criteria: sequence completion times, progressive extinction of learning cues, and autonomy from the goal as assessed by extinction. Additionally, proficiency was also measured in terms of motor precision. The significant increase in finger variance throughout the training is also a strong indicator of motor performance optimization. According to optimal feedback control theory (Todorov and Jordan, 2002), optimal performance is achieved by allowing variability in redundant (task-irrelevant) dimensions. While still learning, participants tend to be more precise, "freezing" the degrees of freedom of their movements and having a fine-tuned and highly accurate sequence of movements (Bernstein, 1967; Vereijken et al., 1992). With training, as the skill develops into a fluid level of proficiency, motor variance increases because subjects learn that this will not impact successful sequence completion and contributes to an improved speed-accuracy trade-off (Todorov and Jordan, 2002).

In terms of proficiency and automaticity, sequence completion times significantly improved throughout training, reaching asymptotic performance levels between practice blocks 40 and 50. The exponential decay in error rates to an asymptote and further optimization of the speed/accuracy trade-off is clear evidence of learning and skill development. The greater improvement in sequence completion time during the initial 20 blocks corresponds to the “fast learning” mode, typically observed during the goal-based acquisition phase mediated by the associative striatal regions, in coordination with cerebellum, prefrontal, and premotor cortical regions (Hikosaka et al., 2002; Hardwick et al., 2013). The progressive stabilization of the sequence completion times during the remaining blocks of training likely resembles a shift to an autonomous stage of habit development (Hikosaka et al., 1999), hypothetically linked to a devolution of control to sensorimotor striatal regions (Hikosaka et al., 1999; Lehericy et al., 2005), and progressive disengagement of cognitive control hubs in the frontal and cingulate cortices (Bassett et al., 2015). The asymptotic performance attained with our task indicates that proficiency was attained as one criterion of response sequence development. Of special note also is the significantly longer sequence completion time of the first trial compared with subsequent trials within each block that occurred only during the later stages of training when asymptotic performance was observed. This may reflect the initial retrieval of the memory of the motor program into working memory and its subsequent priming on succeeding trials. This cognitive mechanism may be an initial step underlying the “chunking” process, by which elements of the motor sequence are most efficiently ordered into a motor program, well known in motor learning research (Graybiel, 1998; Sakai et al., 2003).

The preservation of this skilled behavior after extinction of the external cues, maintaining the same high level speed-accuracy trade-off, is an additional sign of automaticity and habitual control. Our findings are consistent with Hardwick et al. (2019), who also demonstrated that practice influences habits by modulating the likelihood of habit expression *via* reducing the average time of movement initiation (Hardwick et al., 2019). We also found that in later stages of the training, our participants’ response preparation times were extremely brief and unlikely to enable expression of goal-directed responses.

One test of habitual control effected in this task was extinction, involving the omission of explicit reward feedback. The removal of rewarding feedback on the 21st day of training mainly affected errors. Although there was a small effect on sequence completion time (only in the practice condition), this was much less significant, possibly indicating that performance had indeed attained a degree of autonomy from the goal. This suggests that the motor sequence had become habitual in part but still retained some sensitivity to goal despite extinction (and hence goal-directed control). Of course, this extinction manipulation did not remove all forms of motivation from performance because of the degree of intrinsic motivation that humans exhibit in such research studies.

Although one could expect different learning patterns as consequence of different reward schedules, we did not observe significant effects of the reward feedback schedule (i.e., continuous versus variable) on habit development. There was, however, a

selective effect of reward schedule in performance during the switch test. The detrimental effect of extinction on this switch test depended on the previous schedule of reward feedback, specifically occurring in the continuous condition only. A possible explanation for this might be that pitting two habits against one another in an explicit choice situation recruits executive processing, hence re-engaging the goal-directed system, which may be more vulnerable to extinction in the continuously rewarded condition because the change in reward contingency is more immediate and explicit than for the variable schedule. Future studies may seek to vary the nature of intermittency of the reward schedule by explicitly comparing random ratio versus random interval schedules, the latter being associated with greater habitual control (Dickinson et al., 1983), although making such a comparison is challenging for response sequences as distinct from single actions.

This study has a few limitations and challenges to consider. Its ecological nature, enabling people to conduct the task in the comfort of their homes, including it in their everyday schedule, routines, and environment and at their own pace, partly solves the major problem of the artificial nature of previous studies. However, this feature limits the study to its behavioral nature, making it more difficult if one wants to investigate the neural basis of habit and skill development using functional imaging. There were also some technical difficulties to deliver the app on android phones, confining our recruitment to Apple users, which obviously decreased our recruitment pool of subjects. Several iPods were purchased for lending to participants in order to facilitate recruitment. Additionally, the study required careful monitoring by the researchers on a daily basis, to track participants’ commitment, gauge motivation, and send reminders when needed. Conducting this study with clinical populations might be challenging, given that some patients may not be so motivated as healthy volunteers. However, this concern has not in fact been the case with patients with OCD we have also begun to recruit, following the same procedures. Despite all the technical challenges, which also included a continuous update of the online server for data collection, such an advanced methodology was worth pursuing since it facilitated the acquisition of a large dataset, without requiring much effort from our participants.

In conclusion, this article aimed to validate a novel behavioral method for measuring motor response sequence habits in the real world using a mobile phone app. The analysis provided here is preliminary but sufficient to show that proficiency and automaticity is attained according to several different criteria. When tested in clinical populations, the method may provide new insights into the mechanisms underlying abnormal habit learning and corticostriatal functioning in psychiatric disorders and their putative contributions to compulsive behavior. Ongoing research is using this novel app to investigate the neural mechanisms of compulsive behavior in patients with OCD. More generally, this app-based approach could be deployed in a wide variety of discrete sequential production paradigms including dexterity, music, and memory training. It could also be used in the studies of individual differences, for example, to investigate whether aspects of the Big 5 predict how quickly habit/skill is developed.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

All participants were given a letter of information, gave written informed consent prior to participation, in accordance with the Declaration of Helsinki, and were financially compensated for their participation. This study was approved by the Cambridge South Research Ethics Committee.

AUTHOR CONTRIBUTIONS

PB and TR conceived the idea and designed the method. PB and TP developed the app. PB carried out the experiment. PB and DM conducted the data analyses. QL helped with the data preparation. PB, DM, and TR wrote the manuscript.

REFERENCES

- Adams, C. D., and Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol.* 33, 109–121. Available at: https://scholar.google.com/scholar_lookup?hl=en&publication_year=1981&author=C.+D.+Adams&author=A.+Dickinson&author=N.+E.+Spear&author=R.+R.+Miller&title=Actions+and+habits%3A+Variations+in+associative+representations+during+instrumental+learning (Accessed November 1, 2018).
- Averbeck, B. B., Chafee, M. V., Crowe, D. A., and Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *PNAS* 99, 13172–13177. doi: 10.1073/pnas.162485599
- Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419. doi: 10.1016/S0028-3908(98)00033-1
- Bargh, J. (1994). “The four horsemen of automaticity: awareness, efficiency, intentions and control” in *Handbook of social cognition*. eds. R. Wyer and T. Srull (New York: Psychology Press), 1040.
- Bassett, D. S., Yang, M., Wymbs, N. F., and Grafton, S. T. (2015). Learning-induced autonomy of sensorimotor systems. *Nat. Neurosci.* 18, 744–751. doi: 10.1038/nn.3993
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., and Erbaugh, J. (1961). An inventory for measuring depression. *Arch. Gen. Psychiatry* 4, 561–571. doi: 10.1001/archpsyc.1961.01710120031004
- Bernstein, N. I. (1967). *The coordination and regulation of movements*. Oxford: Pergamon.
- de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., et al. (2018). Shifting the balance between goals and habits: five failures in experimental habit induction. *J. Exp. Psychol. Gen.* 147, 1043–1065. doi: 10.1037/xge0000402
- Dezfouli, A., and Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* 35, 1036–1051. doi: 10.1111/j.1460-9568.2012.08050.x
- Dezfouli, A., and Balleine, B. W. (2013). Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Comput. Biol.* 9:e1003364. doi: 10.1371/journal.pcbi.1003364
- Dickinson, A., Nicholas, D. J., and Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Q. J. Exp. Psychol. B* 35, 35–51. doi: 10.1080/14640748308400912
- Dickinson, A., and Weiskrantz, L. (1985). Actions and habits: the development of behavioural autonomy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 308, 67–78. doi: 10.1098/rstb.1985.0010

FUNDING

This work was supported by the Sir Henry Wellcome Trust Postdoctoral Research Fellowship (204727/Z/16/Z) to PB, the Sir Henry Wellcome Trust Postdoctoral Research Fellowship (110257/Z/15/Z) to DM, the NIHR Cambridge Biomedical Research Centre (Mental Health Theme) to TP, the National Natural Science Foundation of China (grant 81873909 and 81930095), Natural Science Foundation of Shanghai (grant 17ZR1444400), National Key Research and Development Program of China (2018YFC0910503), Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01), and ZJLab to QL, and the Wellcome Trust Senior Investigator Award (104631/Z/14/Z) to TR.

ACKNOWLEDGMENTS

We would also like to thank the intern students Samantha De La-Rocque, Emilia Szmyrgala, and Rebecca Richards for their help piloting the task. PB is a postdoctoral research fellow at Hughes Hall College, Cambridge and during the revision of this manuscript QL was a Visiting Fellow at Clare Hall, Cambridge, UK.

- Ersche, K. D., Lim, T.-V., Ward, L. H. E., Robbins, T. W., and Stoeh, J. (2017). Creature of habit: a self-report measure of habitual routines and automatic tendencies in everyday life. *Personal. Individ. Differ.* 116, 73–85. doi: 10.1016/j.paid.2017.04.024
- Fournier, M., d'Arripe-Longueville, F., and Radel, R. (2017). Testing the effect of text messaging cues to promote physical activity habits: a worksite-based exploratory intervention. *Scand. J. Med. Sci. Sports* 27, 1157–1165. doi: 10.1111/sms.12730
- Gardner, B., Abraham, C., Lally, P., and de Bruijn, G.-J. (2012). Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the self-report habit index. *Int. J. Behav. Nutr. Phys. Act.* 9:102. doi: 10.1186/1479-5868-9-102
- Garr, E., and Delamater, A. R. (2019). Exploring the relationship between actions, habits, and automaticity in an action sequence task. *Learn. Mem.* 26, 128–132. doi: 10.1101/lm.048645.118
- Gillan, C. M., Otto, A. R., Phelps, E. A., and Daw, N. D. (2015). Model-based learning protects against forming habits. *Cogn. Affect. Behav. Neurosci.* 15, 523–536. doi: 10.3758/s13415-015-0347-6
- Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* 70, 119–136. doi: 10.1006/nlme.1998.3843
- Haith, A. M., and Krakauer, J. W. (2013). Model-based and model-free mechanisms of human motor learning. *Adv. Exp. Med. Biol.* 782, 1–21. doi: 10.1007/978-1-4614-5465-6_1
- Haith, A. M., and Krakauer, J. W. (2018). The multiple effects of practice: skill, habit and reduced cognitive load. *Curr. Opin. Behav. Sci.* 20, 196–201. doi: 10.1016/j.cobeha.2018.01.015
- Hardwick, R. M., Forreance, A. D., Krakauer, J. W., and Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. *Nat. Hum. Behav.* 1–11. doi: 10.1038/s41562-019-0725-0
- Hardwick, R. M., Rottsch, C., Miall, R. C., and Eickhoff, S. B. (2013). A quantitative meta-analysis and review of motor learning in the human brain. *NeuroImage* 67, 283–297. doi: 10.1016/j.neuroimage.2012.11.020
- Hikosaka, O., Nakahara, H., Rand, M. K., Sakai, K., Lu, X., Nakamura, K., et al. (1999). Parallel neural networks for learning sequential procedures. *Trends Neurosci.* 22, 464–471. doi: 10.1016/S0166-2236(99)01439-3
- Hikosaka, O., Nakamura, K., Sakai, K., and Nakahara, H. (2002). Central mechanisms of motor skill learning. *Curr. Opin. Neurobiol.* 12, 217–222. doi: 10.1016/S0959-4388(02)00307-0
- Krakauer, J. W., and Mazzoni, P. (2011). Human sensorimotor learning: adaptation, skill, and beyond. *Curr. Opin. Neurobiol.* 21, 636–644. doi: 10.1016/j.conb.2011.06.012

- Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., and Wardle, J. (2010). How are habits formed: modelling habit formation in the real world. *Eur. J. Soc. Psychol.* 40, 998–1009. doi: 10.1002/ejsp.674
- Lehericy, S., Benali, H., Van de Moortele, P.-F., Pelegrini-Issac, M., Waechter, T., Ugurbil, K., et al. (2005). Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *Proc. Natl. Acad. Sci.* 102, 12566–12571. doi: 10.1073/pnas.0502762102
- Lin, P.-Y., Wood, W., and Monterosso, J. (2016). Healthy eating habits protect against temptations. *Appetite* 103, 432–440. doi: 10.1016/j.appet.2015.11.011
- McCloskey, K., and Johnson, B. T. (2019). Habits, quick and easy: perceived complexity moderates the associations of contextual stability and rewards with behavioral automaticity. *Front. Psychol.* 10. doi: 10.3389/fpsyg.2019.01556
- Montgomery, S. A., and Asberg, M. (1979). A new depression scale designed to be sensitive to change. *Br. J. Psychiatry* 134, 382–389. doi: 10.1192/bjp.134.4.382
- Nusbaum, H. C., Uddin, S., Van Hedger, S. C., and Heald, S. L. (2018). Consolidating skill learning through sleep. *Curr. Opin. Behav. Sci.* 20, 174–182. doi: 10.1016/j.cobeha.2018.01.013
- Rhodes, B. J., Bullock, D., Verwey, W. B., Averbeck, B. B., and Page, M. P. A. (2004). Learning and production of movement sequences: behavioral, neurophysiological, and modeling perspectives. *Hum. Mov. Sci.* 23, 699–746. doi: 10.1016/j.humov.2004.10.008
- Robbins, T. W., and Costa, R. M. (2017). Habits. *Curr. Biol.* 27, R1200–R1206. doi: 10.1016/j.cub.2017.09.060
- Sakai, K., Kitaguchi, K., and Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. *Exp. Brain Res.* 152, 229–242. doi: 10.1007/s00221-003-1548-8
- Todorov, E., and Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.* 5, 1226–1235. doi: 10.1038/nn963
- Tricomi, E., Balleine, B. W., and O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29, 2225–2232. doi: 10.1111/j.1460-9568.2009.06796.x
- Vereijken, B., van Emmerik, R. E. A., Whiting, H. T. A., and Newell, K. M. (1992). Free(z)ing degrees of freedom in skill acquisition. *J. Mot. Behav.* 24, 133–142. doi: 10.1080/00222895.1992.9941608
- Verplanken, B., and Wood, W. M. (2006). Interventions to break and create consumer habits. *J. Public Policy Mark.* 25, 90–103. doi: 10.1509/jppm.25.1.90
- Walker, M. P., Brakefield, T., Hobson, J. A., and Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature* 425, 616–620. doi: 10.1038/nature01930
- Wood, W. M., and Rünger, D. (2016). Psychology of habit. *Annu. Rev. Psychol.* 67, 289–314. doi: 10.1146/annurev-psych-122414-033417

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Banca, McNamee, Piercy, Luo and Robbins. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Law of Recency: An Episodic Stimulus-Response Retrieval Account of Habit Acquisition

Carina G. Giesen¹, James R. Schmidt² and Klaus Rothermund^{1*}

¹ Department of Psychology, Friedrich Schiller University Jena, Jena, Germany, ² Department of Psychology, Université Bourgogne Franche-Comté, Dijon, France

OPEN ACCESS

Edited by:

John A. Bargh,
Yale University, United States

Reviewed by:

Robert J. Lowe,
University of Gothenburg, Sweden
Nart Bedin Atalay,
TOBB University of Economics
and Technology, Turkey
David Luque,
University of New South Wales,
Australia

*Correspondence:

Klaus Rothermund
klaus.rothermund@uni-jena.de

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 18 March 2019

Accepted: 11 December 2019

Published: 15 January 2020

Citation:

Giesen CG, Schmidt JR and
Rothermund K (2020) The Law of
Recency: An Episodic
Stimulus-Response Retrieval Account
of Habit Acquisition.
Front. Psychol. 10:2927.
doi: 10.3389/fpsyg.2019.02927

A habit is a regularity in automatic responding to a specific situation. Classical learning psychology explains the emergence of habits by an extended learning history during which the response becomes associated to the situation (learning of stimulus-response associations) as a function of practice (“law of exercise”) and/or reinforcement (“law of effect”). In this paper, we propose the “law of recency” as another route to habit acquisition that draws on episodic memory models of automatic response regulation. According to this account, habitual responding results from (a) storing stimulus-response episodes in memory, and (b) retrieving these episodes when encountering the stimulus again. This leads to a reactivation of the response that was bound to the stimulus (c) even in the absence of extended practice and reinforcement. As a measure of habit formation, we used a modified color-word contingency learning (CL) paradigm, in which irrelevant stimulus features (i.e., word meaning) were predictive of the to-be-executed color categorization response. The paradigm we developed allowed us to assess effects of global CL and of an instance-based episodic response retrieval simultaneously within the same experiment. Two experiments revealed robust CL as well as episodic response retrieval effects. Importantly, these effects were not independent: Controlling for response retrieval effects eliminated effects of CL, which supports the claim that habit formation can be mediated by episodic retrieval processes, and that short-term binding effects are not fundamentally separate from long-term learning processes. Our findings have theoretical and practical implications regarding (a) models of long-term learning, and (b) the emergence and change of habitual responding.

Keywords: law of recency, law of exercise, law of effect, habit acquisition, stimulus-response binding, event files, episodic response retrieval, contingency learning

INTRODUCTION

In the cafeteria, you might notice that you bought some fries for lunch – yet again – instead of the much healthier salad. After a long day at work, you might find yourself taking the way home to your old place rather than the new one you recently moved to. Everyone knows situations like these, in which we behave by mere force of habit, sometimes even against our good intentions. But how did we acquire these habits? What is the source of habitual behavior? Psychologists have pondered over the processes underlying habit formation for over a century now.

Currently, the theoretical terrain on habit acquisition is dominated by two accounts, based on either the “law of effect” or the “law of exercise” (for overviews, see, e.g., Wood and R  nger, 2016; Wood, 2017; Miller et al., 2019). Early accounts explained habit acquisition in terms of operant conditioning (Thorndike, 1898; Hull, 1943). According to Hull (1943), habit strength is a direct function of the reinforcement history of a particular response in a specific situation. Whereas responding is initially based on the trial-and-error principle, the likelihood of showing a particular response again in a given situation will increase if the response was rewarded, but will decrease if the response was punished in the past. This emergence of habits for behaviors that were reinforced before is called the “law of effect” (Thorndike, 1898). Learning psychology has seen some debates of what counts a reward or reinforcer, with suggestions ranging from stimuli that reduce states of deprivation of biological needs and that are adaptive for survival (Hull, 1943), to more formal definitions focusing on the transitionally stable quality of a stimulus to increase the probability of different behaviors of a specific organism (Meehl, 1950), to opportunities to execute behaviors that are chosen with high frequency under free-choice conditions (Premack, 1965). A detailed discussion of these accounts is beyond the scope of this article, but it is evident that rewards can also be subtle effects and qualities of the behaviors that are studied. We will take up this important point again in the General Discussion (section “What Is a Reward?”).

Even early learning psychology, however, already had another explanation of habit acquisition that was independent of reinforcement: According to the “law of exercise,” habits can emerge as a mere result of repeating the same behavior in the same situation over and over again (Thorndike, 1898). Since reinforcement and repetition are typically confounded, the outcome devaluation paradigm has been used in order to assess habitual behavior that is independent of reward or valuable outcomes (Dickinson, 1985). Several studies have shown that although outcomes have a strong influence on instrumental behavior, behavior that has been highly overlearned in many repetitions continues to be shown even in the absence of reward or after the outcome has lost all its reinforcing qualities. For instance, the behavior might still be present after having paired the outcome with shock or after providing so much of the reward (e.g., food) that the animal is completely satiated, resulting in a refusal to consume the previously rewarding outcome when it is available (e.g., Rescorla, 1991; Colwill, 1993). These findings provide unambiguous evidence that mere repetition of a response can produce habitual behavior independently of expected reward or reinforcement. In sum, then, the concept of a habit captures the fact that behaviors eventually are elicited in a more or less automatic fashion by situational cues, even in the absence of rewards and intentions.

The concept of a habit can be broadly defined to reflect automatic operant behavior that is elicited by certain stimuli or situations. According to this definition, habitual behavior is necessarily characterized as being automatic, although the reverse does not hold: Behaviors can share features of automaticity, without necessarily reflecting habitual behavior (e.g., Amodio and Ratner, 2011). For instance, behavior that

is based on instincts or autonomous reflexes (“respondent behavior”) can operate automatically without being habitual, and automatic processes without a behavioral component are also not considered to reflect habits (e.g., automatic semantic activation). Thus, a crucial feature that characterizes habits on top of their reflecting features of automaticity is that habits refer to operant behaviors that result from some kind of learning or experience.

Importantly, this definition describes what a habit is, but it does not imply specific assumptions regarding its explanation. That is, a habit can be observed regardless of whether the behavior was reinforced in a certain situation or whether it was just executed (repeatedly or just once) in this situation (without necessarily having been reinforced). Relatedly, the definition of habitual behavior is mute with regard to its underlying causes. Habits might reflect associations between situational cues and responses that will emerge gradually as a consequence of repeated and/or rewarded pairings, as early learning theories have assumed. Again, however, alternative conceptions are possible that explain habitual behavior by automatic memory processes, without necessarily drawing on the concept of associations. Whatever the correct theoretical explanation is, characterizing a behavior as habitual implies that it is assumed to share some features of automaticity (e.g., goal-independence, efficiency, speed, unawareness; Bargh, 1994; Moors and De Houwer, 2006), that it is categorized as operant behavior, and that it is somehow related to learning/experience.¹

The present study proposes an alternative view according to which habit acquisition can be explained by recent cognitive accounts of automatic action regulation that draw on episodic memory models (indeed, this view is also suggested by Wood and R  nger, 2016). In line with such a perspective, we propose the “law of recency” as another route to habitual behavior. According to this instance-based account of habit acquisition, having executed a behavior in a specific situation increases the likelihood of executing the same behavior in the same situation again when it is encountered the next time, even in the absence of reward and although the behavior was executed only once (i.e., in the absence of multiple repetitions). The core focus of our study is to provide a test of the law of recency, and to dissociate influences of an instance-based retrieval of the behavior that was executed during the last encounter with the current situation from alternative explanations in terms of multiple repetitions (global contingencies) and reward. Specifically, we investigate whether habitual behavior resulting from pairings between a stimulus and a response can be explained in terms of such an

¹Researchers in the tradition of the law of exercise have claimed that outcome devaluation (Adams, 1982) is a necessary criterion for establishing that behavior is habitual (e.g., Dickinson, 1985; Ostlund and Balleine, 2007). Although we agree that demonstrating the stability of a behavior against outcome devaluation is important for establishing that a certain behavior is habitual rather than instrumental, we do not think that it should be considered to be a necessary criterion to establish behaviors as habitual. In some cases (e.g., the present study), behaviors are not systematically linked to any (positive or negative) outcomes in the first place. If habitual behavior is established in the absence of reinforcement, the outcome devaluation procedure is not directly applicable (if there is no reward that is linked to the habitual behavior in question, then it cannot be devalued). In addition, alternative criteria can be used to establish that the behavior in question shares features of automaticity (e.g., unawareness, efficiency [resource independence]), and thus can be established as being habitual.

episodic retrieval of responses. To provide a pure test of habitual behavior resulting from previous pairings, we used a paradigm that does not contain any kind of rewards, thus effectively ruling out any influence of reinforcement on the emergence of habits in our study.

It is important to note that our study does not claim to show that reinforcement is irrelevant for the emergence of habits. We just want to limit our study to the investigation of mere repetition effects, without making any claims regarding the validity of the “law of effect” or its underlying causes. Even if we fully succeeded in explaining effects of practice on the basis of episodic response retrieval, this would still leave room for the possibility of reinforcement having an independent, additional effect on habit acquisition, which may or may not be mediated by episodic retrieval.

Episodic Memory Models of Automatic, Stimulus-Based Action Regulation

The idea of stimulus-response bindings (“event files,” Hommel, 1998) is a central characteristic for stimulus-based action regulation accounts (Logan, 1988; Hommel et al., 2001; Rothermund et al., 2005). Accordingly, whenever a response is executed to a stimulus, their mental codes become integrated, resulting in episodic stimulus-response bindings that are stored in memory. Stimulus repetition on a later occasion triggers retrieval of the response that was bound to the stimulus. This will facilitate or impede performance, depending on whether the retrieved response is appropriate or not on the current trial. To date, a burgeoning amount of findings attests that storage and retrieval of these episodic stimulus-response bindings are pervasive principles of action regulation and apply to a broad scope of stimuli and responses (for an overview, see Henson et al., 2014).

A crucial difference between stimulus-response bindings and stimulus-response associations in standard learning paradigms is that stimuli and responses are typically not correlated in designs which are used to investigate stimulus-response binding and retrieval (SRBR) effects. Specifically, SRBR effects are assessed in a sequential trial design, in which the factors Stimulus Relation (i.e., does the stimulus repeat or change from trial $n-1$ to trial n) and Response Relation (i.e., does the response repeat or change from trial $n-1$ to trial n) are orthogonally manipulated. In other words, there simply is *nothing to learn* over the course of the experiment in these tasks, since each word is presented equally often with each response. Yet, it is an unresolved issue how SRBR effects relate to learning effects. Although this is a much debated and discussed topic, empirical findings so far are scarce and unsystematic (Colzato et al., 2006; Herwig and Waszak, 2012; Moeller and Frings, 2014, 2017; Schmidt et al., 2016, 2019). Some of these studies suggest that SRBR effects are only a transient “by-product” of distributed processing and intentional action planning but are unrelated to persistent learning effects (Colzato et al., 2006; Herwig and Waszak, 2012; Moeller and Frings, 2014, 2017). In turn, other studies favor the view that short-term binding effects and more persistent learning effects are essentially the same thing, only studied at different time scales (Schmidt

et al., 2019). Hence, one could conceive of SRBR effects as “one trial learning” that serves as a founding stone for contingent associations which are stored in memory on a long-term basis. This reasoning is further supported by recent computational modeling simulations (Schmidt et al., 2016) which indicate that both types of effects might result from the same underlying learning mechanism.

An Episodic Account of Habit Acquisition

According to the present account, habitual responding results from (a) storing stimulus-response bindings in memory and (b) retrieving the most recent of these bindings when the stimulus is re-encountered on a later occasion. This leads to a reactivation of the response that was bound to the stimulus during the last occurrence of the stimulus. In other words, habitual responding can be understood as a result of previous stimulus-response bindings that emerged over the course of the experiment. First and foremost, we propose this account – the “law of recency” – as an explanation for habits that are based on repetition. According to this account, it is always the most recent instance of the current stimulus situation that is retrieved on the next occasion, and that influences responding in the current situation via a retrieval of the response that was shown during the previous instance. Our account provides an alternative explanation of repetition effects that competes with association- or frequency-based accounts of repetition-based habits that were proposed in the tradition of the law of exercise (e.g., Miller et al., 2019). The crucial difference between the two accounts is that according to the law of recency, it is the most recent episode that drives responding, whereas according to the law of exercise, the global frequency or contingency of responding to all previous occurrences of this situation is the decisive factor. To distinguish between these accounts, the behavior that was shown during the last occurrence has to be manipulated independently of the global context in which this behavior has been shown.² In the current study, we will manipulate these two factors independently.

Importantly, and in contrast to existing accounts on habit formation, stimulus-response bindings can emerge even in the absence of past reinforcement and hence do not rely on any behavior-reward correlation. Hence, our account predicts that habit formation should be possible even though responses are never reinforced. Importantly, our study is not meant to rule out any effects of reinforcement on habit acquisition (“law of effect”), nor do we test whether any such effect is due to episodic retrieval processes. We just wanted to make sure that the habitual behavior we studied reflects pure repetition effects, which is why we studied behavior in the absence of any tangible rewards.

To test the underlying causes of habit formation in the absence of reinforcement, we used a modified color-word contingency learning (CL) paradigm (e.g., Schmidt et al., 2007; for a review, see MacLeod, 2019). In our task, participants classify the color of printed words (neutral adjectives) on each trial. However, each word is presented most often in two of four colors (high

²A similar rationale is used in studies investigating the competing influence of local and global contexts on thought and action (e.g., Meier and Kane, 2013; MacLellan et al., 2015; Fröber et al., 2018).

contingency combinations) and less often in the remaining two colors (low contingency combinations, see **Table 1**). Although the word meaning is irrelevant for the color categorization task, participants learn the contingencies between word stimuli and color responses. Learning of contingencies served as an index of habit formation and is reflected in faster and more accurate performance on high compared with low contingency combinations (Schmidt et al., 2007; for related work, see Miller, 1987; Carlson and Flowers, 1996).

Deviating from previous research on CL, we chose to study the effects of comparatively weak and complex contingencies on behavior. Previous research already showed that participants produce contingency effects even when unaware of the contingencies, thus establishing the automatic (i.e., habitual) nature of behavior that is driven by the CL (Schmidt et al., 2007). Furthermore, learning in this paradigm is incidental, as participants are not informed in advance of contingencies and the words are irrelevant to the main task of color identification. In our study, we used much weaker contingencies than in the original paradigm, and we employed more complex rules in which one stimulus was systematically paired with two instead of just one response. Through these measures, the contingencies in our study were more subtle and much harder to detect, and they could not be translated into simple $S \rightarrow R$ rules (due to the dual response pairings), making it even less likely that our participants would be able to use the contingencies strategically. By implication, any effect of CL in our study can be taken as evidence for automatic behavior regulation, thus representing an index of habitual responding.

The core idea of our study is that habit acquisition that is based on CL can be explained in terms of an episodic retrieval of previous stimulus-response episodes (cf. Schmidt et al., submitted). For high contingency trials, probabilities are above chance (which is $p = 0.25$ in a four color choice task) that the word of the current trial was presented in the same color also during its last occurrence (in our study, this probability is $p = 0.33$ and $p = 0.40$ for Experiments 1 and 2, respectively), whereas for low contingency trials, probabilities of word-color repetitions are lower than chance ($p = 0.17$ and $p = 0.10$ for

Experiments 1 and 2, respectively). By implication, retrieving the response that was stored together with the word during its last occurrence will facilitate responding for 33% (Experiment 1) or 40% (Experiment 2) of the high contingency trials, but for only 17% (Experiment 1) or 10% (Experiment 2) of the low contingency trials. Likewise, response retrieval of the last episode in which the word was presented will activate a different response and will delay responding for 67% (Experiment 1) or 60% (Experiment 2) of the high contingency trials but for 83% (Experiment 1) or 90% (Experiment 2) of the low contingency trials. Our study aims to test the hypothesis that retrieving the response from the last occurrence of the word stimulus drives the CL effect, and is the underlying mechanism of habit formation. We predicted that controlling for these differences in retrieving either the same or a different response should eliminate the global CL effect (cf. Schmidt et al., 2019).

As a crucial design feature of our study, we aimed to assess episodic response retrieval effects and CL effects *simultaneously*, that is, in the very same experiment. Our study had the following expectations: First, we predicted to find robust CL effects. Second, we predicted to find response retrieval effects (reflected in an effect of response relation regarding the current and previous occurrence of the word). Third, and most central to our research aims, we tested whether response retrieval effects can explain habit formation (i.e., the CL effect). We expected that CL will be substantially reduced (or even eliminated) as soon as we control for differences in response retrieval effects. Such a pattern of results would support the law of recency as an explanation of habitual behavior, while at same time controlling for (and ruling out) an alternative explanation in terms of the law of exercise (i.e., a global, frequency based account of repetition effects).

EXPERIMENT 1

Method

Participants

Thirty native German-speaking FSU Jena students (18 female; $M_{\text{age}} = 23.03$ years; range: 18–30 years) took part in the experiment. *A priori* power calculations (G*Power 3; Faul et al., 2007) showed that we need at least 27 participants to detect a medium sized effects ($d = 0.5$) with sufficient power ($1 - \beta \geq 0.8$). Up to six participants were tested in parallel. Each participant was seated individually in a small cubicle. Sessions lasted 25 min. Participants received €2.50 for their participation plus a chocolate bar or ice cream voucher if they fulfilled criteria for speed (more than 80% of all reaction times [RT] faster than 1000 ms) and accuracy (less than 15% errors) in the experimental trials. In accordance with guidelines of the American Psychological Association, prior to the study, all participants gave their explicit consent to take part via pressing the “j” key of the keyboard (responses to the informed consent were saved for each participant). The study was canceled before any data collection started for participants who did not give their consent. An ethics approval was not required as per applicable institutional and national guidelines and regulations because no cover-story or otherwise misleading or suggestive information was conveyed

TABLE 1 | Example for word-color contingency manipulation in Experiments 1 and 2.

	Color responses	Word stimuli			
		‘warm’	‘klein’	‘ganz’	‘fast’
Exp 1	Red	2 (hc)	2 (hc)	1 (lc)	1 (lc)
	Green	2 (hc)	1 (lc)	2 (hc)	1 (lc)
	Blue	1 (lc)	2 (hc)	1 (lc)	2 (hc)
	Yellow	1 (lc)	1 (lc)	2 (hc)	2 (hc)
Exp 2	Red	4 (hc)	4 (hc)	1 (lc)	1 (lc)
	Green	4 (hc)	1 (lc)	4 (hc)	1 (lc)
	Blue	1 (lc)	4 (hc)	1 (lc)	4 (hc)
	yellow	1 (lc)	1 (lc)	4 (hc)	4 (hc)

Exp, Experiment. Hc, high contingency word-color response combinations; lc, low contingency word-color response combinations.

to participants (this procedure is in accordance with the ethical standards at the Institute of Psychology of the FSU Jena).

Apparatus and Stimuli

The experiment was programmed with E-Prime 3.0. Stimuli were the four neutral monosyllabic German adjectives “warm” (“warm”), “klein” (“small”), “ganz” (“whole”) and “fast” (“almost”). Stimuli were presented in Times New Roman font (16 pts.) on a black background on a 17" inch CRT screen. A response pad, attached to the computer via the parallel port, served to collect responses. Participants responded by pressing four colored keys on the response pad with their middle and index fingers of the left and right hand (key order from left middle to right middle finger: red, green, blue, yellow). A fifth key, operated via (left or right) thumb press, was labeled with “Los” (“go”) and served to start the experiment.

Design

Central to our study, we manipulated the contingency between word stimuli and color responses: Each of the four word stimuli appeared in each of the four colors; however, combinations differed in their frequencies. Specifically, each word appeared twice as often in two colors (high contingency combinations) than in the two remaining colors (low contingency combinations), yielding a contingency ratio of 2:1. Thus, each word was predictive of two colors/responses (high contingency combinations) and non-predictive of the other two colors/responses (low contingency combinations).³

The contingency manipulation resulted in 16 different word-color combinations. Given that high contingency combinations were shown twice as often as low contingency combinations, this amounted to a total of 24 word-color combinations (i.e., 16 word-color combinations plus 8 “duplicates” resulting from the 2:1 contingency manipulation, see **Table 1**). Each word-color combination was presented as stimulus in trial *n*-1 and as stimulus in trial *n*, resulting in a total of $24 \times 24 = 576$ experimental trials.

As another advantage, the chosen design allowed us to analyze immediate trial sequences to assess SRBR effects in a systematic and fully controlled manner. For immediate trial sequences within each experimental list, we realized a maximally balanced 2 (contingency of present trial *n*: high vs. low) \times 2 (contingency of preceding trial *n*-1: high vs. low) \times Stimulus Relation between trial *n* and trial *n*-1 (stimulus repetition [SR; 25%] vs. stimulus change [SC; 75%]) \times Response Relation between trial *n* and trial *n*-1 (response repetition [RR; 25%] vs. response change [RC; 75%]) design. Note that trial sequences for the SR-RR cell are only possible when trial *n*-1 and trial *n* both represent high contingency trials or when both represent low contingency trials (i.e., when the contingency *matches* between a trial sequence). Put differently, if both the stimulus and the response repeat in a given trial *n* from the previous trial *n*-1, then the contingency from the

trial *n*-1 has to repeat as well. In turn, contingency *mismatches* (e.g., high contingency on trial *n*-1, but low contingency on trial *n*, or vice versa) are impossible to create within the SR-RR cell. Thus, to analyze SRBR effects, only trial sequences with matching contingencies were regarded.

Procedure

Instructions were given on screen. Participants were informed that on every trial, a word stimulus would first appear in white font and then change its color to red, green, blue, or yellow. Their task was to categorize the color of each word stimulus by pressing the corresponding key on the response pad. After reading the instructions, participants worked through 24 practice trials that were identical to trials in the experimental blocks. The practice block was repeated if more than 20% errors were committed. If error rates still exceeded 20% after the third run of the practice block, the experiment was terminated (however, this never occurred during data collection). Upon successful completion of the practice block, the main experiment started, consisting of 576 experimental plus 1 filler trial (i.e., trial 1, which had no preceding trial). After 288 trials were completed, participants were given a small, self-paced break. The first trial after the break was identical to the last trial before the break and served as filler. Filler trials were not analyzed. Experimental trials were presented in a continuous fashion. At the end of the experiment, participants were rewarded accordingly.

Each trial started with a fixation cross (500 ms), followed by a white word for a variable duration by randomly selecting one out of five possible durations (150, 200, 250, 300, or 350 ms) after which the word changed its color until key press. Erroneous responses elicited the feedback message “Fehler – reagieren sorgfältiger! Weiter mit ‘Los’ Taste” (“Error – be more accurate! Continue with ‘go’ key...”). Responses slower than 1000 ms elicited the feedback message “Zu langsam – reagieren schneller! Weiter mit ‘Los’ Taste” (“Too slow – respond faster! Continue with ‘go’ key...”). Feedback was displayed in white font on red background until key press. Then, the next trial started.

Results

Trials with erroneous responses (6.8%) and RT outliers⁴ (2.6%) were excluded from all analyses.

Contingency Learning Effects

We compared performance in low contingency ($M_{RT} = 534$ ms; $M_{err} = 6.7\%$) with high contingency trials ($M_{RT} = 528$ ms; $M_{err} = 6.9\%$). For RTs, this comparison yielded a significant CL effect of $\Delta_{low-high} = 6$ ms, $t(29) = 3.13$, $p = 0.004$, $d_z = 0.57$, $BF_{10} = 9.08$. For error rates, the effect was not significant ($\Delta_{low-high} = -0.2\%$, $|t| < 1$).

³ Another reason for making each word predictive of two colors was to investigate stimulus-response binding and retrieval effects for immediate sequences in a design that is maximally balanced with regard to contingency as well as stimulus- and response relations. This aspect of our studies, however, was not the core focus of the current paper.

⁴ Probe RT below 250 ms or more than 1.5 interquartile ranges above the third quartile of the individual RT distribution were regarded as outliers (Tukey, 1977). Note that results were virtually identical when trials were filtered according to the “far out” criterion (i.e., exclusion of RTs more than 3 interquartile ranges above the third quartile of the individual RT distribution) or when sample based RT distributions were used instead of individual RT distributions.

Explaining Contingency Learning Effects by Response Retrieval Effects

We investigated whether response retrieval effects influenced responding, and whether they can explain CL effects. To this end, every trial was referenced back to the *last prior occurrence* of the current stimulus – effectively, this implies that this analysis is based on stimulus repetitions (see **Figure 1B**). Furthermore, stimulus repetition trials were coded with regard to two additional factors: First, we coded the relation between the responses to the word in the current trial as well as during its last occurrence, which could be the same or different (factor Previous Response). Second, we coded how distant the last occurrence was from the present stimulus repetition trial (factor Distance: immediate vs. non-immediate stimulus repetition). Distance was coded as a binary factor with “immediate stimulus repetition” indicating that the present stimulus was repeated from the immediately preceding trial $n-1$. In turn, trials in which the last occurrence of the current word stimulus were further away (i.e., trials $n-2$ to $n-30$) were coded as “non-immediate stimulus repetition” (see **Figure 1B** for illustrations). Only last occurrences in which a correct response was committed were included. Thus, data were analyzed in a 2 (contingency: high vs. low) \times 2 (previous response: same vs. different) \times 2 (distance: immediate vs. non-immediate stimulus repetition) ANOVA on mean RTs (the pattern of means is shown in **Table 3**).

Although we obtained a significant CL effect in our first analysis (without controlling for SRBR effects, see above), the main effect of contingency was no longer significant in the final analysis, $F < 1$, $BF_{01} = 6.79$. Instead, the ANOVA yielded a main effect of previous response, $F(1,29) = 179.96$, $p < 0.001$, $\eta_p^2 = 0.86$, $BF_{10} = 3.817e + 21$, indicating that performance was faster if the current stimulus repetition required the same previous response ($M = 480$ ms) compared with a different previous response ($M = 548$ ms). This pattern of findings confirms our hypothesis that controlling for episodic SRBR effects effectively eliminated the CL effect in Experiment 1. The main effect of the distance factor was also significant, $F(1,29) = 141.22$, $p < 0.001$, $\eta_p^2 = 0.83$, indicating that performance was generally faster for immediate stimulus repetitions ($M = 497$ ms) compared to trials in which the last occurrence of the same word stimulus was more distant ($M = 531$ ms). Main effects were qualified by a Distance \times Previous Response interaction, $F(1,29) = 322.52$, $p < 0.001$, $\eta_p^2 = 0.92$. Follow-up tests showed that response retrieval effects were significantly stronger for immediate stimulus repetitions ($M_{\text{same response}} = 432$ ms; $M_{\text{different response}} = 562$ ms; $t[29] = 16.53$, $p < 0.001$, $d_z = 2.78$), but were also significant for stimulus repetitions of more distant trials ($M_{\text{same response}} = 528$ ms; $M_{\text{different response}} = 534$ ms; $t[29] = 1.76$, $p = 0.045$, one-tailed, $d_z = 0.32$). No other effect was significant (all F s < 2.9 , all p s ≥ 0.10).

Multi-Level Analyses

We also conducted multi-level analyses on the basis of individual trials, treating trials as nested within subjects. In these analyses, CL and response retrieval reflect between factors (on the level of trials), which allows us to simulate a stepwise regression approach

to test whether entering response retrieval as an additional predictor in a second step eliminates effects of CL that had been significant when entered as a single predictor into the regression equation in step 1. The multi-level analyses also allow us to treat distance of the last occurrence as a continuous predictor, so we can calculate at which distance the effect of response retrieval effectively becomes zero.

A multilevel analysis with contingency (high frequency = 1 vs. low frequency = 2) as the only level 1 predictor, allowing for random intercepts and slopes, yields a significant CL effect, $\beta = 6.19$, $t = 3.15$, $p = 0.004$, replicating the effect of the previous analysis. Adding Previous Response (same = 1 vs. different = 2), as an additional level 1 predictor in a second step produced a highly significant effect for this variable, $\beta = 34.21$, $t = 9.30$, $p < 0.001$, and it rendered the effect for the CL variable non-significant, $\beta = 0.59$, $t = 0.28$, $p = 0.78$. Effectively, then, although CL predicts RT when considered in isolation, this effect is fully explained by response retrieval.

Although we were primarily interested in the main effects of CL and response retrieval, the multinomial model also allows us to introduce an interaction term for the two variables (CL \times previous response). Adding the product term in a third step yields a beta that is positive and significant ($t = 2.19$, $p = 0.029$). This interaction indicates that effects of response retrieval were slightly stronger for low contingency trials, that is, responses were slowest for low contingency trials in the “different response” condition. A plausible explanation for this asymmetry is that response retrieval may not only be influenced by the last occurrence of the stimulus but may probably also sometimes retrieve an earlier episode in which the stimulus was presented. For low contingency trials in the “different response” condition, such a retrieval of an earlier episode will retrieve a different response in 83% of these trials. For high contingency trials in the different response condition, only 67% of the previous occurrences of the word contained a different response, 33% of the trials contained an identical response. It is thus possible that in some high contingency trials in the “different response” condition, the correct response was retrieved from an earlier episode (leading to a facilitative effect that counteracted the delay effect in the “different response” condition), even though the last occurrence of the word was paired with a different response.

Another multi-level analysis was used to evaluate the moderating effect of distance on effects of response retrieval. For this purpose, we predicted RT with the previous response factor (pr), distance (d), and their interaction (pr \times d). We also added a squared term for distance (d^2) and the interaction of this term with previous response (pr \times d^2) to allow for a non-linear decline of the influence of response retrieval with increasing distance. The full model yielded significant effects for all predictors (all $p < 0.001$). The regression equation is given by the following set of parameter values: $RT = 341 + 105.31pr + 46.72d - 2.11d^2 - 25.43pr \times d + 1.15pr \times d^2$. Transforming this equation into a form that represents the slope of pr as function of d and d^2 gives: $RT = 341 + (105.31 - 25.43d + 1.15d^2)pr + 46.72d - 2.11d^2$. Setting the quadratic formula in brackets that represents the slope for pr to zero and solving for d yields $d = 5.52$, that

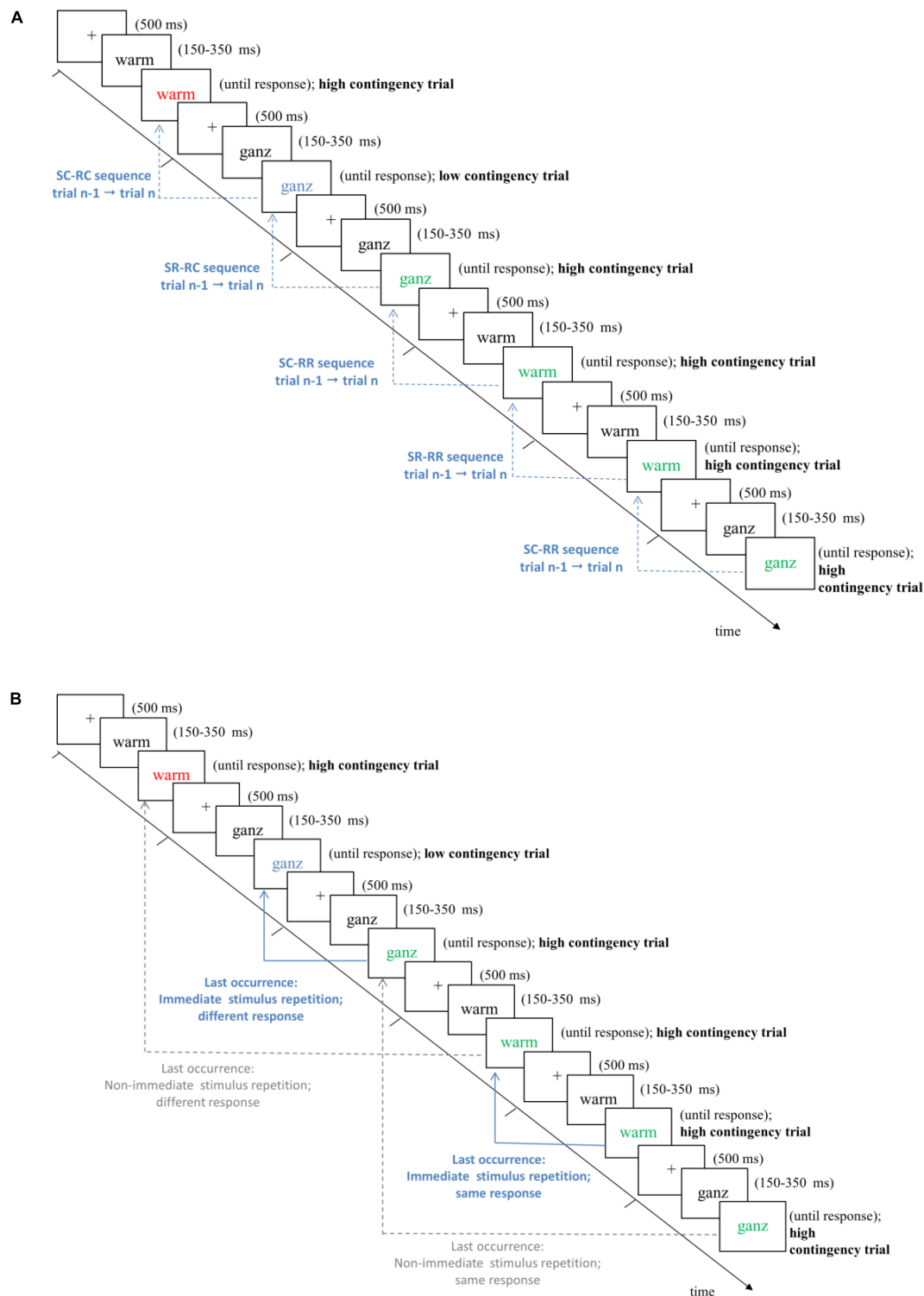


FIGURE 1 | Schematic trial procedure in Experiments 1 and 2. Note that in the experiments, all stimuli were presented on black background in white font or in the respective colors (see **Table 1**). For both figures, we inverted the coloring scheme only for illustrative purposes. Stimuli are not drawn to scale. Trials are classified as high vs. low contingency trials (for details, see **Table 1**). Arrows in **(A)** illustrate different trial types for immediate sequence effects from trial n-1 to trial n to test for immediate SRBR effects (SR, stimulus repetition; SC, stimulus change; RR, response repetition; RC, response change). Arrows in **(B)** illustrate trial classification for the central analyses of interest to explain contingency learning effects by response retrieval effects, i.e., whether a given trial reflected an immediate (solid/blue lines) vs. non-immediate (dotted/gray lines) stimulus repetition trial (factor Distance) with same or different response (factor Previous Response) compared to the last occurrence of the stimulus word. See main text for details.

is, the slope for response retrieval becomes zero at a distance between 5 and 6 trials.

Stimulus-Response Binding and Retrieval Effects

To test for SRBR effects, we analyzed immediate sequence effects from trial $n-1$ to trial n (cf. **Figure 1A**). In these analyses, only sequences with matching contingencies were regarded (see Method section for details). We performed two separate $2 \times 2 \times 2$ repeated measurement analyses of variance (ANOVA) with the factors stimulus relation (stimulus repetition vs. stimulus change from trial $n-1$ to trial n), response relation (response repetition vs. response change from trial $n-1$ to trial n), and type of prime-probe contingency match (both trial $n-1$ and trial n high contingency vs. both low contingency) on trial n performance (i.e., RTs and error rates; see **Table 2** for means).

For RTs, the ANOVA yielded significant main effects of stimulus relation, $F(1,29) = 7.74$, $p = 0.009$, $\eta_p^2 = 0.21$, and response relation, $F(1,29) = 174.16$, $p < 0.001$, $\eta_p^2 = 0.86$, indicating that RTs were faster for stimulus repetition ($M = 495$ ms) compared with stimulus change trials ($M = 505$ ms) and that probe RTs were faster for response repetitions ($M = 444$ ms) than for response changes ($M = 556$ ms). Most importantly, both effects were qualified by a significant Stimulus Relation \times Response Relation interaction, $F(1,29) = 39.62$, $p < 0.001$, $\eta_p^2 = 0.58$, that reflected the typical pattern of SRBR effects. Follow-up tests showed that compared to stimulus change from trial $n-1$ to trial n , stimulus repetition significantly sped up performance by $\Delta_{\text{SCRR-SRRR}} = 29$ ms, $t(29) = 5.48$, $p < 0.001$, $d_z = 1.00$, for response repetition. In turn, stimulus repetition (compared with stimulus change from trial $n-1$ to n) significantly slowed down performance by $\Delta_{\text{SCRC-SRRC}} = -10$ ms, $t(29) = 2.42$, $p = 0.022$, $d_z = 0.44$, for

response changes. No other effect was significant (all $F_s < 1.06$, all $p_s > 0.30$).

For error rates, the same ANOVA yielded only a main effect of response relation, $F(1,29) = 65.81$, $p < 0.001$, $\eta_p^2 = 0.69$, indicating that participants made fewer errors on response repetition ($M = 2.4\%$) than on response change sequences ($M = 7.4\%$). No other effect was significant (all $F_s < 3.2$, all $p_s > 0.08$).

Discussion

The results of Experiment 1 are clear-cut: First, we obtained a CL effect, indicating that participants incidentally learned the word-color response associations over the course of the experiment. Second, we obtained robust response retrieval effects, reflecting faster RTs in the current trial when the same response had been given during the last occurrence of the word stimulus that was also presented in the current trial, compared to trials when a different response had been executed during the last occurrence. Third and most central to our research aims, the CL effect was effectively eliminated after controlling for effects of response retrieval. This pattern of findings emerged both for ANOVA analyses with aggregated data and also in multilevel analyses in which CL and response retrieval were coded on a trial level. Importantly, effects of response retrieval were not limited to the immediately preceding trial, but were found for distances up to 5–6 trials, ruling out alternative explanations of the effect in terms of mere response repetition. For immediate stimulus repetition sequences (distance = 1), effects of response retrieval are identical to effects of response repetition, until sequences in which the stimulus changes are used as a baseline. These analyses replicated the standard pattern of SRBR effects that obtained in many previous studies (Rothermund et al., 2005; see also

TABLE 2 | Results for SRBR effects (probe RT and error rates) in Experiments 1 and 2.

	Type of contingency match trial $n-1 \rightarrow$ trial n	Response relation trial $n-1 \rightarrow$ trial n	Stimulus relation trial $n-1 \rightarrow$ trial n	Probe RT (ms)		Probe errors (%)	
				<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Exp 1	High-high	RR	SR	432	40	1.8	3.3
			SC	456	41	3.9	4.7
		RC	SR	559	58	7.5	4.1
			SC	548	50	8.1	4.8
	Low-low	RR	SR	426	45	1.4	4.2
			SC	460	38	2.6	5.3
		RC	SR	562	60	6.5	9.5
			SC	554	49	7.6	4.7
Exp 2	High-high	RR	SR	409	33	4.2	4.7
			SC	452	38	5.5	4.4
		RC	SR	537	53	9.3	6.0
			SC	529	52	8.5	5.6
	Low-low	RR	SR	415	50	3.5	7.0
			SC	451	61	7.0	12.6
		RC	SR	553	77	13.3	19.2
			SC	551	54	11.1	11.0

Exp, Experiment. RR, response repetition. RC, response change. SR, stimulus repetition. SC, stimulus change.

Frings et al., 2007; Giesen and Rothermund, 2014), rendering explanations of response retrieval effects in terms of mere response repetition unlikely. Together, findings from Experiment 1 support predictions derived from the law of recency that episodic retrieval of responses from the most recent occurrence of the stimulus represents a central process underlying habit formation (i.e., learning of word-response contingencies). Effects of global SR contingencies were completely eliminated after controlling for an influence of the most recent last episode, which rules out frequency-based explanations (law of exercise) of habitual responding in the current study.

The CL effect observed in Experiment 1 was smaller than in previous studies [$d_z = 0.57$, reflecting a medium-sized effect according to Cohen (1969) compared with effect sizes between $d_z = 0.62$ up to $d_z = 1.24$, reflecting medium-to-large- to very-large-sized effects in Schmidt et al., 2007]. In our view, this is probably due to the fact that Experiment 1 had a contingency ratio of only 2:1, which is a rather weak contingency manipulation in and of itself and it is known that the magnitude of contingency effects is proportional to the contingency (Forrin and MacLeod, 2018; see also, Schmidt and De Houwer, 2016). The low contingency was chosen on purpose, since we wanted to make sure that contingencies went undetected, and thus could not be applied in a strategic fashion. However, being aware of the fact that single studies pose the risk of being unreliable (Cesario, 2014; see also Tversky and Kahneman, 1971) and that replication is an increasingly important research value (Nosek et al., 2012), we ran a second experiment with the aim to replicate our initial findings from Experiment 1, but with a stronger contingency manipulation (ratio of 4:1) to boost CL effects. By increasing the contingency we wanted to establish that the contingency effect itself is strong beyond any reasonable doubt, so that eliminating the effect by controlling for effects of response retrieval cannot be attributed to the contingency effect being unreliable in the first place. Although the contingency that was chosen in Experiment 2 is stronger than in Experiment 1, we want to emphasize that it is still much weaker than in previous studies that already demonstrated contingency effects in the absence of awareness (Schmidt et al., 2007). Furthermore, Experiment 2 again used contingencies in which one stimulus was predictive of two different responses, preventing a simple strategic use of the contingencies for response preparation. Furthermore, Experiment 2 was preregistered online before any data collection started (see details below).

EXPERIMENT 2

Method

Participants

Forty native German-speaking FSU Jena students (27 female; $M_{\text{age}} = 23.3$ years, range: 18–32 years) took part in the experiment. We decided to recruit a somewhat larger number of participants compared to Experiment 1 in order to be able to detect effects of CL that are even smaller than medium in size ($d = 0.4$) with sufficient power ($1 - \beta \geq 0.8$). Power calculations were conducted with G*Power 3 (Faul et al., 2007).

All participants gave their explicit verbal consent to take part prior to the study. Session duration and payment of participants were similar to Experiment 1.

Apparatus, Stimulus, Design, and Procedure

Apparatus, stimuli, design, and procedure were similar to Experiment 1 except for the following changes. In Experiment 2, we used a stronger contingency manipulation: Each word appeared four times more often in two colors (high contingency combinations) than in the two remaining colors (low contingency combinations, see Table 1), resulting in a contingency ratio of 4:1. As in Experiment 1, each word was predictive of two colors/responses (high contingency combinations), only more strongly in the present experiment, and non-predictive of the other two colors/responses (low contingency combinations).

The contingency manipulation resulted in 16 different word-color combinations plus 24 “duplicates” resulting from the 4:1 contingency manipulation, thus amounting to a total of 40 word-color combinations. To control of immediate sequences, each word-color combination was then presented as stimulus in trial $n-1$ and as stimulus in trial n , yielding a total of $40 \times 40 = 1600$ experimental trials. Since this number of experimental trials would have resulted in an experiment of unreasonable length, the total list was always split among a group of three participants, taking care that the orthogonal variation of stimulus relation and response relation was maintained for each participant. This resulted in 535 experimental trials + 1 filler trial per participant. Procedural details were again similar to Experiment 1, with the only exception that whenever a timing or response error was committed, participants had to press the correct response key (instead of the “go” key) to continue the experiment.

Preregistration

Prior to data collection, we preregistered the exact method, design, hypotheses, data preparation, and planned data analyses online at www.aspredicted.org.⁵

Results

According to the same criteria as in Experiment 1, trials with erroneous responses (8.1%) and RT outliers (3.0%) were excluded from all analyses.

Contingency Learning Effects

We compared performance in low contingency ($M_{\text{RT}} = 517$ ms; $M_{\text{err}} = 9.4\%$) with high contingency trials ($M_{\text{RT}} = 508$ ms; $M_{\text{err}} = 7.7\%$). These comparisons yielded significant CL effects for RTs, $\Delta_{\text{low-high}} = 9$ ms, $t(39) = 4.41$, $p < 0.001$, $d_z = 0.70$, $\text{BF}_{10} = 242.99$, and also for error rates, $\Delta_{\text{low-high}} = 1.6\%$, $t(39) = 2.83$, $p = 0.007$, $d_z = 0.45$, $\text{BF}_{10} = 5.64$.

Explaining Contingency Learning Effects by SRBR Effects

We investigated retrieval effects and whether the CL effect is reduced or eliminated as soon as we control for these effects, following the same approach as in Experiment 1. Thus, we

⁵<https://aspredicted.org/blind2.php>

performed a 2 (probe contingency: high vs. low) \times 2 (previous response: same vs. different) \times 2 (distance: immediate vs. non-immediate stimulus repetition) ANOVA on mean RTs (the pattern of means for this analysis is shown in **Table 3**).

Importantly, the main effect of contingency was no longer significant in this analysis, $F(1,39) = 1.34$, $p = 0.254$, $\eta_p^2 = 0.03$, $BF_{01} = 7.26$. Furthermore, the ANOVA yielded additional main effects of previous response, $F(1,39) = 276.64$, $p < 0.001$, $\eta_p^2 = 0.88$, $BF_{10} = 1.078e + 38$, indicating that performance was faster if the current stimulus repetition trial required the same previous response ($M = 452$ ms) compared with a different previous response ($M = 533$ ms). These findings replicate Experiment 1 and show that controlling for response retrieval effects effectively eliminated the CL effect also in Experiment 2. There was also a main effect of distance, $F(1,39) = 64.04$, $p < 0.001$, $\eta_p^2 = 0.62$, meaning that performance was faster if the stimulus was repeated from the immediately preceding trial $n-1$ ($M = 480$ ms) compared with non-immediate stimulus repetitions from more distant trials ($M = 504$ ms).⁶ Both main effects were again qualified by a significant Distance \times Previous Response interaction, $F(1,39) = 198.75$, $p < 0.001$, $\eta_p^2 = 0.84$. Follow-up tests showed that response retrieval effects were stronger for immediate stimulus repetitions ($M_{\text{sameresponse}} = 412$ ms; $M_{\text{differentresponse}} = 548$ ms; $t[39] = 19.07$, $p < 0.001$, $d_z = 3.01$),

⁶This main effect supposedly reflects a combination of two things: (a) response retrieval effects due to the repetition of a word are stronger for short (i.e., immediate repetitions) than for larger distances, and (b) benefits of retrieving a correct response outweigh the costs that are incurred due to retrieval of a different response (for a discussion, see Giesen and Rothermund, 2016). In sum, this leads to a facilitating effect of word repetitions on performance that is stronger for immediate than for distant repetitions.

TABLE 3 | Average RTs (and SDs) for the combinations of contingency (high vs. low), previous response (same vs. different), and distance (immediate vs. non-immediate stimulus repetition) in Experiments 1 and 2.

	Contingency	Distance	Previous response	RT (ms)	
				<i>M</i>	<i>SD</i>
Exp 1	High	Immediate stimulus repetition	Same	432	40
			Different	560	52
		Non-immediate stimulus repetition	Same	533	46
			Different	533	45
	Low	Immediate stimulus repetition	Same	432	49
			Different	564	47
		Non-immediate stimulus repetition	Same	522	39
			Different	536	46
Exp 2	High	Immediate stimulus repetition	Same	409	33
			Different	537	51
		Non-immediate stimulus repetition	Same	500	44
			Different	516	44
	Low	Immediate stimulus repetition	Same	415	50
			Different	559	60
		Non-immediate stimulus repetition	Same	482	74
			Different	520	44

Exp, Experiment.

but were still significant for stimulus repetitions of more distant trials ($M_{\text{sameresponse}} = 491$ ms; $M_{\text{differentresponse}} = 518$ ms; $t[39] = 5.16$, $p < 0.001$, $d_z = 0.82$). Two additional interactions were significant: First, the Distance \times Contingency interaction, $F(1,39) = 11.95$, $p = 0.001$, $\eta_p^2 = 0.24$. Follow-up tests revealed that distance had a stronger facilitating effect for the high contingency trials ($M_{\text{immediate}} = 473$ ms, $M_{\text{non-immediate}} = 508$ ms, $t[39] = 12.81$, $p < 0.001$, $d_z = 2.03$) than for the low contingency trials ($M_{\text{immediate}} = 487$ ms, $M_{\text{non-immediate}} = 501$ ms, $t[39] = 2.55$, $p = 0.015$, $d_z = 0.40$). Second, the interaction between Previous Response \times Contingency was significant as well, $F(1,39) = 7.79$, $p = 0.008$, $\eta_p^2 = 0.17$. Follow-up tests showed that response retrieval effects were stronger for high contingency trials ($M_{\text{sameresponse}} = 454$ ms; $M_{\text{differentresponse}} = 526$ ms; $t[39] = 17.76$, $p < 0.001$, $d_z = 2.81$) than for low contingency trials ($M_{\text{sameresponse}} = 449$ ms; $M_{\text{differentresponse}} = 539$ ms; $t[39] = 12.45$, $p < 0.001$, $d_z = 1.97$).⁷ No other effect was significant ($F < 1$, $p > 0.60$).

Multi-Level Analyses

Like in the previous experiment, we also conducted multi-level analyses on the basis of individual trials, treating trials as nested within subjects. A multilevel analysis with contingency (high frequency = 1 vs. low frequency = 2) as the only level 1 predictor, allowing for random intercepts and slopes, yields a significant CL effect, $\beta = 8.60$, $t = 4.26$, $p < 0.001$, replicating the effect of the previous analysis. Adding Previous Response (same = 1 vs. different = 2), as an additional level 1 predictor in a second step produced a highly significant effect for this variable, $\beta = 44.52$, $t = 15.71$, $p < 0.001$, and it rendered the effect for the CL variable non-significant, $\beta = 0.33$, $t = 0.16$, $p = 0.87$. Effectively, then, although CL predicts RT when considered in isolation, this effect is fully explained by response retrieval.

As in the previous experiment, the product term CL \times previous response was significant again ($t = 7.00$, $p < 0.001$). Again, this interaction indicates that effects of response retrieval were slightly stronger for low contingency trials, that is, responses were slowest for low contingency trials in the “different response” condition. For a possible explanation of this interaction effect, see the corresponding paragraph in the results section of Experiment 1.

Another multi-level analysis was used to evaluate the moderating effect of distance on effects of response retrieval. The full model yielded significant effects for all predictors (all $p \leq 0.001$). The regression equation is given by the following set of parameter values: $RT = 299 + 118.22pr + 47.30d - 2.26d^2 - 26.08pr \times d + 1.24pr \times d^2$. Transforming this equation into a form that represents the slope of pr as function of d and

⁷A possible reason for these additional interactions may be that the asymmetry of benefits and costs that are due to retrieving the correct response are stronger for high than for low contingency trials. The correct responses that are retrieved in the high contingency condition are responses that have often been paired with this word, whereas the correct responses that are retrieved in the low contingency condition are responses that have been paired with this word only rarely. Conversely, the wrong responses that are retrieved in the low contingency condition are mostly those responses that have frequently been paired with this word before.

d^2 gives: $RT = 299 + (118.22 - 26.08d + 1.24d^2) \cdot pr + 47.30d - 2.26d^2$. Setting the quadratic formula in brackets that represents the slope for pr to zero and solving for d yields $d = 6.61$, that is, the slope for response retrieval becomes zero at a distance between 6 and 7 trials.

Stimulus-Response Binding and Retrieval Effects

Like in Experiment 1, when analyzing SRBR effects, only trial $n-1$ to trial n sequences with matching contingencies were regarded. We performed two separate $2 \times 2 \times 2$ repeated measurement ANOVA with the factors stimulus relation (stimulus repetition vs. stimulus change from trial $n-1$ to trial n), response relation (response repetition vs. change from trial $n-1$ to trial n), and type of sequential contingency match (high-high vs. low-low contingency) on performance in trial n (RTs and error rates; see Table 2 for means).

For RTs, the ANOVA yielded significant main effects of contingency type, $F(1,39) = 9.76$, $p = 0.003$, $\eta_p^2 = 0.20$, stimulus relation, $F(1,39) = 24.04$, $p < 0.001$, $\eta_p^2 = 0.38$, and response relation, $F(1,39) = 222.39$, $p < 0.001$, $\eta_p^2 = 0.85$. Accordingly, RTs were faster for high-high contingency trial sequences ($M = 481$ ms) than for low-low contingency trial sequences ($M = 493$ ms); further, RTs were faster for stimulus repetition ($M = 479$ ms) than for stimulus change sequences ($M = 496$ ms); last, RTs were faster for response repetition ($M = 432$ ms) than for response change sequences ($M = 542$ ms). These main effects were qualified by several interactions: Contingency type significantly interacted with response relation, $F(1,39) = 5.41$, $p = 0.025$, $\eta_p^2 = 0.12$ (however, this interaction is not of theoretical interest and is thus not discussed further). Central to our predictions, the Stimulus Relation \times Response Relation interaction was also significant, $F(1,39) = 38.15$, $p < 0.001$, $\eta_p^2 = 0.49$. Follow-up tests showed that compared with stimulus changes from trial $n-1$ to trial n , stimulus repetition significantly sped up performance by $\Delta_{SCRR-SRRR} = 39$ ms, $t(39) = 7.02$, $p < 0.001$, $d_z = 1.10$, for response repetition sequences. In turn, stimulus repetition (compared with stimulus changes from trial $n-1$ to trial n) descriptively slowed down performance by $\Delta_{SCRC-SRRR} = -5$ ms, $t(39) = 1.19$, $p = 0.24$, $d_z = 0.18$, for response change sequences. No other effect was significant (all $F_s < 1$, all $p_s > 0.36$).

For error rates, the same ANOVA yielded only a main effect of response relation, $F(1,39) = 23.77$, $p < 0.001$, $\eta_p^2 = 0.38$, indicating that participants made fewer errors on response repetition ($M = 5.0\%$) than on response change sequences ($M = 10.5\%$). No other effect was significant (all $F_s < 3.7$, all $p_s > 0.062$).

Discussion

In Experiment 2, we used a stronger contingency manipulation to boost CL effects, which was successful. What is more, findings from Experiment 2 fully replicate the pattern of effects that were obtained in Experiment 1. In detail, we obtained a robust CL effect that was larger ($d_z = 0.70$) than in Experiment 1 ($d_z = 0.57$). Thus, we can conclude that participants incidentally acquired word-response associations over the course of the experiment. Second, we obtained robust response retrieval effects. Third, the

CL effect was again effectively eliminated when we controlled for response retrieval. Data from both experiments thus support the law of recency according to which habit formation is mediated by episodic retrieval processes, which lead to a reactivation of the response that was stored in episodic memory together with the stimulus during the most recent occurrence of the current situation. Fourth, as in the previous experiment, the influence of response retrieval was not limited to the previous trial but was visible for distances up to 6–7 trials. Finally, standard SRBR effects for immediate ($n-1 > n$) sequences in which stimulus changes are used as a baseline condition were replicated also in this experiment, indicating that response retrieval effects cannot be attributed to mere response repetition.

GENERAL DISCUSSION

The present study provides initial evidence that response retrieval effects may fully explain effects of CL (see also Schmidt et al., 2019) and thus provide a potential explanation for learning processes that eventually lead to habitual behavior. In this respect, our study supports the claim that habit formation can be mediated by episodic response retrieval processes regarding the most recent previous occurrence of the current situation (law of recency). These conclusions are supported by data of two experiments, which yielded robust evidence of the following effects: First, participants of both experiments acquired contingencies between stimulus words and color responses over the course of each experiment, leading to faster and more correct responses in trials with high frequency compared to low frequency combinations of word and color. Importantly, participants were never explicitly informed about any contingency relation between words and responses. However, incidental knowledge about the inbuilt word-response contingency was acquired nonetheless and impacted performance, leading to habitual responding. What is more, we obtained these findings even in the absence of any explicit reinforcement schedule (apart from ordinary feedback regarding errors and slow responses that was given on a negligible number of trials). Second, participants in both experiments also displayed episodic binding and retrieval effects, reflected in performance benefits when the word had been presented in the same color during the current trial and the trial in which the word had been presented during its last occurrence, reflecting a stimulus-based retrieval of the response from the previous trial. Third and most importantly, both effect types were not independent: That is, when we controlled for response retrieval effects in joint analyses, the CL effect was effectively eliminated in both experiments. Together, the present findings support the view that episodic binding effects and persistent forms of learning (e.g., habit acquisition) might result from the same underlying learning mechanism (i.e., episodic binding and retrieval). Our findings support the law of recency that explains habit acquisition as an instance-based process. According to this principle, habitual behavior emerges by retrieving and repeating a behavior that was executed during the last encounter with the current situation.

Theoretical Implications

The present study exemplifies that habit acquisition that is based on CL can be explained in terms of an episodic retrieval of previous stimulus-response episodes (for further recent evidence, see also Schmidt et al., 2019). In this respect, habitual responding can be understood as resulting from the retrieval of stimulus-response bindings that were stored in memory during the last occurrence of the situation that is now encountered again (law of recency).

Behavioral Signatures of the Law of Recency

The law of recency has a characteristic behavioral signature that has been demonstrated in numerous studies (and also in the present study) that revealed effects of SRBR. Basically, the core finding attesting to the law of recency consists in an interaction of stimulus relation and response relation in successive trials of a forced-choice reaction task: Repeating the prime stimulus in the probe leads to facilitation for response repetition sequences, but produces interference for response change sequences (Rothermund et al., 2005; see also Hommel, 1998; Mayr and Buchner, 2006; Frings et al., 2007; Giesen et al., 2012; Giesen and Rothermund, 2014). This pattern can be explained by a retrieval and reactivation of the response information of the prime during the probe. The current study demonstrates that SR binding and retrieval also plays a role in a CL paradigm (Schmidt et al., 2007), and – crucially – that effects of episodic SR binding during the last occurrence are what underlies the CL effect.

The law of recency can be used to generate alternative explanations for a wide range of experimental paradigms that investigated effects of global contexts on behavior (e.g., context effects in interference paradigms, Logan and Zbrodoff, 1979). In many of these paradigms, global effects can be tested against local effects of episodic retrieval in order to see whether the law of recency can account for these effects.

Comparing the current study with a large literature on habit acquisition addressing habit formation mostly in animals reveals a crucial difference: Classical studies typically focus on response frequencies as an outcome variable, whereas our study used performance indices (response speed and accuracy) as dependent variables. Relatedly, our study used a forced-choice color categorization task, whereas standard studies focus on a single qualitative response, the frequency of which is counted (e.g., lever pressing). The paradigm that is chosen to investigate habits may influence the results, so it is perhaps hard to compare findings across these very different experimental approaches. Despite its dissimilarity to the paradigms that were typically employed in habit research in the animal literature, focusing on RTs (instead of response frequencies) may offer some advantages for the study of habitual behavior in humans. It is no coincidence that implicit measures aiming at assessing, for instance, implicit prejudice or stereotypes typically rely on RT measures (e.g., Wittenbrink and Schwarz, 2007; Gawronski and Payne, 2010; Klauer et al., 2012). The reason is that the speed of responding is much less controllable than the execution of a specific response, and thus provides a “window to the

mind” and into automatic influences of human behavior (see Wentura and Rothermund, 2007).

Mathematical Modeling Approaches to the Law of Recency

Processes of an instance-based retrieval of previous stimulus-response episodes have been modeled mathematically within the Parallel Episodic Processing model (PEP; Schmidt et al., 2016). The model is specialized to simulate RT and error data in speeded response time tasks and has been shown to be a powerful tool that successfully simulates and explains experimental findings across a wide variety of experimental paradigms with just one mathematical architecture. For details regarding the modeling approach, we refer the reader to the original articles in which the PEP is presented (e.g., Schmidt et al., 2016). Although the PEP has been developed to account for RT data, its basic rationale might also be applied to model frequency data for single responses (e.g., lever pressing), which will require only minor adjustments in the periphery of the model, which might be a promising endeavor for the future development of the PEP and also for transferring the law of recency to the large literature that uses response frequency as the main dependent variable.

Habits Based on Repetition vs. Reinforcement

Given that CL in our experiments was obtained without linking responses to rewards, the resulting behavior reflects an instance of repetition-based rather than reward-based habits (Thorndike, 1898). Establishing habits without linking behaviors to rewards is an interesting finding in and of itself, showing that reinforcement is not a necessary condition for habit acquisition. On the other hand, we cannot say anything definitive on the possible effects that rewards may have (or not have) on episodic response retrieval processes on the basis of our study, since we did not manipulate rewards.

Separating effects of rewards from repetition can be difficult since reinforcement cannot be applied in the total absence of behavior and typically leads to a higher frequency of showing the respective behavior in the situation in which it was rewarded. As soon as frequent repetitions are involved, however, episodic response retrieval may come into play, and may explain the resulting effects. Our experimental paradigm offers an elegant solution to address this problem in future studies: Systematically varying the response that had to be performed during the last occurrence of a stimulus and either rewarding (or punishing) it or not allows for a systematic investigation of the effects of episodic binding/retrieval, reinforcement, and their interaction (preliminary evidence of a recent study, however, suggests that reinforcement of SR combinations does not have a positive effect on the strength of the resulting binding and retrieval effects; Hauber, 2019).

What Is a Reward?

Although responses that are based on the contingencies of the task are not reinforced by tangible external rewards (e.g., money), it may still be the case that they are reinforced more indirectly, in that responding in line with the contingencies on average leads to performance benefits (i.e., faster responding). Our findings

demonstrated that responding is faster in trials that confirm the contingency in comparison to those trials that are exceptions to the rule. Given that trials confirming the rule are more frequent, this effectively leads to a performance advantage. In our view, however, this difference is not yet evidence for a general performance benefit due to the contingency. The RT difference between high and low frequency trials does not reflect a difference between responses following the contingency rule and those that do not. Instead, both responses follow the contingency rule. The fundamental rationale of the CL paradigm is that contingencies affect responding in all trials, since participants do not know which sort of trial will be presented next. If participants were not influenced by the contingencies also in the low frequency trials, there probably would have been no difference because there were no costs. The claim that participants profit (overall) from applying the contingency rule requires a comparison with a no contingency condition, which was not part of our design. We thus can only speculate on whether it is plausible to assume that behavior in line with the contingency rule is rewarded. In our view, this is unlikely for our study, for the following reasons. First, due to the weak contingencies that we applied in our study, the difference in RTs between high and low frequency trials (which is the upper limit for a performance benefit in comparison to a no contingency baseline) were very small (less than 10 ms), and might not even be perceptible for participants. Second, although there may be a (negligible) performance benefit with regard to speed, responding in line with the contingencies also comes with a somewhat less negligible cost regarding errors. In Experiment 2, absolute error rates were 1.6% higher in the low frequency trials, which is a 20% increase given that overall, about 8% errors were made. Of course, we cannot exclude with certainty that the contingency might also have a beneficial effect on accuracy in the high frequency trials, but such an effect is somewhat unlikely, because the contingency favored not just one but two responses, which should increase error percentages even in high contingency trials.

We also investigated whether the contingency manipulation has an effect on the immediate trial-by-trial feedback that participants received during the task, and whether this might have affected the resulting CL effects. Participants received feedback (a) when their response was slow (i.e., above 1,000 ms) or (b) when they responded incorrectly. With regard to the speed-related feedback (i.e., “too slow” messages), the contingency conditions did not differ significantly in either of the two experiments, due to the fact that the CL effect was small in absolute size and responses were faster than the response deadline in most of the cases for both high and low frequency trials. For error feedback, there was no difference in errors between the contingency conditions in Experiment 1 (and thus no difference in error-related feedback either), but there was a small but significant effect (1.6%) in errors between the high and low frequency conditions (corresponding to a difference in error-related feedback) in Experiment 2. Controlling for this difference at the subject level, however, did not alter the CL effect for RTs at all, nor did it change any of the results of the other analyses regarding the effects of the previous occurrence. Most importantly, the interaction between CL (at

the trial level) and the error-related feedback effect of the contingency manipulation (at the person level) did not interact ($t < 1$), indicating that the CL effect was completely independent of the difference in error-related feedback that participants received. Apparently, the CL effect is unrelated to any feedback participants received.

Assuming for a moment that contingencies may nevertheless come with overall performance benefits (compared to a no contingency baseline that was not part of our study), since they reduce uncertainty, then what can one do about it? A straightforward solution would be to eliminate (extinction) or even reverse (countercondition) the contingency for some time, similar to an outcome devaluation procedure in a study of operant conditioning (cf. Schmidt et al., submitted). Based on our findings, however, this is probably not a promising strategy, since episodic retrieval is influenced by responses that were given during the last occurrence of the situation (“law of recency”). Eliminating or reversing the contingency should thus eliminate or reverse the direction of retrieval processes, effectively destroying the effect. Another somewhat speculative possibility might be to incentivize speed or accuracy in different parts of the experiment, but to keep the contingencies constant. Assuming that contingencies produce mostly gains in speed but mostly costs in accuracy, this should effectively reverse the reinforcement logic, but will keep the basic S-R contingencies intact.

The Question of Automaticity

A crucial question regards the implicitness or non-intentional (i.e., non-instrumental) nature of CL, since this is a precondition of considering it as an instance of habit formation. As we have explained in the introduction, habits reflect stimulus-driven operant behaviors that are characterized by features of automaticity. Research on habit acquisition often relies on using outcome devaluation as a crucial test for establishing the habitual character of a behavior (e.g., Moors et al., 2017; De Houwer et al., 2018). This criterion is of utmost importance when behaviors have previously been reinforced or are still followed by certain outcomes. Without establishing persistence and stability of the behavior in question independently of the rewards (i.e., after outcome devaluation), a core criterion of automaticity cannot be claimed, which is goal-independence. The resulting behavior may thus still have an instrumental character, which speaks against its purely habitual character. In our view, however, outcome devaluation is not a necessary criterion of habit acquisition. Only when behaviors are or have been linked to rewards can the criterion of outcome devaluation be directly applied. If habitual (i.e., automatic) operant behavior can be established via learning or experience without involving reinforcement (as we would argue is the case for the current study of CL without tangible rewards), then the test of outcome devaluation is not directly applicable (if there is no reward, then it cannot be devalued). Although tests of outcome (in-)dependence can be *added* to investigate the reward sensitivity (goal dependence) of a behavior, such a test cannot question the reward independence of the original behavior, which has been established in the absence

of rewards. Demonstrating an influence of reinforcement does not explain why habitual responding was found in the absence of rewards in the first place. This becomes immediately evident when considering outcome devaluation procedures, where outcome devaluation typically does have a strong effect on responding – the crucial aspect is that it does not eliminate behavior *completely*.

However, if the question of goal-independence and the test of outcome devaluation do not directly apply to our study, because contingencies were not rewarded in the first place, what is the basis on which we claim that CL results in a habit, that is, is automatic? CL has been shown to operate in the absence of awareness, which is a major criterion for automaticity (Schmidt et al., 2007). In the current studies, we used weak contingencies, which should be much harder to detect than the contingencies that were used in the study by Schmidt et al. (2007). In addition, we made the contingencies more complex, by making each word predictive of two instead of only one color, which should prevent participants from translating the contingencies into simple response strategies (cf. Schmidt and De Houwer, 2016). Finally, our study capitalized on yet another criterion of automaticity, which is speed. By introducing a response deadline of 1,000 ms, we exerted time pressure on participants during the task, which limits controlled processes during the task to a minimum, and has been shown to foster habitual responding (Hardwick et al., 2018; Luque et al., 2019). In sum, we thus feel justified in claiming that the CL effects that were obtained in our study reflect the operation of automatic processes, and thus can be characterized as being implicit. Of course, we have to acknowledge the limitation that we did not include any direct measures in our study that allowed us to conduct an empirical test for one or more criteria of automaticity within our experiments (Moors and De Houwer, 2006).

To sum up, we want to emphasize that our study is based on a broad conception of habits that categorizes operant behavior as habitual if it is stimulus-bound and shares some features of automaticity. This usage differs from a more narrow conception of habits that has been proposed by some researchers in the field (most notably, Dickinson, 1985), who argued that goal-independence is the core criterion of a habit, and that outcome devaluation is a necessary test to establish the habitual character of a behavior. It is important to interpret our findings against this background. Since we employed different criteria of habitual behavior, our core finding that habits can be explained in terms of episodic response retrieval may not generalize to habits that were established in terms of outcome independence. Further research is needed to clarify whether this functional explanation can be transferred also to behavior that has been shown to be goal-independent.

Stimulus Dependence, and Relevance of the Situational Cues

On a more general level, our results also bear some important implications for our understanding of habit formation. In particular, our findings highlight that situational cues play a crucial part in the acquisition and maintenance of habits, even when these situational cues are completely irrelevant for the

performed behavior. This is supported by the fact that word meaning was irrelevant for the color categorization task in the present study. However, participants' performance showed that they were sensitive to the co-occurrence of words and responses, and automatically retrieve the episodic instance in which the current word was presented most recently. Our findings reveal that effects of CL do not imply that participants were making strategic use of these regularities (cf. Schmidt et al., 2007; Giesen and Rothermund, 2015; see our arguments above). Apparently, all it takes to produce these effects is retrieval of the last occurrence of the word from episodic memory in order to simulate global CL effects (Schmidt et al., submitted).

Moderating Effects of Distance

Our study provides support for the law of recency by demonstrating that habits can emerge on the basis of retrieving just one single episode, which is one in which the person has responded to the current stimulus when it had been encountered during its last occurrence. In a situation where the last encounter has been fairly recent, this effect is strong enough to override all other previous occurrences of this situation that occurred before the last occurrence, rendering global contingencies irrelevant. However, as our data show, the last occurrence of a stimulus/situation quickly loses its influences on behavior with increasing distance to the current situation. Our findings revealed that after 5–6 intervening trials the influence of the last occurrence already vanishes. It remains unclear what happens if the last occurrence exceeds this distance: Instance-based retrieval might either break down completely for long intervals; alternatively, retrieval might still operate but might no longer be restricted to the very last occurrence (cf. Schmidt et al., 2016). According to the latter alternative, the last episode becomes less distinct with increasing distance and will more easily be confused with other instances. The predictions of these two alternatives are starkly different: According to the first variant, global contingencies will not influence behavior at all after controlling for the last occurrence, whereas the second account would predict that effects of mere frequency and/or global contingencies become visible when the last occurrences of the situation is distant. In this case, contingency effects would still reflect retrieval, but retrieval becomes less selective and will resemble more and more the probabilities and contingencies that are inherent in the entire set of previous episodes that share features with the current situation.

Relation Between the Laws of Recency, Exercise, and Effect

Our findings should not be taken to indicate that large frequencies of executing the same behavior over and over again ("law of exercise") have no influence on habitual behavior. For one thing, we did not test any influence of massive repetitions in our studies. We do not have any evidence on this, but it might well be that repeating a response for, say, more than 500 times might result in such a strong habit that inserting one counter-example might not suffice to overcome it. In fact, the influence of massive repetitions might be mediated by

a different pathway, and might operate independently from episodic retrieval process. On the other hand, instance-based retrieval processes and the law of recency might also play an important role for the explanation of overlearned behaviors. To test such an assumption, experiments should vary the similarity between the contexts in which the behavior was repeated and when it is tested. If exercise-based habits are shown to be context-dependent, then retrieval processes might also play a role in explaining these effects, but as we said, that remains to be investigated in future studies.

Finally, we also want to highlight that our findings do not rule out that instrumental behavior is influenced by rewards and incentives (“law of effect”). Demonstrating habitual behavior in the absence of reward just shows that reinforcement is not a necessary ingredient of habitual behavior (similar to what previous research has already shown with the outcome devaluation test). It could well be that reinforcement has a strong influence on responding also in the CL task, and it could also be that processes of episodic retrieval and CL are influenced by systematically rewarding or punishing certain combinations of stimuli and responses (but see Hauber, 2019). Demonstrating habitual behavior in the absence of rewards, however, attests to the fact that reinforcement is not a necessary ingredient of habits.

Practical Implications

The present findings also have important practical implications for the emergence and change of habitual responding. As shown in the present experiments, (irrelevant) situational cues play a major role in the acquisition and maintenance of habits. With regard to practical implications, this insight renders “exposure management” or “situation control” as another key variable of habit change. This claim is supported by research showing that a change of context reduces habitual responding in rats (Thrailkill and Bouton, 2015) and also in humans (e.g., Wood et al., 2005; Verplanken et al., 2008). Interestingly, gaining control over situational retrieval cues (e.g., creating a “seating habit” of sitting with one’s back to an all-you-can-eat buffet; Wansink and Payne, 2008) has the potential to become a new, desirable habit that counteracts undesirable habits (like unhealthy eating) in the future.

The core finding of our study is that the most *recent* stimulus-response bindings are crucial for the maintenance of habitual behavior, attesting to the law of recency. This reasoning is supported by the finding of the current study, as well as others (Schmidt et al., submitted), that response retrieval is much stronger for short distances, and that CL effects seem heavily influenced by more recent bindings. Put differently, in our study it was not the *frequency* of a pairing (reflecting global SR contingencies) but the *recency* of the episode that determines the direction of the habitual impulse. Our findings thus attest to the enormous importance of the very last occurrence of a certain situation in determining the response that is retrieved. Each word stimulus occurred hundreds of times during each experiment, and was paired with four different responses, two of which were highly frequent. Still, response retrieval was driven more or less completely

by the last occurrence of the word, and focusing on only the last occurrence was sufficient to fully explain CL, that is, habitual responding.

The strong effects of recency and in particular the behavior that was shown during the last occurrence of a certain situation offers important insights that can be inspiring for interventions targeted at creating desirable or breaking undesirable habits (for an overview, see Wood and R  nger, 2016): Executing a new behavior only once should already have a strong effect on subsequent behavior in this same situation. This strong effect is well-known for piano players who often have the (deplorable) experience that a specific error which occurred for the first time (and only once) at a certain point in a piece of music then has an extremely strong tendency to repeat at the next time, and to become chronic (see Marx, 1971; Marx et al., 1973; Marx and Marx, 1980; for a review, see Koppenaal, 1960).

On the other hand, this strong effect of a single episodic occurrence also offers a chance to change a bad habit into a good one by changing behavior only once. Breaking or overcoming existing habits typically requires effort and concentration (executive control). Our findings support the view, however, that spontaneous retrieval kicks in after only one occurrence and that execution of a response in a situation then impacts later behavior when the situation is encountered again.

At the same time, however, it would probably be na  ve to assume that a strong habit is already formed just by changing behavior once, and then trusting in retrieval of the last occurrence. Although we would assume that such a strategy may work remarkably well for the context in which the behavior is changed for the first time, it may not work anymore once the behavior has been interrupted by some other activity. It is not that episodic retrieval would not operate across large temporal distances (see previous section). Quite the contrary, the fact that habits are so robust already shows that time alone does not interfere with retrieval. What is different with increasing time is that the advantage of the last response episode – compared to the other episodes that were stored in memory before the last episode – is eliminated. The sharp decay function of episodic retrieval yields a clear advantage for the last episode across short time intervals; across longer intervals, however, the overall contingency should determine retrieval probabilities. That is, changing a habit once will typically be followed by immediate marked changes in behavior. To change it in the long run, however, will require repeated attempts in each new situation until the overall contingency has switched toward the new behavior.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

In accordance with ethical standards at the FSU Jena, the study was exempt from further ethical approval because no

cover-story or otherwise misleading or suggestive information was conveyed to participants.

AUTHOR CONTRIBUTIONS

CG co-developed the research idea, study, design, organized the data collection and analyses, and prepared the manuscript. JS involved in manuscript preparation. KR co-developed the

research idea, study, design, involved in data analyses and manuscript preparation.

ACKNOWLEDGMENTS

We thank Nils Meier for his help in programing the experiments and Christoph Hauber for his help in collecting the data.

REFERENCES

- Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Quar. J. Exp. Psychol. B* 34B, 77–98. doi: 10.1080/14640748208400878
- Amodio, D. M., and Ratner, K. G. (2011). A memory systems model of implicit social cognition. *Curr. Direct. Psychol. Sci.* 20, 143–148. doi: 10.1177/0963721411408562
- Bargh, J. A. (1994). “The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition,” in *Handbook of social cognition*, eds R. S. J. Wyer, and T. K. Srull, (Hillsdale, NJ: Erlbaum), 1–40.
- Carlson, K. A., and Flowers, J. H. (1996). Intentional versus unintentional use of contingencies between perceptual events. *Percept. Psychophys.* 58, 460–470. doi: 10.3758/BF03206821
- Cesario, J. (2014). Priming, replication, and the hardest science. *Perspect. Psychol. Sci.* 9, 40–48. doi: 10.1177/1745691613513470
- Cohen, J. (1969). *Statistical Power Analysis for the Behavioural Sciences*. New York, NY: Academic Press.
- Colwill, R. M. (1993). An associative analysis of instrumental learning. *Curr. Direct. Psychol. Sci.* 2, 111–116. doi: 10.1111/1467-8721.ep10772598
- Colzato, L. S., Raffone, A., and Hommel, B. (2006). What do we learn from binding features? evidence for multilevel feature integration. *J. Exp. Psychol.* 32, 705–716. doi: 10.1037/0096-1523.32.3.705
- De Houwer, J., Tanaka, A., Moors, A., and Tibboel, H. (2018). Kicking the habit: why evidence for habits in humans might be overestimated. *Motiv. Sci.* 4, 50–59. doi: 10.1037/mot0000065
- Dickinson, A. (1985). Actions and habits: the development of behavioral autonomy. *Philos. Trans. R. Soc. Lond.* 308, 67–78. doi: 10.1098/rstb.1985.0010
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/bf03193146
- Forrin, N. D., and MacLeod, C. M. (2018). Contingency proportion systematically influences contingency learning. *Attent. Percept. Psychophys.* 80, 155–165. doi: 10.3758/s13414-017-1424-4
- Frings, C., Rothermund, K., and Wentura, D. (2007). Distractor repetitions retrieve previous responses to targets. *Quar. J. Exp. Psychol.* 60, 1367–1377. doi: 10.1080/17470210600955645
- Fröber, K., Raith, L., and Dreisbach, G. (2018). The dynamic balance between cognitive flexibility and stability: the influence of local changes in reward expectation and global task context on voluntary switch rate. *Psychol. Res.* 82, 65–77. doi: 10.1007/s00426-017-0922-2
- Gawronski, B., and Payne, B. K. (eds) (2010). *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications*. New York, NY: Guilford.
- Giesen, C., Frings, C., and Rothermund, K. (2012). Differences in the strength of distractor inhibition do not affect distractor-response bindings. *Mem. Cogn.* 40, 373–387. doi: 10.3758/s13421-011-0157-1
- Giesen, C., and Rothermund, K. (2014). Distractor repetitions retrieve previous responses and previous targets. experimental dissociations of distractor-response and distractor-target bindings. *J. Exp. Psychol.* 40, 645–659. doi: 10.1037/a0035278
- Giesen, C., and Rothermund, K. (2015). Adapting to stimulus-response contingencies without noticing them. *J. Exp. Psychol.* 41, 1475–1481. doi: 10.1037/xhp0000122
- Giesen, C., and Rothermund, K. (2016). Multi-level response coding in stimulus-response bindings: Irrelevant distractors retrieve both semantic and motor response codes. *J. Exp. Psychol.* 42, 1643–1656. doi: 10.1037/xlm0000264
- Hardwick, R. M., Forrence, A. D., Krakauer, J. W., and Haith, A. M. (2018). Time-dependent competition between habitual and goal-directed response preparation. *BioRxiv* [Preprint] doi: 10.1101/201095v2.
- Hauber, C. (2019). *The Influence of Affective Consequences on Stimulus-Response Binding and Contingency Learning*. Master thesis, Jena: Friedrich-Schiller Universität Jena.
- Henson, R. N., Eckstein, D., Waszak, F., Frings, C., and Horner, A. J. (2014). Stimulus-response bindings in priming. *Trends Cogn. Sci.* 18, 376–384. doi: 10.1016/j.tics.2014.03.004
- Herwig, A., and Waszak, F. (2012). Action-effect bindings and ideomotor learning in intention- and stimulus-based actions. *Front. Psychol.* 3:1–18. doi: 10.3389/fpsyg.2012.00444
- Hommel, B. (1998). Event files: evidence for automatic integration of stimulus-response episodes. *Visual Cogn.* 5, 183–216. doi: 10.1080/713756773
- Hommel, B., Müseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–937. doi: 10.1017/S0140525X01000103
- Hull, C. L. (1943). *Principles of Behavior: an Introduction to Behavior Theory*. Oxford: Appleton-Century.
- Klauser, K. C., Voss, A., and Stahl, C. (eds) (2012). *Cognitive Methods in Social Psychology*. New York, NY: Guilford.
- Koppelaar, R. J. (1960). Repetition of errors in human multiple-choice learning. *Psychol. Rep.* 7, 269–286. doi: 10.2466/PRO.7.6.269-286
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychol. Rev.* 95, 492–527. doi: 10.1037/0033-295X.95.4.492
- Logan, G. D., and Zbrodoff, N. J. (1979). When it helps to be misled: facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Mem. Cogn.* 7, 166–174. doi: 10.3758/bf03197535
- Luque, D., Molinero, S., Watson, P., López, F. J., and Le Pelley, M. (2019). Measuring habit formation through goal-directed response switching. *PsyArXiv* [Preprint] doi: 10.31234/osf.io/xnmfc
- MacLellan, E., Shore, D. I., and Milliken, B. (2015). Contextual control over selective attention: evidence from a two-target method. *Psychol. Res.* 79, 556–569. doi: 10.1007/s00426-014-0593-1
- MacLeod, C. M. (2019). Learning simple associations. *Can. Psychol.* 60, 3–13. doi: 10.1037/cap0000170
- Marx, M. H. (1971). Increased probability of error repetition as a function of number of successive prior repetitions. *Percept. Motor Skills* 32, 544–546. doi: 10.2466/pms.1971.32.2.544
- Marx, M. H., and Marx, K. (1980). Repetition of errors in learning and memory as a function of their prior associative strength. *Bull. Psychon. Soc.* 16, 435–438. doi: 10.3758/BF03329591
- Marx, M. H., Witter, D. W., and Farbray, J. (1973). Greater repetition of errors under performance compared to observation in multiple-choice human learning. *Percept. Motor Skills* 37, 949–950. doi: 10.2466/pms.1973.37.3.949
- Mayr, S., and Buchner, A. (2006). Evidence for episodic retrieval of inadequate prime responses in auditory negative priming. *J. Exp. Psychol.* 32, 932–943. doi: 10.1037/0096-1523.32.4.932
- Meehl, P. E. (1950). On the circularity of the law of effect. *Psychol. Bull.* 47, 52–75. doi: 10.1037/h0058557
- Meier, M. E., and Kane, M. J. (2013). Working memory capacity and Stroop interference: global versus local indices of executive control. *J. Exp. Psychol.* 39, 748–759. doi: 10.1037/a0029200

- Miller, J. (1987). Priming is not necessary for selective-attention failures: Semantic effects of unattended, unprimed letters. *Percept. Psychophys.* 41, 419–434. doi: 10.3758/BF03203035
- Miller, K. J., Shenhav, A., and Ludvig, E. A. (2019). Habits without values. *Psychol. Rev.* 126, 292–311. doi: 10.1037/rev0000120
- Moeller, B., and Frings, C. (2014). Long term response-stimulus associations can influence distractor-response bindings. *Adv. Cogn. Psychol.* 10, 68–80. doi: 10.5709/acp-0158-1
- Moeller, B., and Frings, C. (2017). Dissociation of binding and learning processes. *Atten. Percept. Psychophys.* 79, 2590–2605. doi: 10.3758/s13414-017-1393-7
- Moors, A., Boddez, Y., and De Houwer, J. (2017). The power of goal-directed processes in the causation of emotional and other actions. *Emot. Rev.* 9, 310–318. doi: 10.1177/1754073916669595
- Moors, A., and De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychol. Bull.* 132, 297–326. doi: 10.1037/0033-2909.132.2.297
- Nosek, B. A., Spies, J. R., and Motyl, M. (2012). Scientific utopia: II restructuring incentives and practices to promote truth over publishability. *Perspect. Psychol. Sci.* 7, 615–631. doi: 10.1177/1745691612459058
- Ostlund, S. B., and Balleine, B. W. (2007). Selective reinstatement of instrumental performance depends on the discriminative stimulus properties of the mediating outcome. *Learn. Behav.* 35, 43–52. doi: 10.3758/bf03196073
- Premack, D. (1965). “Reinforcement theory,” in *Nebraska Symposium on Motivation*, ed. D. Levine, (Lincoln, NE: University of Nebraska Press), 123–180.
- Rescorla, R. A. (1991). Associative relations in instrumental learning: the eighteenth Bartlett memorial lecture. *Quar. J. Exp. Psychol.* 43B, 1–23.
- Rothermund, K., Wentura, D., and De Houwer, J. (2005). Retrieval of incidental stimulus-response associations as a source of negative priming. *J. Exp. Psychol.* 31, 482–495. doi: 10.1037/0278-7393.31.3.482
- Schmidt, J. R., Crump, M. J. C., Cheesman, J., and Besner, D. (2007). Contingency learning without awareness: evidence for implicit control. *Conscious. Cogn.* 16, 421–435. doi: 10.1016/j.concog.2006.06.010
- Schmidt, J. R., and De Houwer, J. (2016). Contingency learning tracks with stimulus-response proportion: No evidence of misprediction costs. *Exp. Psychol.* 63, 79–88. doi: 10.1027/1618-3169/a000313
- Schmidt, J. R., De Houwer, J., and Rothermund, K. (2016). The parallel episodic processing (PEP) model 2.0: a single computational model of stimulus-response binding, contingency learning, power curves, and mixing costs. *Cogn. Psychol.* 91, 82–108. doi: 10.1016/j.cogpsych.2016.10.004
- Schmidt, J. R., Giesen, C., and Rothermund, K. (2019). Contingency learning as binding? testing an exemplar view of the colour-word contingency learning effect. *Quar. J. Exp. Psychol.*
- Thorndike, E. L. (1898). Animal intelligence: an experimental study of the associative processes in animals. *Psychol. Rev. Monogr. Suppl.* 2, i–109. doi: 10.1523/JNEUROSCI.1615-08.2008
- Thrallkill, E. A., and Bouton, M. E. (2015). Contextual control of instrumental actions and habits. *J. Exp. Psychol.* 41, 69–80. doi: 10.1037/xan0000045
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Tversky, A., and Kahneman, D. (1971). Belief in the law of small numbers. *Psychol. Bull.* 76, 105–110. doi: 10.1037/h0031322
- Verplanken, B., Walker, I., Davis, A., and Jurasek, M. (2008). Context change and travel mode choice: combining the habit discontinuity and self-activation hypotheses. *J. Environ. Psychol.* 28, 121–127. doi: 10.1016/j.jenvp.2007.10.005
- Wansink, B., and Payne, C. R. (2008). Eating behavior and obesity at Chinese buffets. *Obesity* 16, 1957–1960. doi: 10.1038/oby.2008.286
- Wentura, D., and Rothermund, K. (2007). “Paradigms we live by. A plea for more basic research on the IAT,” in *Implicit Measures of Attitudes*, eds B. Wittenbrink, and N. Schwarz, (New York, NY: Guilford Press), 195–215.
- Wittenbrink, B., and Schwarz, N. (eds) (2007). *Implicit Measures of Attitudes*. New York, NY: Guilford Press.
- Wood, W. (2017). Habit in personality and social psychology. *Pers. Soc. Psychol. Rev.* 21, 389–403. doi: 10.1177/1088868317720362
- Wood, W., and Rünger, D. (2016). Psychology of habit. *Annu. Rev. Psychol.* 67, 289–314. doi: 10.1146/annurev-psych-122414-033417
- Wood, W., Tam, L., and Witt, M. G. (2005). Changing circumstances, disrupting habits. *J. Pers. Soc. Psychol.* 88, 918–933. doi: 10.1037/0022-3514.88.6.918

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Giesen, Schmidt and Rothermund. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Ideomotor Action: Evidence for Automaticity in Learning, but Not Execution

Dan Sun*, Ruud Custers*, Hans Marien and Henk Aarts

Department of Psychology, Utrecht University, Utrecht, Netherlands

OPEN ACCESS

Edited by:

Wendy Wood,
University of Southern California,
United States

Reviewed by:

Poppy Watson,
University of New South Wales,
Australia
Barbara Jean Knowlton,
University of California, Los Angeles,
United States

*Correspondence:

Dan Sun
d.sun@uu.nl
Ruud Custers
r.custers@uu.nl

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 16 April 2019

Accepted: 27 January 2020

Published: 14 February 2020

Citation:

Sun D, Custers R, Marien H and
Aarts H (2020) Ideomotor Action:
Evidence for Automaticity in Learning,
but Not Execution.
Front. Psychol. 11:185.
doi: 10.3389/fpsyg.2020.00185

Human habits are widely assumed to result from stimulus-response (S-R) associations that are formed if one frequently and consistently does the same thing in the same situation. According to Ideomotor Theory, a distinct but similar process could lead to response-outcome (R-O) associations if responses frequently and consistently produce the same outcomes. This process is assumed to occur spontaneously, and because these associations can operate in a bidirectional manner, merely perceiving or thinking of an outcome should automatically activate the associated action. In the current paper we test this automaticity feature of ideomotor learning. In four experiments, participants completed the same learning phase in which they could acquire associations, and were either explicitly informed about the contingency between actions and outcomes, or not. Automatic action selection and initiation were investigated using a free-choice task in Experiment 1 and forced-choice tasks in Experiment 2, 3a, and 3b. An ideomotor effect was only obtained in the free-choice, but not convincingly in the forced-choice tasks. Together, this suggests that action-outcome relations can be learned spontaneously, but that there may be limits to the automaticity of the ideomotor effect.

Keywords: action control, automaticity, goal-directed behavior, ideomotor, implicit learning

INTRODUCTION

Habits are often regarded to be the result of stimulus-response (S-R) associations that are assumed to be formed if people repeatedly and consistently perform the same behavior in the same situation, often because there is an incentive to do so (Wood and Rünger, 2016). As a consequence, the situation may trigger the associated response in an automatic fashion, leading to habitual behavior that is no longer guided by deliberative processes (Aarts and Dijksterhuis, 2000), but controlled by the environment. A relevant but distinct line of research proposes a similar mechanism in which behaviors can become associated with the situations or events that *follow* actions: Ideomotor theory proposes that if a behavioral response is repeatedly and consistently followed by the same perceptual outcome, thinking about or activating the mental representation of that outcome can to a certain extent prepare or trigger the behavior through bi-directional response-outcome (R-O) associations. This mechanism of ideomotor action has been used to explain various instances in which the environment triggers behaviors in an automatic fashion, such as mimicry, or behavior from affordances (Iacoboni, 2008; Custers and Aarts, 2010).

Ideomotor-action could be relevant to the understanding of habitual behavior in at least two ways. First, it may help to understand how the environment could trigger behaviors that look like habits, but may not be the result of classic habit formation processes (i.e., not resulting from

S-R associations). Second, it may help to understand the implementation of seemingly abstract S-R associations. That is, many behaviors that are regarded as habits (reading the newspaper on Saturday morning, having coffee after dinner, reading a book before going to sleep) are not directly represented at the motor level, but representations include a rich collection of experiences of the consequences of executing the behavior and allow for an abstract representation of the behavior. Research indeed suggests that people represent behaviors in a hierarchical way, in which more abstract representations of the behavior are often the outcomes of the lower-level actions that produced them (Vallacher and Wegner, 1987; Kruglanski et al., 2002; Cooper and Shallice, 2006). Representing behaviors in terms of their outcomes may therefore help to produce the same behavioral outcome (e.g., reading the newspaper) under slightly different conditions (e.g., picking up the paper from a slightly different location on the doormat each time and finding an empty chair to read it; Powers, 1973; Custers and Aarts, 2010).

Although action-outcome representations may be indispensable for human behaviors, and especially goal-directed actions, it is less clear how these associations are acquired. Moreover, although contemporary approaches to ideomotor action (Hommel, 2013) assume that bi-directional R-O associations could trigger responses in an automatic fashion, there are few rigorous tests that demonstrate this. In the present paper we put the automaticity in the formation and execution of ideomotor action within the classic ideomotor paradigm to the test. We first review current evidence for the automatic nature of ideomotor action and evidence for spontaneous ideomotor learning. We then investigate whether or not learning relations between actions and outcomes can occur spontaneously, by merely executing actions and observing following events, and without specific instructions. Three different ideomotor tests are used to gain insight in the degree to which potentially resulting ideomotor actions are automatic.

Ideomotor Theory

The notion of ideomotor action dates back to the 19th century (Carpenter, 1852; Lotze, 1852; James, 1890), aiming to explain how thought can trigger action (for reviews see, Stock and Stock, 2004; Shin et al., 2010). The central idea of early ideomotor theory was that merely envisioning an action triggers that action to a certain extent (James, 1890), even in the absence of a conscious intention to act (Ansfield and Wegner, 1996). Embracing the idea that thinking of an action includes envisioning its anticipated outcomes, Greenwald (1970) proposed that ideomotor action relies on bi-directional R-O associations. That is, thinking about an actions involves thinking about the perceptual experiences that have become associated with particular motor programs (see also., Zwaan and Taylor, 2006). While such associations enable response selection based on outcomes of actions (i.e., goal-directed behavior), the strong version of ideomotor theory (see Shin et al., 2010) holds that once the association is formed, thinking (ideation) of an outcome, or merely perceiving a related stimulus, is enough to trigger the associated action. This backward activation appears to be a robust and general phenomenon which has been observed for many different

action and stimuli, such as auditory stimuli (e.g., Elsner and Hommel, 2001), faces (Herwig and Horstmann, 2011), locations (Hommel, 1993), and letters (Ziessler and Nattkemper, 2002; Hommel et al., 2003).

In the last two decades, the Theory of Event Coding (TEC) (Hommel et al., 2001) has revived interest in ideomotor action, by providing a cognitive-perceptual framework for understanding these effects. This framework holds that both actions and their perceived sensory effects are cognitively represented in a similar distributed fashion and that their feature codes become intricately linked in action-stimulus representations that contain information about both. As these representations can be used bi-directionally, observing or thinking of an outcome activates the representation of the corresponding action, explaining phenomena such as mimicry (Iacoboni, 2008) action priming (Dijksterhuis and Bargh, 2001), and goal priming (Custers and Aarts, 2010). According to TEC, representations of effects and basic motor movement already become intertwined in early infancy (Hommel et al., 2001; Heyes, 2010). It appears, then, that R-O associations emerge spontaneously as a result of acting and observing, giving rise to representations that can drive behavior in an automatic, habit-like fashion.

Ideomotor Research

Following Greenwald (1970), tests of ideomotor learning typically contain two-phases: An acquisition phase in which action-outcome associations are acquired, and a test phase that tests whether these stimuli (i.e., outcomes) facilitate associated actions. In a classic study, Elsner and Hommel (2001) had participants freely choose in the first phase (i.e., free-choice acquisition phase) between left and right key presses that were each consistently followed by a specific tone (high or low pitch). Importantly, participants were explicitly informed that the tones were irrelevant to the task. In the second phase (forced-choice test phase, Experiments 1a, 1b), participants had to press left or right keys preceded by the tones that mapped on the earlier learned responses (non-reversal group), whereas for the other group the Response-Outcome mapping was reversed (reversal group). Results showed that actions were performed faster when the mapping was consistent with that in the acquisition phase, rather than reversed. Follow-up experiments (Experiments 2–4) revealed a similar consistency effect in a free-choice test phase that required subjects to press left and right keys randomly: Actions that were consistent with the Response-Outcome mapping were more frequently selected after the tones, showing a response bias in free choice as a result of outcome priming.

Later studies have systematically compared the effects of free- and forced-choice learning phases. Herwig et al. (2007) used a forced choice test-phase in which participants were allocated to a non-reversal or reversal group. They found that effects of ideomotor learning between actions and resulting outcomes only occurred when participants voluntarily selected actions in the learning phase (free-choice learning), and not when the required responses were forced by cues (forced-choice learning). These findings suggest that participants more readily represented the stimuli (tones) as outcomes of their actions when they engaged

in free-choice learning, whereas merely responding to cues did not produce such a psychological process. Hence, even though actions were followed by stimuli in exactly the same way in free- and forced-choice learning phases, the stimulus information appears to have been encoded differently during learning.

Subsequent work by Pfister et al. (2011) suggested that it may not be the encoding in the acquisition phase, though, that makes the difference, but rather the mode in which people control their behaviors in the test phase. Using a free-choice test phase, they found evidence for ideomotor effects, regardless of whether learning took place in a free- or a forced choice phase. They concluded that ideomotor learning takes place whenever actions are followed by events, regardless of the acquisition task, but that participants need to be engaged in “intention-based control” in the test phase (that is, selecting outcome-related actions), for ideomotor effects to arise. This would suggest that while learning of habitual action-outcome relations may be spontaneous, it may be conditional on a certain mind set or task set (i.e., conditional automaticity; see Aarts and Dijksterhuis, 2000).

Instruction Effects

Although the research discussed above suggests that ideomotor learning occurs spontaneously whenever events follow actions, this “spontaneous learning” always occurs within the experimental setting. As it happens, though, task instructions in the acquisition phase often explicitly mention the presence of outcomes in the task, stating that they are irrelevant and should be ignored (e.g., Elsner and Hommel, 2001). Whilst it is not always clear which exact instructions are provided in the acquisition phase in ideomotor research, Eder and Dignath (2017) have recently demonstrated in a task in which learning and testing of ideomotor action are intertwined, that such task instructions matter a lot. Based on recent insights in the power of instruction effects (see Liefvoeghe et al., 2018), Eder and Dignath provided instructions to ignore, attend, learn, or intentionally produce action outcomes in one combined learning/test phase. Results showed that instructions affect the task set with which action-stimulus relations are learned (Custers and Aarts, 2011), but that unlike the learning and intention instructions, instructions to ignore or attend to outcomes did not lead to ideomotor learning, at least not in this experimental setting.

In the present paper, we investigate whether ideomotor learning occurs spontaneously in the standard two-phase paradigm with auditory stimuli. In four studies, we manipulated instructions in a free-choice learning phase, either saying nothing at all about tones that followed actions, or emphasizing their relationship in terms of actions and outcomes. All experiments used a free-choice acquisition phase, as previous research suggests that action-outcome relations are more strongly acquired and subsequently used (Herwig et al., 2007; Pfister et al., 2011). Given the complexity of obtaining clear and reliable ideomotor effects, and in order to gain more insight in what is learned in the acquisition phase, we employed three different ideomotor tests in four separate experiments. In Experiment 1, we used a free-choice test phase, as earlier work has suggested that ideomotor effects are most likely to occur under such conditions (e.g., Pfister et al., 2011).

However, as the free-choice ideomotor test is - by definition - open to influences of conscious deliberation and choice, we follow up in Experiment 2, 3a, and 3b with a forced-choice ideomotor test. While Experiment 2 used a 2-block design where participants received opposite instructions on the different blocks that forced them to react to outcome stimuli either in line with the acquired action-outcome mapping, or the opposite mapping, Experiment 3a and 3b used an interference paradigm with imperative cues (presented together with outcome stimuli) to force people's choice on trial level. These forced-choice ideomotor tests would provide stronger evidence for the automatic initiation of actions than the free-choice test, with Experiment 3a and 3b being the least susceptible to alternative explanations. As such, the current line of experiments not only tests, but also aims to verify the automatic nature, of potential ideomotor actions arising from spontaneous ideomotor learning.

EXPERIMENT 1: FREE-CHOICE IDEOMOTOR TEST

Method

Participants and Design

Sample sizes on previously published ideomotor learning studies which varied from 12 (e.g., Kühn et al., 2009, Experiment 1) to 20 participants per condition (e.g., Herwig and Waszak, 2009, Experiment 1–3), and given the fact that small sample sizes can counterintuitively inflate effect size, we decided prior to data collection to test at least 20 participants per condition in each experiment.

Fifty participants took part in the experiment in exchange for a small monetary payment or extra course credits. Participants with attention-related disorders or those who were on related medication were excluded beforehand. The experimental design consisted of one between-subjects factor: Instructions (No-Instructions vs. Instructions). After signing the informed consent, participants were randomly assigned to either the No-Instructions condition or the Instructions condition.

Data of one participant were lost because of a technical issue, and two participants were excluded due to the unbalanced proportion of key presses during the acquisition phase (outside of the range of a left-to-right ratio of 40 to 60%), which was defined before data collection. Data of the remaining 47 participants (No-Instructions condition: $n = 23$, Instructions condition: $n = 24$) were included in the analyses [37 females, mean age: 24 years (18–30 years), no left-handed and 2 ambidextrous participants].

Procedure

Participants were told that they would perform two tasks on a computer and were asked to read the instructions carefully. The present study used the same design as the third experiment of Elsner and Hommel (2001), consisting of an acquisition phase and a test phase. Both phases featured a Go – No-Go paradigm, and the auditory stimuli following responses in the acquisition phase [i.e., a low tone (400 Hz) and a high tone (800 Hz)] were presented again in the test phase upon which participants were

to freely choose a left or a right response. After the acquisition phase, they continued with the second task (i.e., the test phase).

After the two main phases, participants filled out a short questionnaire that tested their knowledge about Response-Outcome mappings acquired in the learning phase and measured the representation levels on four hierarchically different levels of self-causation (i.e., association, prediction, causality, and agency level of Response-Outcome relations, see below) to check whether the instructions induced the desired processing goals differently. Response-Outcome mappings were counterbalanced among the participants. That is, for half of the participants, the left key was followed by the high tone and the right key by the low tone (Response-Outcome mapping A), whereas for the other half, the opposite key-tone mappings (Response-Outcome mapping B) were used.

Acquisition Phase

After general task instructions, all participants read the following specific instruction for the acquisition phase:

"In this part you have to press a key with your left or right index finger, depending on the instructions on the screen: If you see "<<>>", you can choose yourself to press the left key ("z"), or the right key ("l"). You can choose freely, but try, on average, to press left and right equally often. If you see "xxxx," however, you should not press any key."

Participants in the Instructions condition were then given detailed additional information about the R-O mappings – which depended on the counterbalancing of the mapping – and were provided with processing goals through descriptions of the relationship between the responses and their outcomes in ascending levels of self-causation (i.e., from associative, predictive, to causal) in the acquisition phase:

"Pressing your left key is associated with a High [Low] tone and pressing your right key with a Low [High] tone. This means that upon pressing your left key you can predict a High [Low] tone and upon pressing your right key a Low [High] tone. In other words: pressing your left key causes a High [Low] tone and pressing your right key causes a Low [High] tone."

It is important to keep in mind that in the No-Instructions condition the tones are just stimuli that consistently followed key-presses, without any related mention about the occurrence of the tones, and that in the Instructions condition the stage was set for processing the tones as outcomes of self-chosen actions.

The trial procedures of the acquisition phase are depicted in **Figure 1A**. Each trial of the acquisition phase started with a fixation asterisk (*) for 500 ms on the middle of the screen, followed by a 200-ms Go (i.e., "<<>>") or No-Go (i.e., "xxxx") signal. Participants were asked to press the left or right key freely as soon as they saw the Go signal and were asked not to respond in No-Go trials. The program waited up to 1,000 ms for a response. On Go trials, reaction times over 1,000 ms were treated as omissions and responses faster than 100 ms as anticipations. Only reaction times in the valid range (100–1000 ms) triggered the contingent tone, which started after a 50-ms lag from the onset of the keypress and was presented for 200 ms. Incorrect trials (i.e., omissions, anticipations, and responses to No-Go trials)

were recorded, and were signaled to the participant by a 1000-ms warning messages on the screen saying: "too slow", "too fast," or "No-Go trial, respectively. All incorrect trials were repeated in random order by the end of the first task. Participants had to redo all the incorrect trials until all required responses were valid.

The acquisition phase consisted of three practice trials and 300 valid trials, divided into 10 blocks. Every two blocks, there was a 10 s break, during which participants were informed about how often they had pressed the left and right keys. In the Instructions condition, the extra processing information about the Response-Outcome mappings was also repeated (e.g., "Each specific key causes a specific tone. The left key causes a High tone and the right key causes a Low tone").

Test Phase

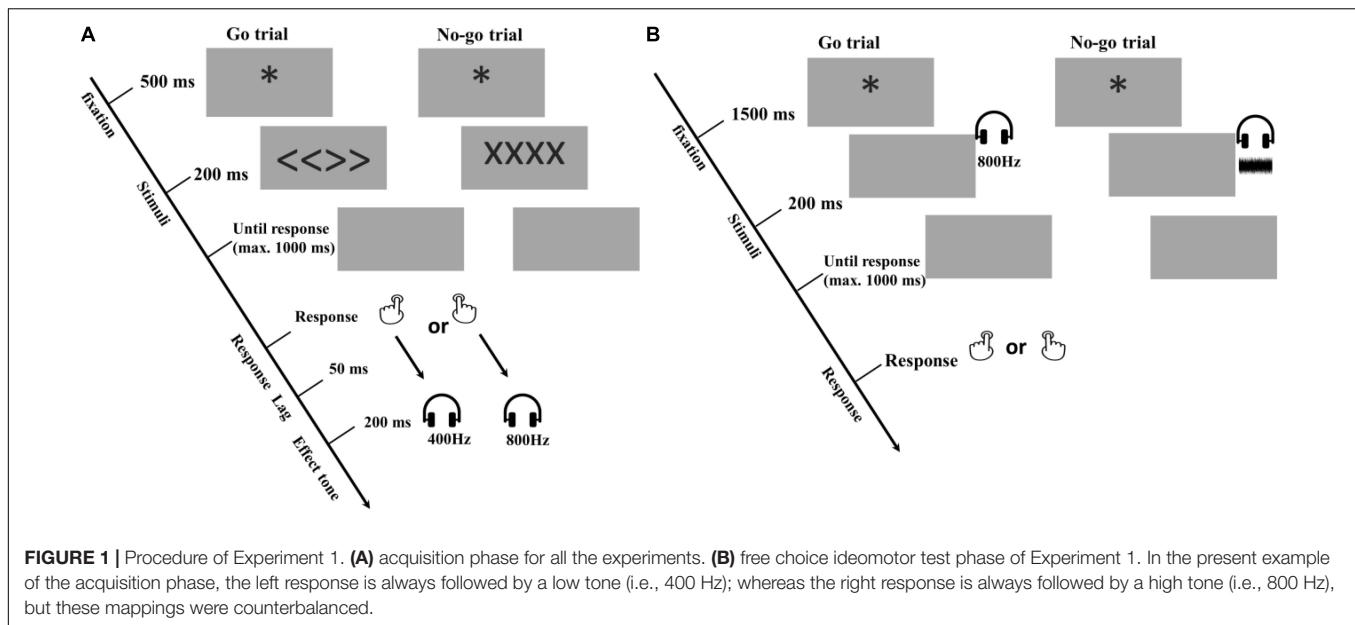
The test phase was similar to the acquisition phase, also using the Go – No-Go paradigm. This time, however, two tones that previously served as outcomes were presented as cueing stimuli (see **Figure 1B**). Participants were instructed to press the left or right key randomly in response to the tone. In addition, as suggested by Elsner and Hommel (2001), to add response uncertainty and prevent participants from responding before the tone appeared, a novel sound (i.e., a 200-ms white noise signal) was presented in one third of the test trials, serving as a No-Go signal after which participants were to withhold their response. Each test trial started after an inter-trial interval of 1,500 ms with an asterisk on the center of the screen, followed by a 200-ms sound (i.e., a high tone, a low tone or a white noise signal), which were presented in a random order. Then the program waited up to 1,000 ms for an appropriate response. Response omissions and anticipations were defined in the same way as in the acquisition phase. However, this time no error message was presented and participants worked through six practice trials and 288 valid trials, divided into 8 blocks, including 96 No-Go trials in total. Again, every two blocks, there was a 10 s break. This time no extra information about the Response - Outcome mappings was provided during the break.

Manipulation Check of R-O Mappings

After the test phase, participants answered two questions that tested their knowledge about the relationship between the responses (i.e., left/right key presses) and the corresponding outcomes (i.e., low/high tones) in the acquisition phase, to check whether participants were able to report which tone followed which response. There were four answer options to each mapping question. For instance, when asked: "Which tone did the left key press produce?", response options were: (1) the left key press produced the High tone, (2) the left key press produced the Low tone, (3) the left key press produced both tones, (4) the left key press was irrelevant to the tones" (see **Supplementary Appendix S1** for more details).

Manipulation Check of Instructions

Subsequently, participants filled out a questionnaire designed to measure changes in the representation of the response-outcome relations as a result of the instructions manipulation. The questionnaire probed the four levels of the hierarchical



representation used in the Instructions condition (i.e., association, prediction, causality, and agency level of Response-Outcome relations). Specifically, for each level, three items probed representations using a 9-point scale. The complete questionnaire can be found in the Appendix (see **Supplementary Material**). A difference between instruction conditions on these measures would indicate that the manipulation changed the way in which participants represented the response-outcome relations.

Data Analysis Plan

Data were analyzed using R 3.5 (R Core Development Team, 2014). Visualizations of raw data points were built with the raincloud plots (Allen et al., 2018). ANOVA's were calculated using the aov_ez function and Type III sums of squares (afex package Version 0.22-1 in R) (Singmann et al., 2016). When assumptions of sphericity were violated Greenhouse-Geisser (GG) correction was utilized in the ANOVA model. In this case, we reported uncorrected degrees of freedom and corrected *p*-values. To further draw conclusions about the support of null effects, we also calculated Bayesian factors (BFs) with the default prior setting in JASP (version 0.9, JASP Team 2018) (van Doorn et al., 2019). The advantage of BFs is that it quantifies evidence in favor of one (e.g., null) hypothesis compared to another (e.g., alternative) hypothesis given the observed data.

Results

Acquisition Phase

First, we excluded all acquisition trials with anticipations (No-Instructions: 0.01%, Instructions: 0.01%) and omissions (No-Instructions: 0.04%, Instructions: 0.09%). Failures to withhold responses on the No-Go trials were calculated and all participants fell below the pre-set criteria of less than 20% (No-Instructions: 2.89%, Instructions: 2.55%). After that, response proportions (left/right keypress) were calculated. To make sure

the participants had followed the general instruction to press the left and right key randomly but equally often, participants with proportions outside the 40% to 60% range were excluded (see section "Participants and Design"). The mean left/right response proportions were equal in each condition – No-Instructions condition: 49.9% vs. 50.1%; Instructions condition: 49.6% vs. 50.4%.

The mean RTs of the participants did not differ between the No-Instructions, $M = 362.94$ ms, $SD = 60.24$ ms, and Instructions condition, $M = 362.41$ ms, $SD = 39.01$, $F(1,45) = 0.00$, $p = 0.97$. The mean RTs of right responses, $M = 360.75$ ms, $SD = 52.40$ ms, were marginally faster than the mean RTs of left responses $M = 364.59$ ms, $SD = 48.49$ ms, $F(1,45) = 2.87$, $p = 0.10$. This difference was not qualified by an interaction with the between-subjects factor Instructions, $F(1, 45) = 0.75$, $p = 0.39$.

Test Phase

Test trials with response anticipations (No-Instructions: 0.05%, Instructions: 0.06%) and omissions (No-Instructions: 1.31%, Instructions: 0.91%) were excluded from data analysis and the percentage of responses that were consistent with the previously acquired Response-Outcome mapping was calculated for each participant.

As expected, in the No-Instructions condition the mean proportion of consistent responses was significantly larger than chance (i.e., 50%), $M = 61.49\%$, $SD = 22.61\%$, $t(22) = 2.44$, $p = 0.012$ (one-tailed), Cohen's $d_z^1 = 0.508$, and the Bayesian one sample *T*-Test resulted in $BF_{+0} = 4.80$, which means that the data are approximately 4.8 times more likely to occur under H_+ (i.e., proportion in consistent condition is higher than chance level, that is, larger than 50%), than under H_0 (i.e., proportion in consistent condition is at chance level). This

¹Cohen's d_z is the standardized mean difference effect size, for a detailed calculation see Lakens (2013).

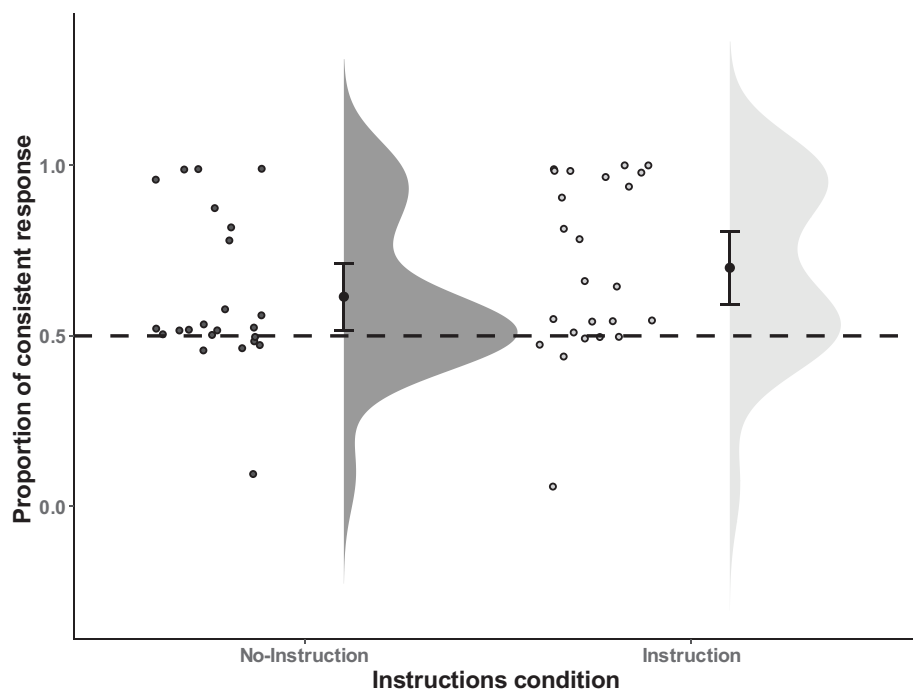


FIGURE 2 | Distribution of the proportion of consistent responses of all the individual data points in Experiment 1. Error bars represent the 95% confidence intervals.

result indicates moderate evidence in favor of H_+ . The same effect was observed for the Instructions condition: $M = 69.98\%$, $SD = 25.42\%$, $t(23) = 3.85$, $p = 0.0004$ (one-tailed), Cohen's $d_z = 0.786$, and the Bayesian one sample T -Test result is $BF_{+0} = 83.90$, which indicates strong evidence in favor of H_+ . Finally, we tested whether instructions affected the proportion of consistent responses, but the direct comparison between the two conditions did not reveal any significant difference, $t(45) = -1.21$, $p = 0.23$ (two-tailed), and the Bayesian Independent samples T -Test result equals (BF_{01}) 1.91, which only slightly favors the null hypothesis (H_0 : The Instructions condition has no effect on response preference) over the alternative hypothesis (H_1 : the Instructions condition biases response selection). In sum, while there was very strong support for an ideomotor effect in the Instructions condition and substantial evidence in the No-Instructions condition, evidence for no difference between Instructions conditions was only anecdotal (see **Figure 2** for distribution).

Furthermore, we compared RTs for consistent and inconsistent trials in both Instructions conditions. There was no difference between consistent ($M = 502.42$ ms, $SD = 95.89$ ms) and inconsistent trials ($M = 503.64$ ms, $SD = 99.22$ ms) in the No-Instructions condition, $t(22) = -0.14$, $p = 0.894$; nor between consistent ($M = 510.68$ ms, $SD = 70.11$ ms) and inconsistent trials ($M = 506.64$ ms, $SD = 69.98$ ms) in the Instructions condition, $t(21) = 0.77$, $p = 0.453$. The corresponding BF also indicates moderate evidence for the null hypothesis (H_0 : The reaction times are not different between consistent and inconsistent trials) over the alternative hypothesis (H_1 : The reaction times are different between consistent and inconsistent trials) in

No-Instruction condition ($BF_{01} = 4.535$) and Instruction condition ($BF_{01} = 3.448$), respectively.

Manipulation Check of R-O Mappings

Most participants (85% in total) were able to explicitly report the correct mapping of responses and subsequent stimuli they were exposed to in the acquisition phase. Six people missed the response-stimulus mapping in the No-Instructions condition (6 out of 23), and only one participant failed in the Instructions condition (1 out of 24).

Manipulation Check of Instructions

In order to assess whether there were differences in how people represented the relation between responses and outcomes in the acquisition phase, the average of each of the three questions measuring association, prediction, causality, and agency was calculated. The mean scale ratings were analyzed as a function of Instructions conditions and as a function of representation level (i.e., the hierarchical levels explained before). Only a main effect of representation level was found, $F(3,135) = 7.97$, $p_{[GG]} < 0.001$, $\eta_p^2 = 0.15$, which merely showed that collapsed over Instructions conditions, there were significant differences in ratings between the four level of representation constructs (**Table 1** and **Supplementary Table S1.1** presents more details of the responses to the scales).

Discussion

These results provide support for an ideomotor effect, in the sense that tones followed responses in the acquisition phase

TABLE 1 | Means and Standard deviations of the four different representation levels collapsed over Instructions conditions for all three experiments.

Representation level	Exp1: free	Exp2: block-based	Exp3: trial-based
Association	7.45 ± 2.18	6.90 ± 2.63	7.55 ± 2.20
Prediction	6.67 ± 2.86	6.99 ± 2.73	7.32 ± 2.53
Causality	6.99 ± 2.78	7.03 ± 2.75	7.60 ± 2.23
Agency	6.26 ± 2.92	6.16 ± 3.10	6.83 ± 2.75

The means of each scale is relatively high, indicating that in both the Instructions and No-instructions condition participants processed the learning task in line with the Instructions manipulation.

were more likely to evoke these responses in the test phase. Moreover, this effect occurred regardless of instructions about the relation between responses and tones, which demonstrates that ideomotor learning – at least in the current paradigm – unfolded spontaneously.

Although the ideomotor effect was observed within both instruction conditions, it appeared more pronounced in the instructions condition. Bayesian tests, however, revealed slightly more support for the absence of difference between the two conditions. While it cannot be ruled out that instructions can strengthen ideomotor learning, it is clear that instructions were not necessary for learning to occur in the acquisition phase. This finding is further corroborated by an absence of a difference in the representation-level checks.

While the observed ideomotor effect obtained in the test phase seems comparable in size with other ideomotor studies (c.f., Elsner and Hommel, 2001), the free choice test phase does not provide strong evidence for the automatic nature of the effect (i.e., that the responses are triggered automatically by the stimuli that served as outcomes in the acquisition phase), as this task allows for deliberate responses in the test phase as well. On closer inspection, the response data show a bimodal distribution, with the majority of people responding at chance level and a considerable amount of people demonstrating a very large bias, with some participants showing near perfect consistence with the mapping acquired in the acquisition phase. This could suggest that the observed effect was not so much produced by the tones triggering the corresponding actions in the test phase, but by some people deliberately responding in line with the mapping learned in the acquisition phase. We return to this issue in the general discussion.

To rule out these more deliberate sources of the compatibility effect and to investigate whether spontaneously learned action-outcome associations can cause outcome stimuli to trigger ideomotor action directly, Experiments 2, 3a, and 3b used a forced-choice task, in which responses required by imperative cues or instructions were accompanied by tones that – according to the mapping learned in the acquisition phase – should trigger either compatible or incompatible responses. While compatible and incompatible trials were presented in separate blocks in Experiment 2, they were intermixed in Experiments 3a and 3b.

EXPERIMENT 2: BLOCK-BASED INTERFERENCE IDEOMOTOR TEST

In Experiment 2, we used a block-based interference ideomotor test in which participants completed two test blocks. In the compatible block, participants received instructions to respond to tones that were compatible with the earlier acquired mapping. In the incompatible block, the instructions were reversed. The order of the two test blocks was counterbalanced across participants. We expected to observe significantly reduced RTs and lower error rates in compatible blocks compared to incompatible blocks.

Method

Participants and Design

Fifty participants took part in the experiment in exchange for a small monetary payment or extra course credits. Participants with attention related disorders or those who were on related medication were excluded beforehand. Participants were randomly assigned to a cell of the 2 (Instructions: No-Instructions vs. Instructions) * 2 (Compatibility: Compatible vs. Incompatible) mixed factorial design, with Compatibility as a within-participants variable. The order of the compatible and incompatible blocks was counterbalanced across participants.

Three participants were excluded due to the unbalanced proportion of key pressing during the learning phase, that is, the balanced left-to-right key ratio (i.e., 40–60%). Data of the remaining 47 participants (No-Instructions condition: $n = 24$ vs. Instruction condition: $n = 23$) were analyzed in the test phase [23 Females, mean age: 22 years (18–31 years), no left-handed and two ambidextrous participants].

Procedure

The procedure was similar to Experiment 1. After finishing the unchanged acquisition phase, participants came to the interference ideomotor task with the compatibility manipulated on the block level. With regard to the acquired R-O mapping, the response rule participants received on one block was compatible, whereas on the other block it was incompatible. For example, if the participant got the R-O mapping A (left key – high tone, right key – low tone), the compatible block meant that participants were asked to press left key when hearing a high tone, and right key for a low tone; while the response rule in the incompatible block was reversed, that is, pressing left key for a low tone, and right key for a high tone.

Acquisition Phase

The acquisition phase was as identical as the first task of Experiment 1 (see **Figure 1A**).

Test Phase

Both the compatible and the incompatible block, consisted of 4 sub-blocks of 24 trials (see **Figure 3**). The order of the blocks was counterbalanced between participants. Each trial began with a 1500-ms fixation with an asterisk (“*”) centered in the screen, and then one of the two effect tones (i.e., the one learned in acquisition phase) was presented for 200 ms. The program would wait up to 1,000 ms to accept a response. On the first block, participants

were instructed to respond according to either the compatible or incompatible response rule. Before switching to the second block with the opposite rule of responding, participants had to perform two example trials in which the responding requirements were explained as well as four practice trials without any clues.

Manipulation Check of R-O Mappings

The questions were the same as in Experiment 1.

Manipulation Check of Instructions

The questionnaire was the same as in Experiment 1.

Results and Discussion

Acquisition Phase

Trials with response omissions (No-Instructions condition: 0.05%, Instruction condition: 0.11%) or anticipations (No-Instructions condition: 0.05%, Instructions condition: 0.05%) were excluded. After that, response proportions (left vs. right keypress) were calculated for each group. The mean left/right response proportions were equal in each condition (No-Instructions condition, 50.2% vs. 49.8%; Instruction condition: 49.6% vs. 50.4%).

The mean RTs of the participants did not differ between the No-Instructions, $M = 374.38$ ms, $SD = 33.98$ ms, and Instructions condition, $M = 376.57$ ms, $SD = 37.73$ ms, $F(1,45) = 0.04$, $p = 0.83$. The mean RTs of right responses $M = 375.82$ ms, $SD = 33.81$ ms, were not faster than the mean RTs of left responses $M = 375.09$ ms, $SD = 37.83$ ms, $F(1,45) = 0.13$, $p = 0.72$. There was also no interaction with the between-subjects factor Instructions, $F(1,45) = 0.92$, $p = 0.34$.

Test Phase

Participants who failed to meet the response criteria in the acquisition phase were excluded (3 participants). Furthermore, this time there were no trials with response anticipations

(No-Instructions condition: 0%, Instructions condition: 0%), and trials with omissions (No-Instructions condition: 1.60%, Instructions condition: 1.22%) were excluded from data analysis.

Error rates

A 3-way mixed 2 (Instructions: No-Instructions vs. Instructions) * 2 (Order: Compatible First vs. Incompatible First) * 2 (Compatibility: Compatible vs. Incompatible) ANOVA yielded a main effect of Order, $F(1,43) = 5.51$, $p = 0.02$, $\eta_p^2 = 0.11$. Neither the main effect of Instructions, $F(1,43) = 0.04$, $p = 0.85$, nor that of Compatibility, $F(1,43) = 0.00$, $p = 0.97$, was significant. No significant interaction effects between Instructions * Order $F(1,43) = 0.99$, $p = 0.32$, between Instructions * Compatibility, $F(1,43) = 0.25$, $p = 0.62$, or between Instructions * Order * Compatibility, $F(1,43) = 0.02$, $p = 0.89$, were found. Only a 2-way interaction between Order and Compatibility, $F(1,43) = 4.36$, $p = 0.04$, $\eta_p^2 = 0.09$, was found showing that the direction of the Compatibility effect was different for the two Order conditions. However, the Compatibility effect was not significant in the Compatible first condition, $t(43) = 1.46$, $p = 0.15$, nor was the Compatibility effect significant in the Incompatible first condition, $t(43) = -1.49$, $p = 0.14$.

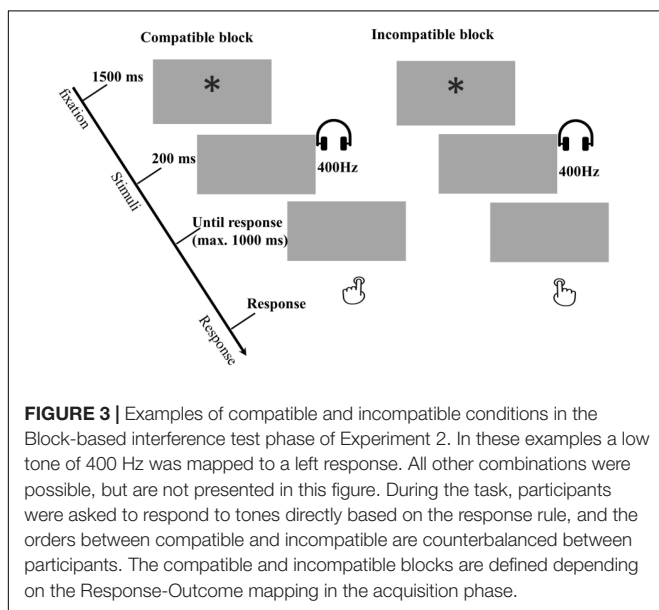
To further evaluate the evidence for the absence of a compatibility effect, the compatibility effect on error rates was calculated for all participants regardless of Instructions. If anything, errors showed a reversed compatibility effect ($M_{CE} = -0.07324$, $SD_{CE} = 0.042$) and the independent *T*-Test results, $t(46) = -0.012$, $p = 0.51$ (one-tailed), $BF_{0+} = 6.373$, indicated moderate evidence for the null hypothesis (i.e., there is no difference between compatible and incompatible condition, namely, $CE = 0$) against the one-sided alternative hypothesis (i.e., the incompatible condition has more error rates than the compatible condition, namely, $CE > 0$).

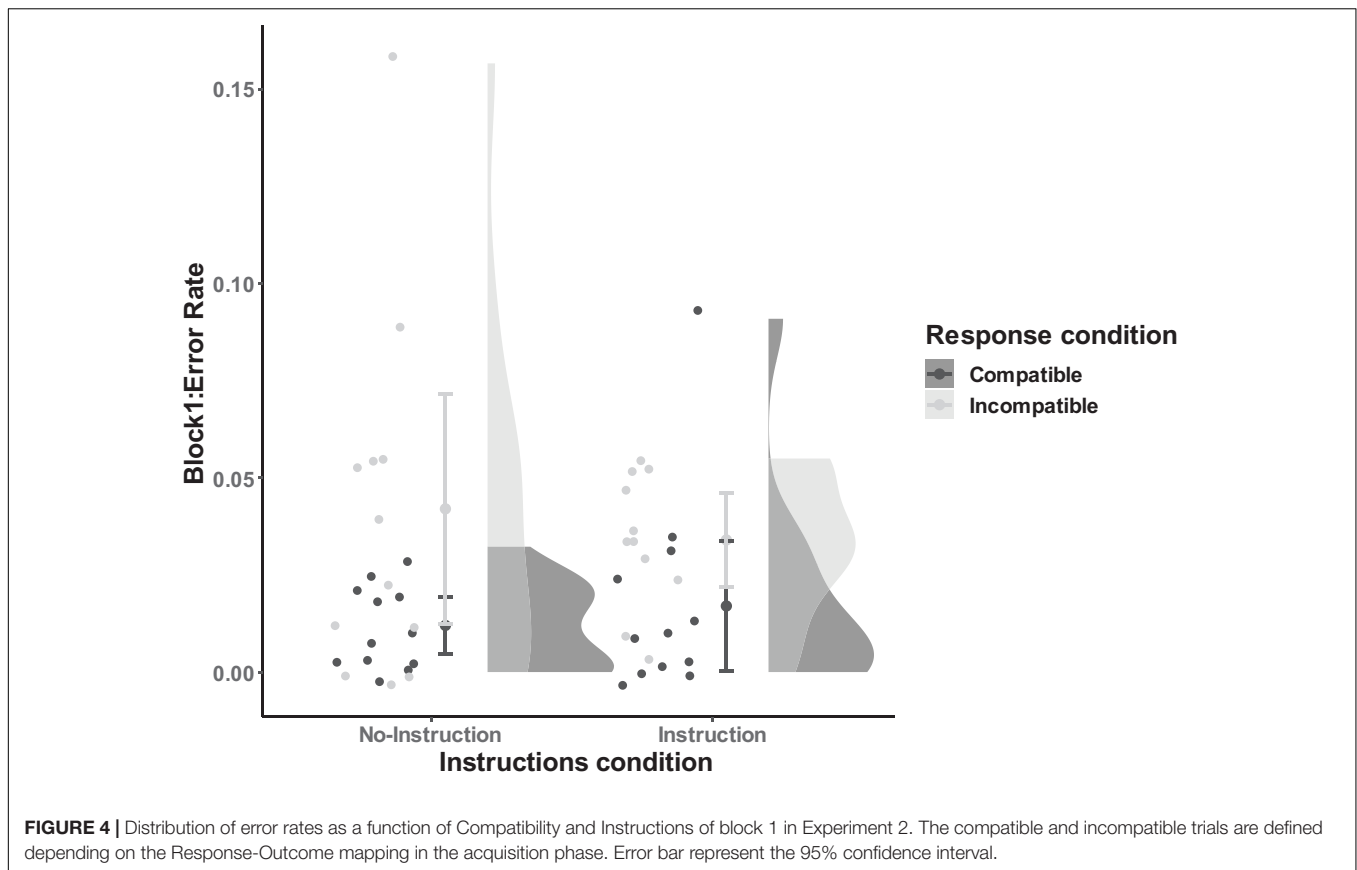
Previous research tested the compatibility effect in a between-subjects design with a non-reversal and reversal group (e.g., Elsner and Hommel, 2001, Experiment 1a, 1b). In such a design, there is only one test block and participants just receive a compatible or incompatible response rule. To perform a comparable analysis on our data we zoomed in on the first block only, with Compatibility as a between-subjects factor.

For the first block, we conducted a 2-way between-subjects ANOVA ($M_{no-instruction_compatible} = 0.012$, $SD = 0.012$; $M_{no-instruction_incompatible} = 0.042$, $SD = 0.046$; $M_{instruction_compatible} = 0.017$, $SD = 0.026$; $M_{instruction_incompatible} = 0.034$, $SD = 0.018$). The results found a significant effect of compatibility, $F(1,43) = 7.47$, $p = 0.009$, $\eta_p^2 = 0.15$, but no main effect of Instructions, $F(1,43) = 0.02$, $p = 0.90$, nor an interaction, $F(1,43) = 0.66$, $p = 0.42$ (see Figure 4 for error rates in the first block visualized distribution).

Reaction times

Mean RTs for correct trials were subjected to a 3-way 2 (Instructions: No-Instructions vs. Instructions) * 2 (Order: Compatible First vs. Incompatible First) * 2 (Compatibility: Compatible vs. Incompatible) mixed measure ANOVA, that along with the between-participants factor Instructions and





the within-participants factor Compatibility also included the counterbalancing between-participants factor Order. No main effects of Instructions, $F(1,43) = 1.67$, $p = 0.20$, Order, $F(1,43) = 0.03$, $p = 0.88$, and Compatibility, $F(1,43) = 0.18$, $p = 0.67$, were found. Furthermore, the Instruction * Order, $F(1,43) = 0.10$, $p = 0.75$, Instruction * Compatibility, $F(1,43) = 0.46$, $p = 0.50$, and Order * Compatibility, $F(1,43) = 0.65$, $p = 0.42$, interactions were not significant, neither was the 3-way interaction, $F(1,43) = 0.00$, $p = 0.99$ (see Figure 5 for visualized distribution).

To further evaluate the evidence for the absence of a compatibility effect, the compatibility effect was calculated for all participants regardless of Instructions and Order. If anything, the compatibility effect was reversed, $M_{CE} = -2.65$ ms, $M_{CE} = 42.34$ ms, and the independent T -Test results, $t(46) = -0.43$, $p = 0.665$, $BF_{0+} = 8.54$, provided relevant moderate evidence for the null hypothesis (i.e., there is no difference between compatible and incompatible condition, namely, $CE = 0$) against the one-sided alternative hypothesis (i.e., the reaction time in the incompatible condition is longer than the compatible condition, namely, $CE > 0$).

To further explore the data we zoomed in on the first block only, with Compatibility as a between-subjects factor, comparable to earlier ideomotor research. The RTs were subjected to a 2-way between-subjects ANOVA ($M_{\text{no-instruction_compatible}} = 492.46$ ms, $SD = 81.76$ ms; $M_{\text{no-instruction_incompatible}} = 475.46$ ms, $SD = 73.38$ ms; $M_{\text{instruction_compatible}} = 454.18$ ms, $SD = 57.94$ ms;

$M_{\text{instruction_incompatible}} = 459.18$ ms, $SD = 72.06$ ms). Again, no significant results were found, Instructions, $F(1,43) = 1.69$, $p = 0.20$; Compatibility: $F(1,43) = 0.08$, $p = 0.78$; Interaction: $F(1,43) = 0.28$, $p = 0.60$.

Manipulation Check of R-O Mappings

Not all participants (only 60% correct, 28 out of 47) were able to explicitly report the correct mapping of actions and outcomes they were exposed to in the acquisition phase. In the No-Instructions condition, 10 out of 24 participants failed, either forming a reversed R-O mapping, or randomly guessing the R-O mapping. The Instructions condition has similar pattern, 9 out of 23 participants missed the correct R-O mapping rule. This number may be lower than in Experiment 1, though, as the test phase also featured the opposite mapping, which may have confused participants.

Manipulation Check of Instructions

In order to assess whether there were differences in how people represented the relation between responses and outcomes in the acquisition phase, the average of each three questions measuring association, prediction, causality, and agency was calculated. The 2 (Instructions condition: No-Instruction vs. Instructions) * 4 Representation level ANOVA only found a main effect of Representation level, $F(3,135) = 4.25$, $p_{[GG]} = 0.01$, $\eta_p^2 = 0.09$, which merely showed that collapsed over Instructions conditions, there were significant differences in ratings between the four level

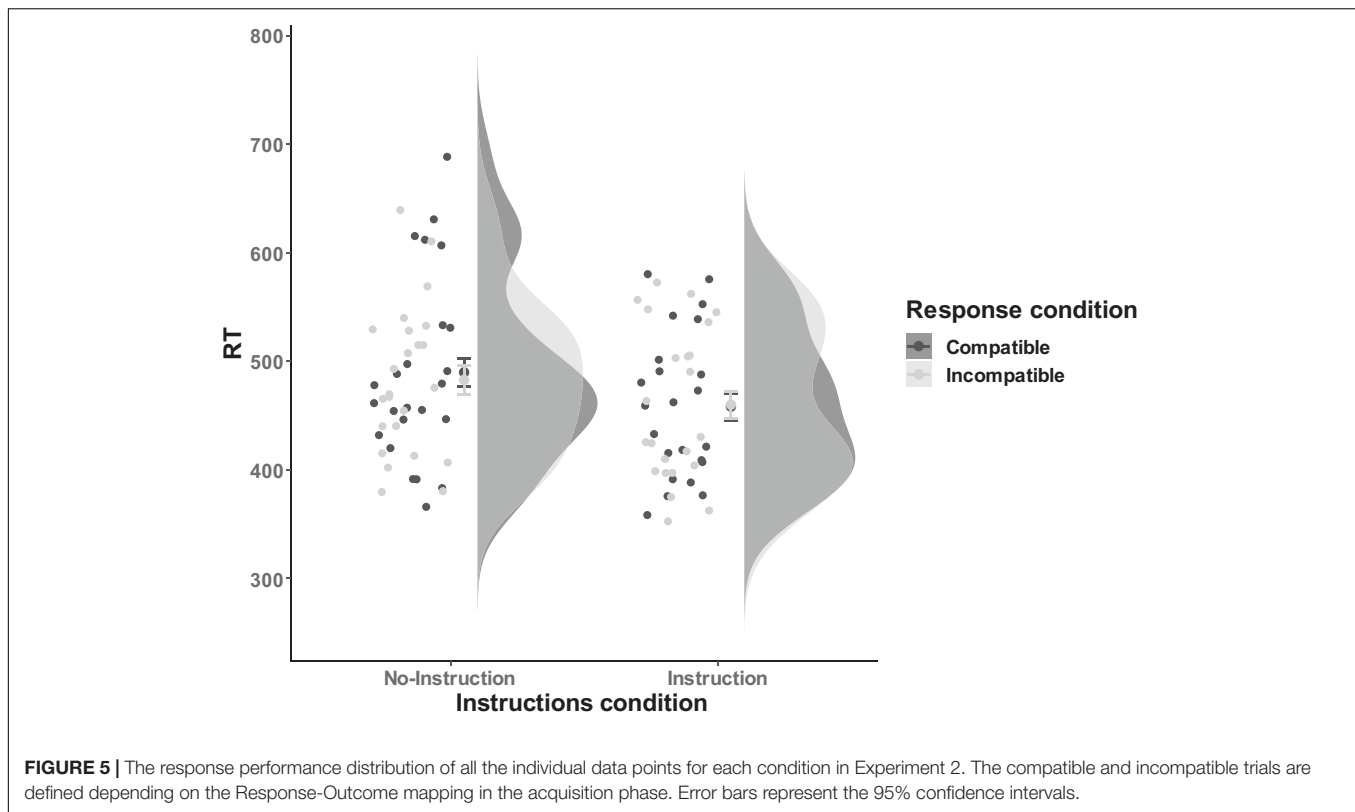


FIGURE 5 | The response performance distribution of all the individual data points for each condition in Experiment 2. The compatible and incompatible trials are defined depending on the Response-Outcome mapping in the acquisition phase. Error bars represent the 95% confidence intervals.

of representation constructs (**Table 1** presents more details of the responses to the scales).

Discussion

The block-based compatibility paradigm only provided limited support for an ideomotor effect. While no effects on RTs were found, participants made more errors on incompatible than compatible trials, though only on the first block. With no difference between instructions, this effect on errors at first glance seems to replicate the finding of Experiment 1, that ideomotor learning occurs spontaneously, also in the absence of instructions.

This compatibility effect – especially in the first block – could, however, also emerge as a result of a task switch (Monsell, 2003) that required participants who started with the incompatible block to use a new mapping, whereas participants in the compatible condition could still rely on the mapping that was learned in the acquisition phase. This effect should be less pronounced – or non-existing – in the second block, as participants in both order conditions would have to switch mappings. Note that an ideomotor effect based on an R-O association forged in the acquisition phase would predict a compatibility effect on the second block as well, as participants who entered the compatible after the incompatible block would benefit from the automatic responses triggered by the primes.

Evidence for a within-participants compatibility effect, however, was not obtained. A closer inspection of the pattern revealed that while participants who moved from a compatible to an incompatible block made more errors on the second block, showing a classic compatibility effect, participants who

moved from the incompatible to the compatible block also made more errors on the second block. This suggests that the switch in instructions from block 1 to block 2 created more errors, regardless of whether the new rule was compatible or incompatible with the acquisition phase. This may indicate that people simply struggled to switch to a new response rule.

In order to rule out this possibility Experiment 3a and 3b were conducted, in which the compatibility effect was tested at trial level. This time, participants were instructed to react to imperative cues, but were at the same time presented with stimuli that had followed responses in the acquisition phase. These stimuli should interfere with participants' responses if they are associated responses that are incompatible with the imperative cues. Such a trial-based interference ideomotor test would be the most rigorous test and cannot be regarded as a task-switch effect.

EXPERIMENT 3: TRIAL-BASED INTERFERENCE IDEOMOTOR TEST

Experiment 3a

Method

Participants and design

Sixty participants took part in the experiment in exchange for a small monetary payment or extra course credits. Participants with attention-related disorders or those who were on related medication were excluded beforehand. The experimental design consisted of one between-subjects factor: Instructions (No-Instructions vs. Instructions), and one within-subjects factor:

Compatibility (Compatible vs. Incompatible). After signing the informed consent, participants were randomly assigned to either the Instructions condition or the No-Instructions condition.

Data of one participant were lost because of a technical issue, and five participants were excluded due to the unbalanced proportion of key presses during the learning phase (outside of the range of a left-to-right ratio of 40 to 60%), which was defined before data collection. Data of the remaining 54 participants (No-Instructions condition: $n = 25$ vs. Instructions condition = 29) were analyzed in the test phase [35 females, mean age: 23 years (18–37 years), 7 left-handed and 3 ambidextrous participants].

Stimuli and procedure

We used the same sounds as in Experiment 1, plus a standard Landolt “C” ring and its mirror image, as the target for the interference ideomotor task in the test phase. We selected these stimuli because they are clearly different from the arrow stimuli in the acquisition phase, making sure that imperative cues were not associated with responses (Muhle-Karbe and Krebs, 2012). Procedures were similar to Experiment 1, including an acquisition phase and a test phase.

Acquisition phase

The acquisition phase was as identical to the one used in Experiment 1 (see Figure 1A).

Test phase

In the test phase participants were asked to perform an interference task, namely, the compatibility task, consisting of eight main blocks of 24 trials. Each trial started with a 1500-ms fixation (“*”), and then one of the former effect sounds was simultaneously presented with the Landolt “C” (see Figure 6). The duration of the prime and the target were 200 ms and 250 ms, respectively. Participants were told to detect and respond to the opening direction of Landolt “C” ring as fast and accurately as possible. Pressing the left key (“z”) for a left opening, and the right

key (“/”) for a right. The program waited up to 1,000 ms for a response. Response omissions and anticipations were defined in the same way as in the acquisition phase. There was no response feedback in the test phase.

Based on the R-O mapping in the acquisition phase, the test trials were categorized as a compatible trial when the to-be-executed response was the same as the response that was followed by the primed tone in the acquisition phase and incompatible trials when the to-be-executed response was the opposite of the response that was followed by the primed tone in the acquisition phase. For instance, if one had received the response – outcome mapping “left key – low tone, right key – high tone”, a trial was compatible when a left opening “C” ring was presented together with a low tone, and when a right opening “C” ring was presented with a high tone. A trial was incompatible when a left opening “C” ring accompanied by a high tone, and a right opening “C” ring with a low tone.

Manipulation check of R-O mappings

The questions were the same as in Experiment 1.

Manipulation check of instructions

The questionnaire was the same as in Experiment 1.

Data Analysis Plan

Analyses were similar to Experiment 1, RTs and error rates in the test phase were analyzed as a function of Instructions and Compatibility conditions.

Results

Acquisition phase

First, we excluded all acquisition trials with anticipations (No-Instructions: 0.09%, Instructions: 0.09%) and omissions (No-Instructions: 0.05%, Instructions: 0.08%). The remaining mean error rate for the No-Instructions condition was 4.78%, whereas for the Instructions condition it was 3.63%. After that, response proportions (left vs. right keypress) were calculated for each group. The mean left/right response proportions were equal in each condition (No-Instructions condition: 49.6% vs. 50.4%; Instructions condition: 49.8% vs. 50.2%).

The mean RTs of the participants did not differ between the No-Instructions, $M = 344.61$ ms, $SD = 51.34$ ms, and the Instructions condition, $M = 358.07$ ms, $SD = 43.48$ ms, $F(1,52) = 1.10$, $p = 0.30$. The mean RTs of right responses $M = 349.00$ ms, $SD = 46.91$ ms, were significantly faster than the mean RTs of left responses $M = 354.67$ ms, $SD = 48.43$ ms, $F(1,52) = 7.03$, $p = 0.01$, $\eta_p^2 = 0.12$. This effect was not qualified by an interaction with the between-subjects factor Instructions, $F(1,52) = 0.60$, $p = 0.44$.

Test phase

Participants who failed to meet the response criteria in the acquisition phase were excluded (five participants). Furthermore, trials with response anticipations (No-Instructions condition: 0.02%, Instructions condition: 0.036%) and omissions (No-Instructions condition: 0.0%, Instructions: 0.018%) were excluded from data analysis.

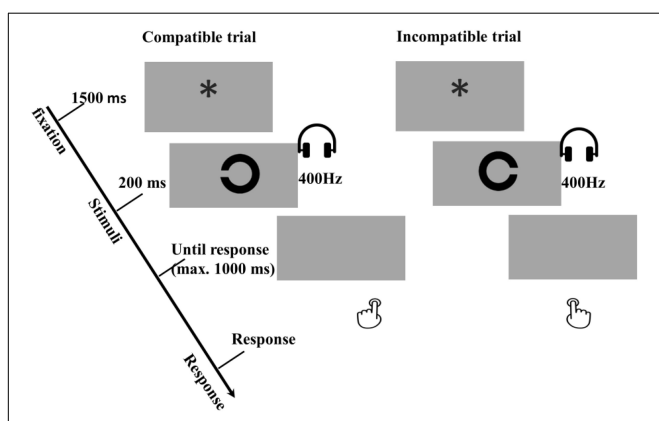
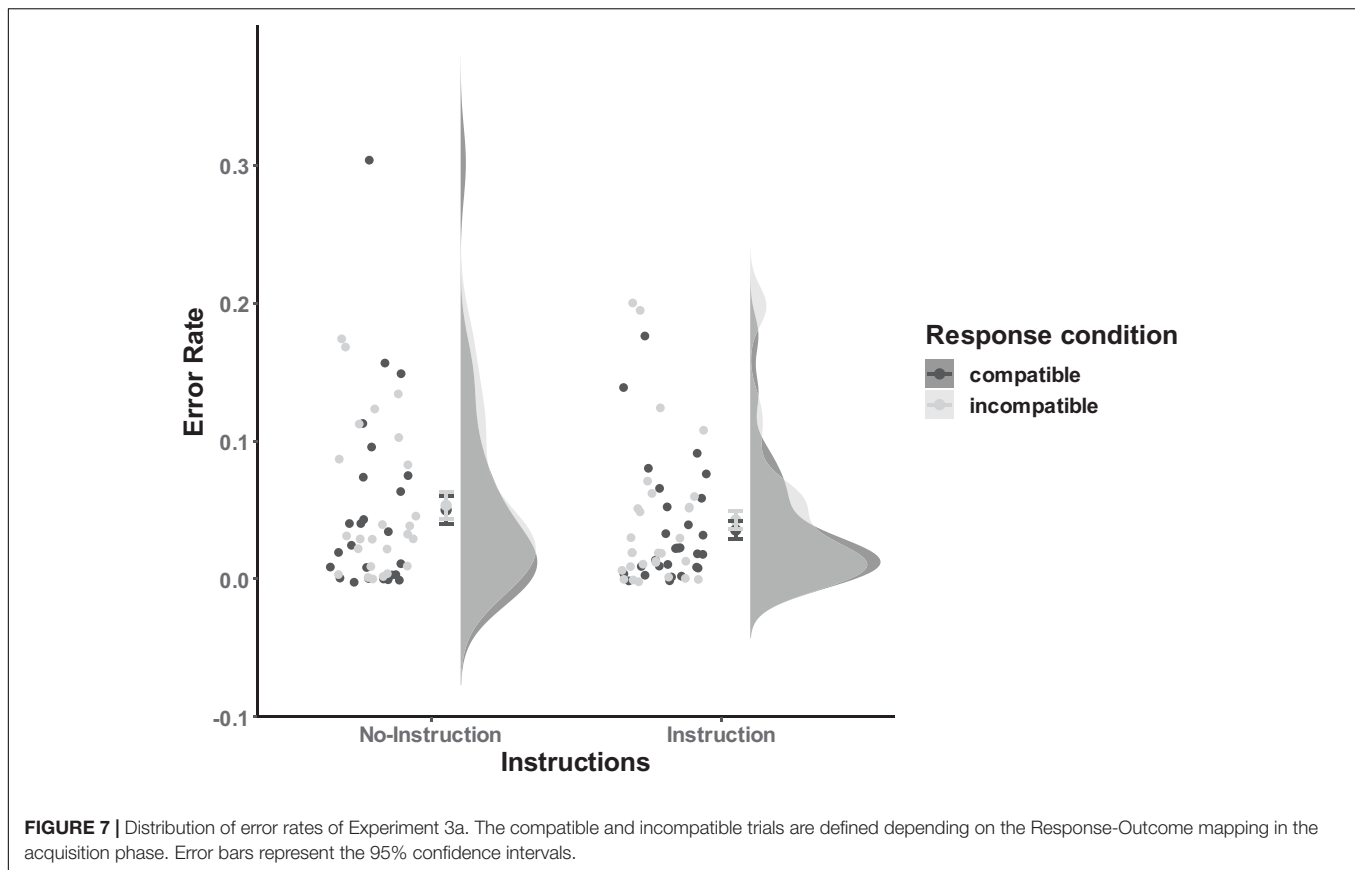


FIGURE 6 | Examples of compatible and incompatible conditions in the trial-based interference test phase of Experiment 3a and 3b. In these examples a low tone of 400 Hz was mapped to a left response, which depending on the R-O mapping in the acquisition phase. All other combinations were possible, but are not presented in this figure. The main task is the orientation discrimination task with the tones as primes, and the compatible and incompatible are intermixed in trials level.



Error rates. Error rates were analyzed based on all trials. As **Figure 7** shows, participants were relatively accurate, and most of the error rates per condition were less than 10% ($M_{\text{no-instruction_compatible}} = 0.050$, $SD = 0.070$; $M_{\text{no-instruction_incompatible}} = 0.053$, $SD = 0.054$; $M_{\text{instruction_compatible}} = 0.035$, $SD = 0.043$; $M_{\text{instruction_incompatible}} = 0.043$, $SD = 0.053$). The 2 (Instructions: No-Instructions vs. Instructions) \times 2 (Compatibility: Compatible vs. Incompatible) mixed ANOVA did not reveal any significant effects [Instructions effect: $F(1,52) = 0.74$, $p = 0.39$; Compatibility: $F(1,52) = 1.67$, $p = 0.20$; Interaction: $F(1,52) = 0.23$, $p = 0.63$].

Thereafter, in further exploratory analyses, we calculated the compatibility effect on error rates by collapsing over the Instructions factor ($M_{CE} = 0.0054$, $SD_{CE} = 0.029$). An independent T -Test, $t(53) = 1.34$, $p = 0.09$, $BF_{0+} = 1.61$, provided relevant moderate evidence for the null hypothesis (i.e., there is no difference between compatible and incompatible condition, namely, $CE = 0$) against the one-sided alternative hypothesis (i.e., the incompatible condition has more error rates than the compatible condition, namely, $CE > 0$).

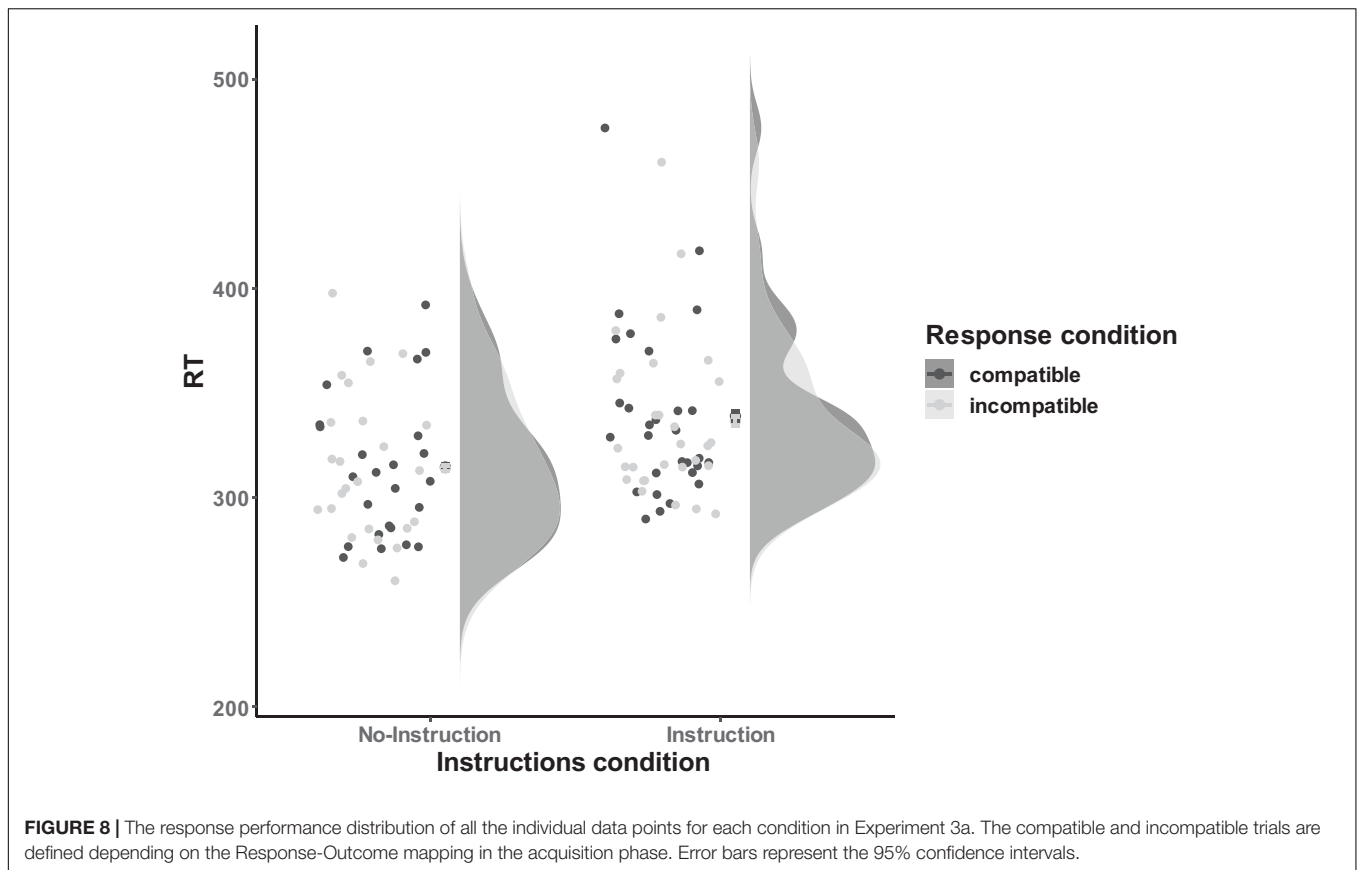
Reaction times. Reaction times (RTs) for remaining correct trials were aggregated over compatible and incompatible trials for each participant (see **Figure 8**, for visual distribution). Subsequently, the mean RTs and error rates were subjected to a 2 (Instructions: No-Instructions vs. Instructions) \times 2 (Compatibility: Compatible vs. Incompatible) ANOVA, with

Instruction as between and Compatibility as within-subjects factor. RTs analysis did not reveal a significant compatibility effect, $F(1,52) = 2.26$, $p = 0.14$. Neither the effect of interaction reached significance $F(1,52) = 0.91$, $p = 0.34$, but we found a main effect of Instruction $F(1,52) = 5.23$, $p = 0.03$, $\eta_p^2 = 0.09$, indicating that participants in the Instructions condition were overall slower to respond ($M_{\text{No-Instructions}} = 314.22$ ms, $SD = 34.57$ ms; $M_{\text{Instruction}} = 337.66$ ms, $SD = 39.65$ ms). If anything, RTs in the compatible condition, $M = 327.57$ ms, $SD = 40.05$ ms, were higher than the incompatible condition, $M = 326.05$ ms, $SD = 38.33$ ms, $t(53) = 1.58$, $p = 0.06$, $BF_{01} = 2.10$.

To further evaluate the evidence for the absence of a compatibility effect, the compatibility effect (CE) was calculated for all participants regardless of Instructions, $M_{CE} = -1.515$ ms, $SD_{CE} = 7.05$ ms. A directional T -Test, $t(53) = -1.58$, $p = 0.94$, $BF_{0+} = 16.31$, provided strong evidence for the null hypothesis (i.e., there is no difference between compatible and incompatible condition, namely, $CE = 0$) against the one-sided alternative hypothesis (i.e., the reaction time in the incompatible condition is longer than the compatible condition, namely, $CE > 0$).

Manipulation check of R-O mappings

Most participants (nearly 91% correct, 49 out of 54) were able to explicitly report the correct mapping of actions and outcomes they were exposed to in the acquisition phase. Participants could still recall previous learned R-O mapping rule. In the No-Instruction condition, 4 out of 25 failed, and only one missed



in the Instructions condition ($n = 29$). Collectively, this suggests that participants indeed acquired R-O knowledge spontaneously, although this did not translate into automatic response priming in the test phase.

Manipulation check of instructions

In order to assess whether there were differences in how people represented the relation between responses and outcomes in the acquisition phase, the average of each three questions measuring association, prediction, causality, and agency was calculated. The 2 (Instructions condition: No-Instructions vs. Instructions) * 4 Representation levels ANOVA only found a main effect of Representation level, $F(3,156) = 3.69$, $p_{[GG]} = 0.03$, $\eta_p^2 = 0.07$, which showed that collapsed over Instructions conditions, there were significant differences between the questions of the four levels. among these four levels (see **Table 1** for more details).

Discussion

The results in the present experiment did not reveal a compatibility effect in any of the two groups, suggesting that the presented outcomes did not trigger associated actions. The paradigm used, though, was designed as the strongest test for automatic action selection, with compatibility being manipulated at the trial level. In such a paradigm, compatibility effects have to arise at the trial level itself, if two stimuli evoke either the same or two conflicting responses. As far as we known, only two articles reported compatibility effects on trial level when

using the classical two-phases paradigm (Kühn et al., 2009; Sato and Itakura, 2013). However, we were not able to replicate these effects regardless of whether we provided participants with instructions to pay attention to R-O mappings in the acquisition phase or not.

Experiment 3b: Replication Trial-Based Interference Ideomotor Test

To make sure that the null findings in the rigorous test of Experiment 3a were not a false negative, we conducted a high-powered replication of the core part of Experiment 3b. Because of practical constraints we could not include the manipulation checks, but we assume based on the previous three experiments that most participants were aware of the correct mapping and that the instructions had no effect on the way participants represented the R-O relations.

Method

Participants and design

Two hundred and two participants ($N = 202$) took part in the experiment in exchange for a small monetary payment or extra course credits. Participants with attention-related disorders or those who were on related medication were excluded beforehand. The experimental design consisted of one between-subjects factor: Instructions (No-Instructions vs. Instructions), and one within-subjects factor: Compatibility (Compatible vs.

Incompatible). After signing the informed consent, participants were randomly assigned to either the Instructions condition or the No-Instructions condition.

Data of ten participants were lost because of a technical issue, and six participants were excluded due to the unbalanced proportion of key presses during the learning phase (outside of the range of a left-to-right ratio of 40 to 60%), which was defined before data collection. Data of the remaining 186 participants were analyzed in the test phase (No-Instructions condition: a total of 90 participants, 63 female, age: $M = 23$ years, $SD = 5$; Instructions condition: a total of 96 participants, 70 female, age: $M = 23$ years, $SD = 4$).

Stimuli and procedure

We used the same stimuli and procedure as mentioned in Experiment 3a, except that the procedure only had an acquisition phase and a test phase. In this experiment the experimenter was also blind to the real research goals, and waited outside the testing room.

Acquisition phase

The acquisition phase was as identical to the one used in Experiment 3a (see Figure 1A).

Test phase

The acquisition phase was as identical to the one used in Experiment 3a (see Figure 6).

Data analysis plan

Analyses were the same as Experiment 3a, RTs and error rates in the test phase were analyzed as a function of Instructions and Compatibility conditions.

Results

Acquisition phase

First, we exclude all acquisition trials with omissions (No-Instructions: 0.10%, Instructions: 0.08%) and anticipations (No-Instructions: 0.15%, Instructions: 0.12%). The remaining mean error rates for the No-Instructions condition was 5.29%, whereas for the Instructions condition it was 5.26%. After that, response proportion (left vs. right keypress) were calculated for each group. The mean left/right response proportions were equal in each condition (No-Instructions condition: 49.8% vs. 50.2%; Instructions condition: 49.5% vs. 50.5%).

The mean RTs of the participants did not differ between the No-Instructions, $M = 360.76$ ms, $SD = 52.70$ ms, and Instructions condition, $M = 356.10$ ms, $SD = 44.84$ ms, $F(1, 184) = 0.43$, $p = 0.51$. The mean RTs of left responses $M = 356.95$ ms, $SD = 49.96$ ms, were significantly faster than the mean RTs of right responses $M = 359.76$ ms, $SD = 47.69$ ms, $F(1, 184) = 4.91$, $p = 0.03$, $\eta_p^2 = 0.03$. This effect was not qualified by an interaction with the between-subjects factor Instructions, $F(1, 184) = 0.05$, $p = 0.83$.

Test phase

Participants who failed to meet the response criteria in the acquisition phase were excluded (six participants). Furthermore, trials with response anticipations (No-Instructions condition:

0.0%, Instructions condition: 0.02%) and omissions (No-Instructions condition: 0.08%, Instructions: 0.11%) were excluded from data analysis.

Error rates. Error rates were analyzed based on all valid trials. Similar to the results in Experiment 2b, participants were relatively accurate ($M_{\text{no-instruction_compatible}} = 0.0578$, $SD = 0.073$; $M_{\text{no-instruction_incompatible}} = 0.059$, $SD = 0.073$; $M_{\text{instruction_compatible}} = 0.0512$, $SD = 0.062$; $M_{\text{instruction_incompatible}} = 0.0588$, $SD = 0.090$). We employed the same 2-way mixed ANOVA with Instructions as between-subjects factor (No-Instruction vs. Instruction) and Compatibility as within-subjects factor (Compatible vs. Incompatible). Again, the results were not significant: Instructions: $F(1, 184) = 0.11$, $p = 0.74$; Compatibility: $F(1, 184) = 2.20$, $p = 0.14$; Interaction: $F(1, 184) = 1$, $p = 0.32$.

Following the analyses in Experiment 3a, we also conduct the same Bayesian one sample *T*-test for the compatibility effect on error rates by collapsing over the Instructions factor ($M_{\text{errorrates_CE}} = 0.0046$, $SD_{\text{errorrates_CE}} = 0.042$). The corresponding BF indicates more support to the null hypothesis (i.e., there is no difference between compatible and incompatible condition, namely, $CE = 0$), $t(185) = 1.516$, $p = 0.07$, $BF_{0+} = 2.13$).

Reaction times. The mean RTs on each condition ($M_{\text{no-instruction_compatible}} = 341.29$ ms, $SD = 56.13$; $M_{\text{no-instruction_incompatible}} = 341.34$ ms, $SD = 54.61$; $M_{\text{instruction_compatible}} = 339.36$ ms, $SD = 52.15$; $M_{\text{instruction_incompatible}} = 338.97$ ms, $SD = 53.62$) were also subjected to the same 2 (Instructions: No-Instructions vs. Instructions) * 2 (Compatibility: Compatible vs. Incompatible) ANOVA, with Instruction as between-subjects and Compatibility as within-subjects factors. No effects approached significance [Instructions: $F(1, 184) = 0.07$, $p = 0.79$; Compatibility: $F(1, 184) = 0.04$, $p = 0.84$; Interaction: $F(1, 184) = 0.07$, $p = 0.79$].

To further evaluate the evidence for the absence of a compatible effect, the CE was calculated for all participants regardless of Instructions ($M_{\text{CE}} = -0.179$ ms, $SD_{\text{CE}} = 11.49$), and the Bayesian one sample *T*-test still give strong evidence for the null hypothesis (H_0 : $CE = 0$), $t(185) = -0.213$, $p = 0.584$, $BF_{0+} = 14.34$).

Discussion

The results in present experiment provide a powerful replication of the effects obtained in Experiment 3a, namely strong evidence for the absence of a compatibility effect and no effects of the Instruction manipulation.

GENERAL DISCUSSION

Habits are often understood as actions that are automatically triggered by stimuli or situations through S-R associations resulting from repeated and consisted coactivation. In the present paper, we explored whether repeated and consistent coactivation of actions and effects can result in similar structures (R-O associations) by which mere perception of stimuli can then elicit the associated response (i.e.,

ideomotor action). Specifically, we investigated whether learning of R-O associations can occur spontaneously and whether as a result, these stimuli can automatically trigger associated responses. Accordingly, in four experiments, we tested automaticity in ideomotor learning in the standard two-phases paradigm that required participants to perform actions (pressing keys) that lead to specific outcomes (tones). In each experiment, we manipulated instructions in a free-choice learning phase, either making no mention in any way of the tones that followed actions, or induced a processing goal that explicitly emphasizing the relation between responses and the subsequent stimulus. In Experiment 1, evidence for ideomotor action was observed in a free-choice test phase, regardless of instructions. Experiment 2, 3a, and 3b, however, which employed forced-choice tasks to test for automaticity, provided little evidence for ideomotor effects. Together, these results don't support the strong version of ideomotor theory. That is, they suggest that ideomotor learning can occur spontaneously, but that there are limits to the automatic effect on behavior.

Mixed Evidence for Automatic Ideomotor Effects

The findings of Experiment 1 demonstrate that ideomotor learning can take place in the absence of explicit instructions that emphasize the relation between actions and outcomes. Although this finding matches with the literature on implicit learning (e.g., Cleeremans et al., 1998) and may indicate that associations have been formed as a result of coactivation of response and resulting stimulus representations, this does not necessarily mean that learning occurred outside of awareness (Melnikoff and Bargh, 2018). Indeed, given the fact that the large majority of participants could indicate which outcome was produced by which action in the acquisition phase, and the relatively high scores on these R-O mapping checks in the no-instruction and instruction conditions, it seems to be the case that although learning was spontaneous and may have resulted in associations, the acquired knowledge was clearly propositional in nature (Mitchell et al., 2009). This effect on the R-O mapping checks was consistent across all experiments, although reports were understandably less accurate when the mapping was changed during the test phase in Experiment 2. In sum, while learning occurred spontaneously, it seems that participants had explicit knowledge about which action caused which outcome in the learning phase.

The results of the different test phases across our experiments at first seem liked a mixed bag. While Experiment 1 produced a healthy ideomotor effect consistent in size with the ideomotor literature (c.f., Elsner and Hommel, 2001), Experiments 2, 3 did not provide such evidence. An exception is the effect in Experiment 2 on error rates in the first block of trials of the test phase. Below, we entertain two possible explanations to reconcile these findings.

First, one could argue that on top of the reportable causal knowledge about the action outcome mappings, people did indeed form bi-directional associations, capable of producing ideomotor effects. This explanation is consistent with the findings of Experiment 1. In accordance with the strong version of ideomotor theory (Shin et al., 2010), merely hearing the tones during the free-choice task could have automatically triggered the associated responses, leading to more mapping-consistent responses. As the learning phases and explicit reports were quite similar in Experiments 2 and 3, one would have to assume, though, that the tones at least had the potential to trigger similar responses in the corresponding test phases. Maybe the null effects there could be explained by a lack of power. This seems unlikely, though, as the number of trials is comparable with other reports in the ideomotor literature and the failure to find an effect in the high-powered replication of Experiment 3 seems more in line with the absence of an effect. It could be the case that the test tasks in Experiments 2, as well as Experiment 3a and its replication, were somehow flawed and not able to pick up the ideomotor effect. This seems unlikely as well. The tasks were closely modeled after Elsner and Hommel (2001) and should theoretically have produced the ideomotor effect, at least according to the strong version of the theory.

A theoretical explanation for the null effects in Experiment 2 and 3, though, is that people were able to suppress or inhibit ideomotor responses in the test phase. It has recently been argued that automatic responding may emerge in some tasks, but be overruled in others in which people have the goal to inhibit such responses (Melnikoff and Bargh, 2018). Although the task instructions in the test phases of Experiments 2 and 3 did not explicitly ask people to ignore the tones, it may be the case that people tried to ignore them, or at least suppress responses in order to meet the task goal. That is, responding according to the dictated response rule (Experiment 2), or responding to the visual target (Experiment 3 and its replication). It could indeed be possible that people were able to inhibit ideomotor responses in the task and exactly cancel out the effect, without revealing an opposite inhibition effect, or were fully able to shut out the auditory stimuli in the compatibility tasks, but not the free choice task. However, we believe another explanation is more plausible.

This second explanation follows the opposite line of argument: that bi-directional associations were not formed, at least not strong enough for the tones to trigger responses in an automatic fashion. This would then require and explanation for the findings in Experiment 1. In this experiment, participants engaged in a free-choice task, which – by definition – allows for deliberate control of behavior. It may have been the case that explicit knowledge about the action-outcome relations drove the behavioral effects (Seabrooke et al., 2016). Loersch and Payne (2011) have noted that such biases can occur if primes affect the explicit knowledge that is retrieved and used as input for the decision-making process. Although this does not necessarily imply that participants were aware of this bias, it would entail an indirect priming effect that operates

through biasing conscious decisions rather than by stimuli automatically triggering responses. Although this may suggest that people use knowledge of R-O mappings to freely select their actions, this would not be ideomotor action according to the strong version of the theory. Interestingly, though, such a process fits well with action control models that consider the preparation of human behavior to be rooted in sensorimotor processes that operate under radar of conscious awareness, while the ultimate execution of actions is under the control of a decision making process that selects actions associated with an act of conscious will (Gold and Shadlen, 2007; Brass and Haggard, 2008; Aarts, 2012; Zedelius et al., 2014).

Another explanation for the findings of Experiment 1 is that participants may have used the tones to fulfill the criteria of responding randomly and equally often with the two keys, or have chosen to respond with the keys suggested by the tones simply because it is easier. Random selection of responses is extremely hard and the tones may have provided an easy way out. Note that this explanation still assumes that people use the R-O knowledge that was spontaneously obtained in the acquisition phase. As a considerable number of participants responded consistent with the mapping of the acquisition phase on nearly 100% of the trials (and two individuals in close to 0% of the time, reflecting the use of a reversed mapping; see **Figure 2**), this seems a plausible explanation. Although papers in the ideomotor literature typically don't provide information about the distribution of scores, the means and standard deviations in the present study are remarkably similar to earlier studies (e.g., Elsner and Hommel, 2001) suggesting that these studies may be open to the same explanation.

In Experiment 2, we found no within-participants compatibility effects, but did obtain a difference in error rates in the first block of the experiment. While this effect is consistent with the classic forced-choice effect (e.g., Elsner and Hommel, 2001, Experiments 1a, 1b), these effects could also be interpreted as a task-switching effect (Monsell, 2003). That is, in the light of the explicit knowledge about the R-O mapping in the acquisition phase, the instruction to use the opposite mapping to respond to the outcome stimuli in the test phase could have caused the increase in errors. Hence, the obtained compatibility effect may say more about the challenges of remembering and responding according to reversed task rules, than ideomotor effects. The complexity of obtaining ideomotor effects under forced-choice conditions (Herwig et al., 2007; Pfister et al., 2011), and the relative absence of ideomotor effects in the forced-choice task in the present study indicates that further inquiry is needed to specify when and how ideomotor learning effects emerge in the test paradigms employed so far.

Implications for Habits

Although ideomotor learning can create R-O associations, only weak evidence for the ideomotor effect was obtained. So based on the current data, it seems that S-R associations underlying habits function in a different way than the R-O learning that drove the ideomotor effect in our free-choice test phase in

Experiment 1. This does not necessarily mean that ideomotor action should be discarded as a mechanism by which outcome stimuli can trigger responses, in a similar way as stimuli trigger habitual responses. As the ideomotor effect has been demonstrated across a large literature (although often with less strict tests than in the current experiments), it could be the case that the ideomotor effect holds, but that the learning phase in our experiments was too short for R-O associations to develop through co-activation, and that habit-like structures take longer to develop. Moreover, research on rewards in ideomotor learning has demonstrated that rewarding stimuli that follow responses produce much stronger ideomotor effects in free-choice or instructed compatibility tasks (Muhle-Karbe and Krebs, 2012; Eder et al., in press). It may be the case that ideomotor learning is therefore more likely to occur in daily life, where stimuli following actions are rarely neutral. Interestingly, with this notion, the ideomotor effect becomes similar to the Pavlovian Instrumental Transfer (PIT) effect, which holds that stimuli associated with rewards are found to facilitate instrumental responses that have been followed by those rewards during learning (Watson and de Wit, 2018). Such a mechanism may reflect habitual responses that are still mediated by outcome representations at some level.

Further research is needed, though, to determine how rewards boost responses in the ideomotor and PIT paradigm. As ideomotor studies on this topic (Muhle-Karbe and Krebs, 2012; Eder et al., in press) used a block-based compatibility paradigm, the enhanced effects could still be the results of explicit knowledge, as a result of propositional learning, interfering with conflicting task instructions. Relatedly, recent investigations into the nature of the PIT effect have demonstrated that the PIT effect itself is also dependent on propositional learning (Trick et al., 2011; Seabrooke et al., 2016, 2017). As here it is also unclear whether rewards influence learning, response execution, or both, it is hard to predict whether the same results would emerge in a trial-based compatibility task, to provide strong evidence for the automaticity of ideomotor action.

Conclusion

Together, while the current findings do provide evidence for spontaneous ideomotor learning, it is less evident how resulting response-stimulus representations subsequently guide behavior. Rather than automatically facilitating responses, it may be the case that R-O knowledge affects behavior in a less automatic way. While primed outcomes may activate knowledge of associated actions (Bargh et al., 2001; Custers and Aarts, 2005, 2010), they may influence behavior indirectly by biasing conscious choice (see e.g., Custers et al., 2012). As such, responses following outcome primes may be more the result of biased choice than direct response priming. Given the parallels between ideomotor thinking and the study of habitual behavior, the current work suggests that research on habitual behaviors could benefit from more careful experimentation and theorizing (Marien et al., 2018, 2019) to help understand in which ways cues in the environment could elicit habitual behavior.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the principles of the Declaration of Helsinki and the Dutch Code of Conduct for Scientific Practices as determined by the VSNU Association of Universities in Netherlands, Faculty Ethics Review Board (FERB) of the Faculty of Social and Behavioral at Utrecht University with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Faculty Ethics Review Board (FERB) of the Faculty of Social and Behavioral at Utrecht University.

REFERENCES

- Aarts, H. (2012). "Goals, motivated social cognition, and behavior," in *The SAGE Handbook of Social Cognition*, eds S. T. Fiske, and N. C. Marrae, (Thousand Oaks, CA: Sage), 75–95. doi: 10.4135/9781446247631.n5
- Aarts, H., and Dijksterhuis, A. (2000). Habits as knowledge structures: automaticity in goal-directed behavior. *J. Personal Soc. Psychol.* 78, 53–63. doi: 10.1037/0022-3514.78.1.53
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., and Kievit, R. (2018). Raincloud plots: a multi-platform tool for robust data visualization. *PeerJ Prepr.* 6:e27137v1. doi: 10.1044/1092-4388(2010/09-0205
- Ansfield, M. E., and Wegner, D. M. (1996). "The feeling of doing," in *The Psychology of Action: Linking Cognition and Motivation to Behavior*, eds P. M. Gollwitzer, and J. A. Bargh, (New York, NY: Guilford Press).
- Bargh, J. A., Lee-Chai, A., Barndollar, K., Gollwitzer, P. M., and Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *J. Personal. Soc. Psychol.* 81, 1014–1027. doi: 10.1037/0022-3514.81.6.1014
- Brass, M., and Haggard, P. (2008). The what, when, whether model of intention action. *Neurosci.* 14, 319–325. doi: 10.1177/1073858408317417
- Carpenter, B. W. (1852). On the influence of suggestion in modifying and directing muscular movement, independently of volition. *Proc. R. Inst.* 1, 147–154.
- Cleeremans, A., Destrebecqz, A., and Boyer, M. (1998). Implicit learning: news from the front. *Trends Cogn. Sci.* 2, 406–416. doi: 10.1016/S1364-6613(98)01232-7
- Cooper, R. P., and Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychol. Rev.* 113, 887–916. doi: 10.1037/0033-295X.113.4.887
- Custers, R., and Aarts, H. (2005). Positive affect as implicit motivator: on the nonconscious operation of behavioral goals. *J. Personal. Soc. Psychol.* 89, 129–142. doi: 10.1037/0022-3514.89.2.129
- Custers, R., and Aarts, H. (2010). The unconscious will: how the pursuit of goals operates outside of conscious awareness. *Science* 329, 47–50. doi: 10.1126/science.1188595
- Custers, R., and Aarts, H. (2011). Learning of predictive relations between events depends on attention, not on awareness. *Conscious. Cogn.* 20, 368–378. doi: 10.1016/j.concog.2010.05.011
- Custers, R., Eitam, B., and Bargh, J. A. (2012). "Conscious and unconscious processes in goal pursuit," in *Goal-Directed Behavior*, eds H. Aarts, and A. J. Elliot, (London: Psychology Press), 231–266.
- Dijksterhuis, A., and Bargh, J. A. (2001). "The perception-behavior expressway: automatic effects of social perception on social behavior," in Vol.33, *Advances*

AUTHOR CONTRIBUTIONS

DS, RC, and HA designed the study. DS collected and analyzed the data. DS, RC, HM, and HA drafted the manuscript. RC and HM provided the critical revisions. All authors approved the final version of the manuscript for submission.

FUNDING

This research was supported by a China Scholarship Council grant awarded to DS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00185/full#supplementary-material>

- in *Experimental Social Psychology*, (Cambridge, MA: Academic Press), 1–40. doi: 10.1016/s0065-2601(01)80003-4
- Eder, A. B., and Dignath, D. (2017). Influence of verbal instructions on effect-based action control. *Psychol. Res.* 81, 355–365. doi: 10.1007/s00426-016-0745-6
- Eder, A. B., Erle, T. M., and Kunde, W. (in press). Reward strengthens action-effect binding. *Motiv. Sci. Adv. online publ.* doi: 10.1037/mot0000153
- Elsner, B., and Hommel, B. (2001). Effect anticipation and action control. *J. Exp. Psychol.* 27, 229–240. doi: 10.1037//0096-1523.27.1.229
- Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574. doi: 10.1146/annurev.neuro.29.051605.113038
- Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: with special reference to the ideomotor mechanism. *Psychol. Rev.* 77, 73–99. doi: 10.1037/h0028689
- Herwig, A., and Horstmann, G. (2011). Action-effect associations revealed by eye movements. *Psychon. Bull. Rev.* 18, 531–537. doi: 10.3758/s13423-011-0063-3
- Herwig, A., Prinz, W., and Waszak, F. (2007). Two modes of sensorimotor integration in intention-based and stimulus-based actions. *Q. J. Exp. Psychol.* 60, 1540–1554. doi: 10.1080/17470210601119134
- Herwig, A., and Waszak, F. (2009). Intention and attention in ideomotor learning. *Q. J. Exp. Psychol.* 62, 219–227. doi: 10.1080/17470210802373290
- Heyes, C. (2010). Where do mirror neurons come from? *Neurosci. Biobehav. Rev.* 34, 575–583. doi: 10.1016/j.neubiorev.2009.11.007
- Hommel, B. (1993). The relationship between stimulus processing and response selection in the Simon task: Evidence for a temporal overlap. *Psychol. Res.* 55, 280–290. doi: 10.1007/BF00419688
- Hommel, B. (2013). "5 Ideomotor action control: on the perceptual grounding of voluntary actions and agents," in *Action science: Foundations of an Emerging Discipline*, eds W. Prinz, M. Beisert, and A. Herwig, (Cambridge, MA: MIT Press), 113–136. doi: 10.7551/mitpress/9780262018555.001.0001
- Hommel, B., Alonso, D., and Fuentes, L. J. (2003). Acquisition and generalization of action effects. *Vis. Cogn.* 10, 965–986. doi: 10.1080/13506280344000176
- Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behavi. Brain Sci.* 24, 849–878. doi: 10.1017/S0140525X01000103
- Iacoboni, M. (2008). Imitation. Empathy, and Mirror Neurons. *Annu. Rev. Psychol.* 60, 653–670. doi: 10.1146/annurev.psych.60.110707.163604
- James, W. (1890). *The Principles of Psychology*. New York, NY: Henry Holt and Company.
- Kruglanski, A. W., Shah, J. Y., Friedman, R., and Sleeth-keppeler, D. (2002). A theory of goal-system. *Advan. Exp. Soc. Psychol.* 34, 331–378.

- Kühn, S., Elsner, B., Prinz, W., and Brass, M. (2009). Busy doing nothing: evidence for nonaction-effect binding. *Psychon. Bull. Rev.* 16, 542–549. doi: 10.3758/PBR.16.3.542
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863
- Liefoghe, B., Braem, S., and Meiran, N. (2018). The implications and applications of learning via instructions. *Acta Psychol.* 184, 1–3. doi: 10.1016/J.ACTPSY.2017.09.015
- Loersch, C., and Payne, B. K. (2011). The situated inference model: an integrative account of the effects of primes on perception, behavior, and motivation. *Perspect. Psychol. Sci.* 6, 234–252. doi: 10.1177/1745691611406921
- Lotze, H. (1852). *Medizinische Psychologie oder, Physiologie der seele*. Leipzig: Weidmann'sche Buchhandlung.
- Marien, H., Custers, R., and Aarts, H. (2018). "Understanding the formation of human habits: an analysis of mechanisms of habitual behaviour," in *The Psychology of Habit*, ed. B. Verplanken, (Cham: Springer International Publishing), 51–69. doi: 10.1007/978-3-319-97529-0_4
- Marien, H., Custers, R., and Aarts, H. (2019). Studying human habits in societal context: examining support for a basic stimulus–response mechanism. *Curr. Dir. Psychol. Sci.* 28, 614–618. doi: 10.1177/0963721419868211
- Melnikoff, D. E., and Bargh, J. A. (2018). *The Mythical Number Two*. *Trends in Cognitive Sciences*. Amsterdam: Elsevier Current Trends.
- Mitchell, C. J., De Houwer, J., and Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behav. Brain Sci.* 32, 183–198. doi: 10.1017/S0140525X09000855
- Monsell, S. (2003). Task switching. *Trends Cogn. Sci.* 7, 134–140. doi: 10.1016/S1364-6613(03)00028-7
- Muhle-Karbe, P. S., and Krebs, R. M. (2012). On the influence of reward on action-effect binding. *Front. Psychol.* 3:450. doi: 10.3389/fpsyg.2012.00450
- Pfister, R., Kiesel, A., and Hoffmann, J. (2011). Learning at any rate: action-effect learning for stimulus-based actions. *Psychol. Res.* 75, 61–65. doi: 10.1007/s00426-010-0288-1
- Powers, W. T. (1973). Feedback: beyond behaviorism. *Science* 179, 351–356.
- R Core Development Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Development Team.
- Sato, A., and Itakura, S. (2013). Intersubjective action-effect binding: eye contact modulates acquisition of bidirectional association between our and others' actions. *Cognition* 127, 383–390. doi: 10.1016/J.COGNITION.2013.02.010
- Seabrooke, T., Hogarth, L., and Mitchell, C. J. (2016). The propositional basis of cue-controlled reward seeking. *Q. J. Exp. Psychol.* 69, 2452–2470. doi: 10.1080/17470218.2015.1115885
- Seabrooke, T., Le Pelley, M. E., Hogarth, L., and Mitchell, C. J. (2017). Evidence of a goal-directed process in human pavlovian-instrumental transfer. *J. Exp. Psychol.* 143, 377–387. doi: 10.1037/xan0000147
- Shin, Y. K., Proctor, R. W., and Capaldi, E. J. (2010). A Review of Contemporary Ideomotor Theory. *Psychol. Bull.* 136, 943–974. doi: 10.1037/a0020541
- Singmann, H., Bolker, B., Westfall, J., and Aust, F. (2016). *Afex: Analysis of Factorial Experiments*. Available at: <https://cran.r-project.org/package=afex>
- Stock, A., and Stock, C. (2004). A short history of ideomotor action. *Psychol. Res.* 68, 176–188. doi: 10.1007/s00426-003-0154-5
- Trick, L., Hogarth, L., and Duka, T. (2011). Prediction and uncertainty in human pavlovian to instrumental transfer. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 757–765. doi: 10.1037/a0022310
- Vallacher, R. R., and Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychol. Rev.* 94, 3–15. doi: 10.1037/0033-295X.94.1.3
- van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., et al. (2019). The JASP guidelines for conducting and reporting a bayesian analysis. *PsyArxiv*. [preprint]. doi: 10.31234/osf.io/yqxf
- Watson, P., and de Wit, S. (2018). Current limits of experimental research into habits and future directions. *Curr. Opin. Behav. Sci.* 20, 33–39. doi: 10.1016/j.cobeha.2017.09.012
- Wood, W., and Rünger, D. (2016). Psychology of habit. *Annu. Rev. Psychol.* 67, 289–314. doi: 10.1146/annurev-psych-122414-033417
- Zedelius, C. M., Veling, H., Custers, R., Bijleveld, E., Chiew, K. S., and Aarts, H. (2014). A new perspective on human reward research: How consciously and unconsciously perceived reward information influences performance. *Cogn. Affect. Behav. Neurosci.* 14, 493–508. doi: 10.3758/s13415-013-0241-z
- Ziessler, M., and Nattkemper, D. (2002). "Effect anticipation in action planning," in *Common Mechanisms in Perception and Action: Attention and Performance XIX*, eds W. Prinz, and B. Hommel, (Oxford: Oxford University Press), 645–672.
- Zwaan, R. A., and Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *J. Exp. Psychol.* 135, 1–11. doi: 10.1037/0096-3445.135.1.1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Sun, Custers, Marien and Aarts. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



How Sequential Interactive Processing Within Frontostriatal Loops Supports a Continuum of Habitual to Controlled Processing

Randall C. O'Reilly^{1,2*}, Ananta Nair², Jacob L. Russin¹ and Seth A. Herd^{1,2}

¹ Computational Cognitive Neuroscience Lab, Department of Psychology, Computer Science, and Center for Neuroscience, University of California, Davis, Davis, CA, United States, ² eCortex, Inc., Boulder, CO, United States

We address the distinction between habitual/automatic vs. goal-directed/controlled behavior, from the perspective of a computational model of the frontostriatal loops. The model exhibits a continuum of behavior between these poles, as a function of the interactive dynamics among different functionally-specialized brain areas, operating iteratively over multiple sequential steps, and having multiple nested loops of similar decision making circuits. This framework blurs the lines between these traditional distinctions in many ways. For example, although habitual actions have traditionally been considered purely automatic, the outer loop must first decide to allow such habitual actions to proceed. Furthermore, because the part of the brain that generates proposed action plans is common across habitual and controlled/goal-directed behavior, the key differences are instead in how many iterations of sequential decision-making are taken, and to what extent various forms of predictive (model-based) processes are engaged. At the core of every iterative step in our model, the basal ganglia provides a “model-free” dopamine-trained Go/NoGo evaluation of the entire distributed plan/goal/evaluation/prediction state. This evaluation serves as the fulcrum of serializing otherwise parallel neural processing. Goal-based inputs to the nominally model-free basal ganglia system are among several ways in which the popular model-based vs. model-free framework may not capture the most behaviorally and neurally relevant distinctions in this area.

Keywords: habits, goals, controlled processing, automatic processing, computational modeling, frontal cortex, basal ganglia

OPEN ACCESS

Edited by:

John A. Bargh,
Yale University, United States

Reviewed by:

Ion Juvina,
Wright State University, United States
Terrence C. Stewart,
University of Waterloo, Canada

*Correspondence:

Randall C. O'Reilly
oreilly@ucdavis.edu

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 26 June 2019

Accepted: 18 February 2020

Published: 10 March 2020

Citation:

O'Reilly RC, Nair A, Russin JL and
Herd SA (2020) How Sequential
Interactive Processing Within
Frontostriatal Loops Supports
a Continuum of Habitual to Controlled
Processing. *Front. Psychol.* 11:380.
doi: 10.3389/fpsyg.2020.00380

INTRODUCTION

Since its inception, the field of psychology has been fascinated by the distinction between two types of behavior, one that leads us to act relatively automatically, according to well-worn habits, and another that allows us to act with intent and deliberation (James, 1890; Thorndike, 1911; Hull, 1943; Tolman, 1948). These two classes of thought and action have been referred to by several different sets of terminologies, each with slightly varying definitions, which has sown some confusion in the literature (Hassin et al., 2009; Kool et al., 2018; Miller et al., 2018, 2019). Historically, the first terminology applied to this intuitive distinction was *stimulus-response* vs. *cognitive-map guided*

(Thorndike, 1911; Tolman, 1948). This distinction was later replaced by *habitual* vs. *goal-directed behavior* (Tolman, 1948; Balleine and Dickinson, 1998; Dickinson and Balleine, 2002; Killcross and Coutureau, 2003; Balleine, 2005; Yin and Knowlton, 2006; Tricomi et al., 2009), which co-existed alongside *automatic* vs. *controlled* processing (Shiffrin and Schneider, 1977; Cohen et al., 1990; Miller and Cohen, 2001). More recently, a good deal of work has been directed at the distinction between *model-free* and *model-based* reinforcement learning (Sutton and Barto, 1998; Doya, 1999; Doya et al., 2002; Daw et al., 2005).

In this paper, we attempt to clarify the relationships among these terminological distinctions through the lens of a computational model of the underlying brain mechanisms. This model builds on detailed neural recording data available on animal action-selection. One of the major conclusions from this model is that these apparently distinct types of behavior may be manifestations of a core underlying neural system, which evaluates the relative cost/benefit tradeoffs of engaging in more time-consuming, deliberative processing using the same basic mechanisms that drive all the other behavioral decisions that an organism must make. Furthermore, we argue that the neural pathways that support the habitual stimulus-response level behavior are actually an integral part of the same system that supports deliberative, controlled processing. Thus, this framework provides a unified view of action selection and decision making from the most basic habitual level up to the most complex, difficult decisions that people face. In our theory, Type 2 (deliberative) decisions are essentially composed of many Type 1 (automatic) decisions. Thus, it offers a mechanistic explanation of the proposed continuum between them (Melnikoff and Bargh, 2018).

Goal-Driven/Controlled vs. Habitual/Automatic

We first establish some common ground by attempting to define a consensus view about the closely-related distinctions between *goal-driven* vs. *habitual*, and *controlled* vs. *automatic* processing. Of the two, controlled vs. automatic (Shiffrin and Schneider, 1977) is perhaps more clearly defined, by virtue of a history of computational models based on the idea that the prefrontal cortex (PFC) supports controlled processing by maintaining active working memory representations that drive top-down biasing of processing elsewhere in the brain (Cohen et al., 1990; Miller and Cohen, 2001; Herd et al., 2006; O'Reilly, 2006; O'Reilly and Frank, 2006). Cognitive control is needed to support novel, difficult, complex tasks, e.g., to overcome prepotent (i.e., habitual) response pathways in the widely-studied Stroop task. As a task or stimulus-response pathway becomes more strongly practiced, behavior becomes more automatic and free from the need for this top-down biasing support. Thus, automatic and habitual are closely related terms. The connection between goal-driven and controlled processing is somewhat less exact, as one could imagine behaving according to goals that do not require significant cognitive control (Bargh, 1989), and potentially even exerting cognitive control in the absence of clear

goal-driven motivations. Sustained active neural firing of goal-like representations, that can exert an ongoing biasing effect on behavior, is perhaps a more direct mechanistic connection between the two.

Phenomenologically, habitual behavior is typically characterized as being relatively insensitive to the current reward value of actions, and not as strongly under the control of active, conscious goal engagement (Wood and Runger, 2016). On the other hand, it remains a challenge to consider the nature of real-world behaviors that are characterized as habits, as they often involve extended sequences of actions coordinated over reasonably long periods of time (e.g., driving home from work, making coffee, etc.) – these do not seem to be entirely unconscious activities devoid of any cognitive control influences, or contextual sensitivity (Cushman and Morris, 2015). Furthermore, how can it be that subtle, unconscious factors can sometimes strongly shape our overt behavior (Bargh, 2006; Huang and Bargh, 2014)?

Our general answer to these questions, as captured in our computational modeling framework, is that both habits and more controlled, goal-driven behaviors emerge from a shared neural system, and both operate within a common *outer-loop* of overall cognitive control that pervasively shapes and modulates the nature of processing performed in the *inner-loops* associated with specific task performance. This is similar to the hierarchical control framework of Cushman and Morris (2015), except that we postulate a sequential, temporal organization of decision making and control, where the same neural systems iteratively process multiple steps over time, including periodic revisiting of the broader context and goals that we refer to as the *outer-loop*. Thus, habits only drive behavior when permitted by this outer-loop of cognitive control, and indeed the actual unfolding of behavior over time is usually at least somewhat coordinated by the outer-loop. Furthermore, as we'll elaborate below, a crucial factor across all behavior in our framework is a so-called *Proposer* system that integrates many different factors in a *parallel-constraint-satisfaction* system to derive a proposed plan of action at any point in time – the properties of this system may explain how unconscious factors can come to influence overt behavior in the course of solving the *reduction problem* of choosing one plan among many alternatives (Bargh, 2006; Huang and Bargh, 2014).

The Model-Free vs. Model-Based Dichotomy

Within the above context, how does the *model-based* vs. *model-free* (MBMF) framework fit in? This framework has engaged new enthusiasm by offering the promise of a more formal, precise definition of the relevant processes, and by leveraging the direct connection between reinforcement learning principles and properties of the midbrain dopamine system (Montague et al., 1996; Schultz, 2013). Specifically, the model-free component is typically defined as relying on learned, compiled estimates of future reward associated with a given current state (or potential actions to be taken in that state), which have been trained via phasic dopamine-like temporal difference signals, as in the classic

TD and Q-learning reinforcement learning frameworks (Sutton and Barto, 1998). By contrast, the model-based system adds an internal model that can simulate the evolution of the state of the world over multiple iterations, so that action selection can be based on those predicted states. As such, the model-free system is considered to be relatively inflexible to changes in the reward function, including changes resulting from internal state (e.g., not being hungry at the moment), whereas the model-based system can dynamically adjust its predictions based on goal changes and other changes, and is thus more flexible.

Thus, it is this key difference in the relative flexibility of the two systems that maps onto the existing notions of goal-driven vs. habitual behavior. However, there are various other aspects of the MBMF framework which map less well, creating significant confusion when people intend to characterize the goal-driven vs. habitual distinction, but using the MBMF terminology. At a very basic level, there is no principled reason why a model-free system should not have access, as inputs, to internal drive and goal states in addition to external environmental states. If it does, its behavior can also be goal-directed, and sensitive to internal bodily states such as hunger. In addition, model-based is not synonymous with goal-directed, as model-based is defined specifically in terms of models of the external environment. In our framework, a model-free-like system indeed receives internal state and goal inputs, and thus participates in goal-directed behavior. This illustrates an important mismatch between these two terminologies, which are often taken to be interchangeable. More generally, standard reinforcement-learning paradigms tend not to incorporate a significant goal-driven component, and instead generally assume a single overriding goal of maximizing a scalar-valued reward, which is delivered to an entirely externally-motivated agent (O'Reilly et al., 2014). Thus, aside from a few more recent examples (Berseth et al., 2018), standard reinforcement-learning models are not particularly well-suited for describing goal-driven processing in the first place.

Recent reviews by Miller et al. (2018, 2019) point out the following additional issues with the MBMF terminology. First, it is problematic that the model-free system relies on learned *value* estimates to drive action selection, whereas most existing data indicates that habitual behavior is specifically more insensitive to reward value (Wood and R  nger, 2016). Second, the neural substrates associated with MBMF mechanisms are largely overlapping and hard to disentangle, involving the dopaminergic system, the basal ganglia (BG), and the prefrontal cortex (PFC). Whereas the BG was traditionally viewed as being primarily a habit-based motor area (e.g., Miller, 1981; Mishkin et al., 1984; Squire and Zola-Morgan, 1991; Squire, 1992; Packard and Knowlton, 2002) more recent evidence and theorizing suggest that, with the exception of the dorsal-lateral striatum, most of the BG is more clearly involved with non-habitual behavior and deliberative, controlled cognition in novel and challenging tasks (Pasupathy and Miller, 2005; Samejima et al., 2005; Yin et al., 2005; Balleine et al., 2007; Seger and Spiering, 2011; Pauli et al., 2012). Many authors nevertheless continue to assume the simple association of model-free with the BG, in keeping with the traditional habit-based ideas.

Furthermore, while the MBMF distinction is often considered to be dichotomous, more recent work has explored various combinations of these aspects to deal with the computational intractability of full model-based control, further blurring the lines between them (Pezzulo et al., 2013; Cushman and Morris, 2015). Likewise, there are many ways of approximating aspects of model-based predictions of future outcomes that may not fit the formal definition of iterative model-state updating, e.g., using predictive learning in the *successor-representation* framework (Dayan, 1993; Littman and Sutton, 2002; Momennejad et al., 2017; Russek et al., 2017; Gershman, 2018). This may be considered acceptable if the distinction is just that the model-free system has absolutely no model-like element, and the model-based system has any kind of approximation of a world model (Daw and Dayan, 2014), but this may end up straining the value of the distinction. For instance, a successor-representation model is otherwise quite similar to a standard model-free system, but it does use information about outcomes (although they do not usually explicitly predict an outcome).

The above considerations led Miller et al. (2018, 2019) to conclude that MBMF are both aspects of the goal-based, controlled-processing system, based on the prefrontal cortex/basal ganglia/dopamine circuits in the brain, while habitual, automatic processing is supported by an entirely separate system governed by a Hebbian, associative form of learning that strengthens with repetition.

Overview of the Paper

In the remainder of this paper, we present an alternative framework based on computational models of the basal ganglia/prefrontal cortex/dopamine system, which is consistent with the overall critique of MBMF by Miller et al. (2018, 2019), and provides a specific set of ways in which these brain systems can support a continuum of goal-directed, model-based forms of decision making and action selection. The original controlled vs. automatic distinction has always incorporated this notion that these are two poles along a continuum. Our framework goes further in describing how model-based and model-free elements interact in various ways and to varying degrees to provide a rich and multi-dimensional continuum of controlled, goal-driven cognition, which also supports varying degrees and shades of habitual or automatic elements.

This framework contrasts with several others that posit strongly dichotomous and internally homogenous habitual vs. goal-driven pathways, followed by an arbiter system that decides between the two (e.g., Daw et al., 2005; Miller et al., 2019). Instead, we propose that an outer-loop of goal-driven, but model-free, processing is itself essentially an arbiter of how much time and effort to invest in any given decision-making process. It controls the degree of engagement of a broader toolkit of basic decision-making computations to be deployed, as a function of their relative tradeoffs (c.f., Pezzulo et al., 2013). In particular, it controls whether to perform additional steps of predictive modeling down each given branch of the state-space model.

We also address a critical phenomenon for any model in this domain, which is the nature of the transition from controlled to automatic processing (Cohen et al., 1990; Gray et al., 1997;

Hikosaka and Isoda, 2010). Behaviorally, this transition occurs gradually over time and appears to reflect something like the strengthening of habit representations, which offer advantages in terms of speed, resistance to distraction, and the ability to do more in parallel, at the cost of flexibility and sensitivity to current goals – i.e., the fundamental underlying tradeoffs along this dimension. However, due to the multi-component nature of our goal-driven model, there are also various ways in which learning within this system can change these relative tradeoffs, leading to a richer picture of this process of habit formation.

THE PROPOSER-PREDICTOR-ACTOR-CRITIC MODEL

Our theoretical framework has been specified as a neural network model in the Leabra framework (O'Reilly, 1998; O'Reilly and Munakata, 2000; O'Reilly et al., 2016).

The *Proposer-Predictor-Actor-Critic* (PPAC) model (Figures 1, 2; Herd et al., 2019) leverages the prototypical loops descending from all areas of frontal cortex through the basal ganglia and converging back to modulate the function of matching areas of frontal cortex (Alexander et al., 1986; Haber, 2010, 2017; Sallet et al., 2013). Functionally, these BG/PFC loops support the ability to selectively activate and maintain neural activity (i.e., working memory) in the service of supporting top-down control representations (Miller and Cohen, 2001; Frank and O'Reilly, 2006; Herd et al., 2006; O'Reilly, 2006). As such, this system is critical for controlled, goal-driven processing. The PPAC model includes an important distinction among the nature of the cortical input representations into the BG: proposed actions vs. predicted outcomes. Critically, complex decision-making unfolds sequentially across multiple iterations in the model, each of which involves parallel operations across these circuits (i.e., a *serial-parallel* model, in which parallel computations are iterated serially).

In this theory, complex decision-making consists of a series of selections of internal “actions,” each of which consists of an update to working memory and/or episodic memory. Selecting a move in chess or choosing a plane ticket to purchase may each require a large number of belief updates (like “too expensive to fly direct in the afternoon”) and the selection of several new mid-level plans (like “try to threaten a more valuable piece instead of defending the knight”). Each of these can be stored in active memory, which executes controlled processing (by exerting top-down biasing of processing (Cohen et al., 1990; Miller and Cohen, 2001; Herd et al., 2006). Maintaining each plan or belief in working memory can also create an episodic memory trace for later recall and re-use. Our theory holds that each such representation is selected for maintenance (and therefore plan execution) much as motor representations are selected, by distinct but computationally and structurally analogous circuits.

Our theory expands on existing work on action selection in the basal ganglia, and addresses the contributions of cortex to this process. As such, we adopt the terminology of an actor-critic reinforcement learning architecture (Sutton and Barto, 1998; O'Reilly and Frank, 2006) to describe the computational roles

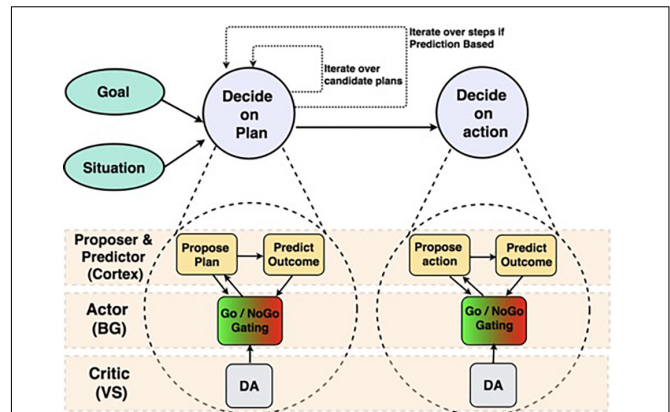


FIGURE 1 | Structure of Proposer-Predictor-Actor-Critic architecture (Herd et al., 2019) across frontal cortex and subcortical areas. We depict two parallel circuits with a hierarchical relationship. The top is a broad functional diagram, emphasizing the serially iterative and hierarchical nature of our proposed decision-making process. The bottom expands those functions, and identifies the brain areas that perform each function.

of basal ganglia and the dopamine system. The basal ganglia functions as an *Actor* that decides which action to take (or in our extended model, which plan to pursue). The dopamine release system, including amygdala, ventral striatum, and related areas, serves as a *Critic* by gauging the success of each action relative to expectations. Phasic dopamine release from this critic system serves as a reward prediction error learning signal for the basal ganglia actor system.

To this existing computational/biological theory we add two new computations, each made by participating regions of cortex. The first is a *Proposer* component. This system takes information about the current situation as input, and produces a single candidate plan representation. This proposer functional role may be less important for laboratory tasks, since they usually have a small set of actions (e.g., levers, yes/no responses), which can be learned thoroughly enough to process all options in parallel routes through the basal ganglia (e.g., Collins and Frank, 2014). However, dealing with unique real-world situations requires coming up with a potential approach before evaluating outcomes (e.g., different plausible routes in a trip planning context). This proposer system could use computations characterized as model-free, stimulus-response, constraint satisfaction, or model-based, depending on the complexity of the situation.

The other cortical addition is a *Predictor* component, which predicts the likely outcome of each proposed plan. In our model as currently implemented, this prediction always took place in two steps: predicting an “Outcome,” and from that outcome, predicting a “Result” or potential reward. We think that this type of prediction is actually performed by a variety of brain systems, using a variable number of steps for different types of decisions; but for the present purposes, it is adequate to simply think of this component as producing a prediction of an outcome by any means. This system’s computation is thus very much “model-based,” according to that terminology.

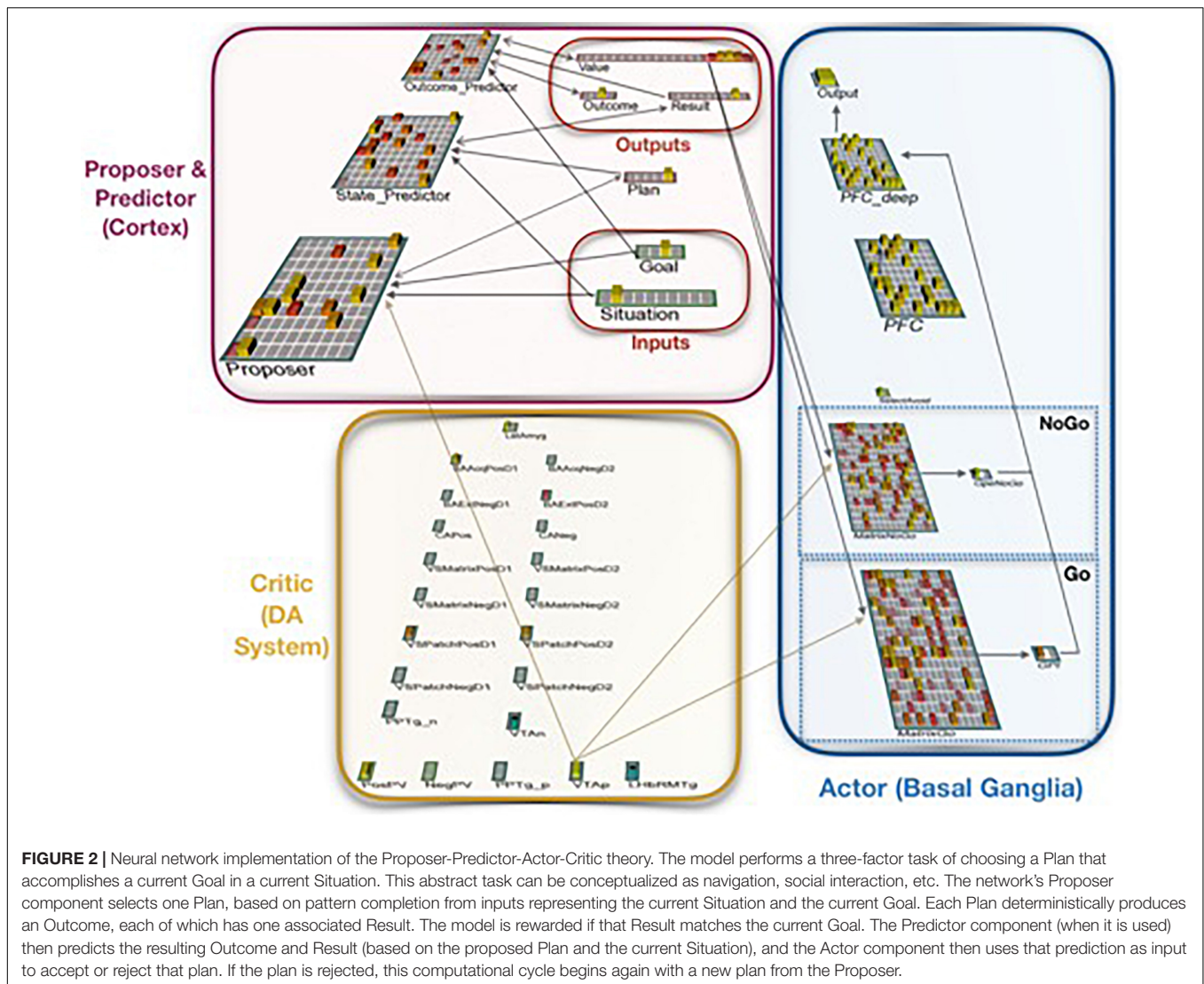


FIGURE 2 | Neural network implementation of the Proposer-Predictor-Actor-Critic theory. The model performs a three-factor task of choosing a Plan that accomplishes a current Goal in a current Situation. This abstract task can be conceptualized as navigation, social interaction, etc. The network's Proposer component selects one Plan, based on pattern completion from inputs representing the current Situation and the current Goal. Each Plan deterministically produces an Outcome, each of which has one associated Result. The model is rewarded if that Result matches the current Goal. The Predictor component (when it is used) then predicts the resulting Outcome and Result (based on the proposed Plan and the current Situation), and the Actor component then uses that prediction as input to accept or reject that plan. If the plan is rejected, this computational cycle begins again with a new plan from the Proposer.

In our system, the Actor uses the predicted outcome (when available) of the proposed plan to either accept or reject that plan. Having this specific outcome prediction greatly simplifies the computational task of the actor component; it need simply accept plans that are predicted to have rewarding outcomes, and reject those that do not. If the proposed plan is rejected, the Proposer component makes a new, different proposal, a new prediction is made by the Predictor, and the Actor again decides to accept or reject that newly proposed plan. This operation proceeds serially until a candidate plan is selected. The serial, one-at-a-time plan consideration is slow, but computationally helpful in making an accurate prediction of outcomes in novel, poorly learned domains. It allows the full power of the cortex to be directed toward each prediction, and avoids binding problems, as we address further in the Discussion section.

This computational approach can attack complex problem spaces by sequentializing a complex decision into many sub-decisions, and allowing the actor component to accept or reject each proposed sub-plan or sub-conclusion. We propose

that our ability to sequentialize a problem into sub-steps and make a binary decision for each is the source of humans' remarkable cognitive abilities relative to other animals. This method of simplification may, however, have particular inherent weaknesses that explain some of humans' notable cognitive biases.

Continuum of Controlled/Goal-Directed vs. Automatic/Habitual

Due to its sequential, hierarchical and multi-component nature, the model provides a mechanistic basis for a continuum of controlled/goal-directed vs. automatic/habitual behavior. At every sequential step, there is the potential for an outer-loop decision about what overall strategy to employ, e.g., whether to engage in further prediction, or iterate to another proposed plan, etc. Within that outer loop, there are more specific decisions regarding what factors to focus on, such as which branches to pursue in prediction, etc.

In cases of high urgency or low stakes, all of that complexity could be elided in favor of a quick thumbs-up (Go gating decision) from the Actor to the Proposer's initial suggestion. This optimization for speed could be created by reinforcement learning in the basal ganglia, with inputs that capture timing and relevant time pressures. We suggest that this may represent the majority of habitual or automatic responding – a fast path through the very same circuits, typically at the lower levels of the abstraction hierarchy (e.g., involving supplementary motor areas and the dorsolateral striatum). Thus, consistent with the continuum perspective, and a surprising difficulty in finding explicit claims and data about what neural substrates uniquely support habitual behavior (e.g., Wickens et al., 2007; Seger and Spiering, 2011), there may be no separate neural substrate associated with habitual behavior – it is just the simplest and fastest mode of processing through the entire decision-making apparatus.

If this is the case, then it would seem to challenge the various attempts to establish strong dichotomies between e.g., model-free vs. model-based, or even value-based vs. value-free or belief-based vs. belief-free (Miller et al., 2018, 2019). In short, even habitual behavior depends on a (usually implicit) decision to not engage in a more controlled form of behavior, and that decision likely depends on assessments of the relevant “stakes” (values or utilities) in the current context, and the estimated cost/benefit tradeoffs in engaging in more effortful levels of control (Pezzulo et al., 2013).

Thus, estimated value is always in play, even in the context of habitual behavior. To reconcile this idea with the finding that habitual behaviors are relatively insensitive to changes in reward, we would need to determine the relative cost/benefit tradeoff estimates associated with the alternative options that might have been taken instead of performing the habitual response. Certainly, if the habitual response would lead to imminent severe harm, and this was obvious to the individual, then we would expect them not to engage in it. Typically when clearly erroneous habitual responding occurs in the real world, it can be traced to a lack of attention being paid to the relevant factors, likely resulting from prior decisions to allocate that attention elsewhere. In other words, taken literally, a purely habitual response presumes that the person is otherwise somewhat of a zombie. Instead, we suggest, consistent with others (e.g., Cushman and Morris, 2015) that habitual responses occur within a broader context (i.e., the outer-loop) of at least some level of cognitive control.

The Model-Free Actor in the Loop

A central feature of our model is that the basal ganglia Actor system provides a value-based final Go/NoGo decision, even (and perhaps especially) under controlled, deliberative situations. The Actor fits the classic description of a model-free reinforcement learning system, and thus our framework says that there is an important model-free component to even high-level goal-driven and controlled behavior. This is consistent with a similar claim in the hierarchical model of Cushman and Morris (2015) and with their more recent experimental results (Cushman et al., this issue). Thus, whether one wants to call this Actor “model-free” or not, even when it receives all manner of

highly-processed goal, internal state, and prediction inputs from the cortex, further challenges the utility of this terminology. Furthermore, as we noted above, the availability of predicted outcome representations from the Predictor component can make the Actor's job very simple, and yet likely much more effective than a typical model-free system.

The Central Role of the Proposer

The function of the Proposer is particularly central to our overall framework, as it serves as the starting point for any action/plan initiation process. As noted, we think it functions through parallel, constraint-satisfaction processing to integrate a large number of different constraints, cues, and other contextual information to arrive at a plausible plan of action in a given situation (O'Reilly et al., 2014). It is precisely through this dynamic integration process that otherwise subtle, unconscious factors may be able to have measurable influences over our behavior (Bargh, 2006; Huang and Bargh, 2014). In addition, this property of the proposer enables even habit-based behavior to be somewhat flexible and capable of incorporating novel constraints from the current environmental state – even habitual actions are not purely ballistic and “robotic” in nature (Cushman and Morris, 2015; Hardwick et al., 2019).

Furthermore, as we'll see next, the incremental shaping of these Proposer representations over the course of learning plays a critical role in the automatization and habitization of behavior. Indeed, as the Proposer gets better and better at generating effective plans for increasingly well-known contexts, the Actor learns to essentially rubber-stamp these plans, thus resulting in fast, efficient habitual behavior. This happens through reinforcement learning shaping the weights from cortex to the basal ganglia Actor system; as the Actor sees more positive, rewarding examples, it becomes more biased toward a Go response. Along with its importance in habitual behavior, the Proposer component is also essential for coming up with plans in novel, challenging situations requiring controlled processing. Thus we argue that these functional distinctions may not have clear corresponding anatomical distinctions: the basal ganglia, Actor component is involved in all types of decisions, and that different areas of cortex may be recruited to play roles as Proposer, Predictor, and even to add more highly-processed inputs to model-free value predictions (Herd et al., 2019).

Transition From Slow and Controlled to Fast and Automatic Processing

One of the main results from our computational model (Herd et al., 2019) is shown in **Figure 3**, where the Proposer component gradually learned to choose a Plan appropriate for the current situation and goal. Initially, without relevant domain knowledge, the Proposer generates plans essentially at random, and a larger number of iterations are required to arrive at a Plan that the Actor approves of. Over the course of learning, the more appropriate initial plans generated by the Proposer reduces the number of iterations required, and thus the overall model gradually transitions from a more serial, iterative mode of processing to a more parallel mode of processing dominated

by the parallel constraint-satisfaction dynamic in generating plans in the Proposer system. This illustrates a continuum of habitization occurring over learning within the same overall system. Furthermore, the Proposer was able to learn only when the remaining systems chose to pursue a given plan; its learning was thus guided by the other systems, including the Predictor component.

Our initial model does not include the outer-loop ability to select which decision-making processes to engage in, so it did not have the ability to further optimize decision making by not engaging the Predictor at all, which would have resulted in even greater speedup, and corresponds with a more purely habitual response mode. We are currently working on a version of the model with this functionality.

Goal-Directed Behavior From a Model-Free System

In our model, the input to the Proposer system includes information about goals, so the behavior produced by this system qualifies as goal-directed, despite the relatively simple computations. Most computational work on model-free reinforcement learning systems addresses systems that do not include current goals as inputs. Those systems can only produce habitual behavior. However, there does not appear to be any strong justification for this assumption, and it seems more reasonable (as well as empirically justified) to assume that the relevant systems in the mammalian brain have access to a variety of useful information, including current goals. Indeed, there has been some discussion of goal-directed habits in other literature (Verplanken and Aarts, 1999; Aarts and Dijksterhuis, 2000).

When we assessed the accuracy with which the Proposer component produced a Plan which accomplished the current Goal (with Situation and Goal chosen at random from ten and four possibilities, respectively), we observed that this component displayed goal-directed behavior by matching Plans to Goals at an

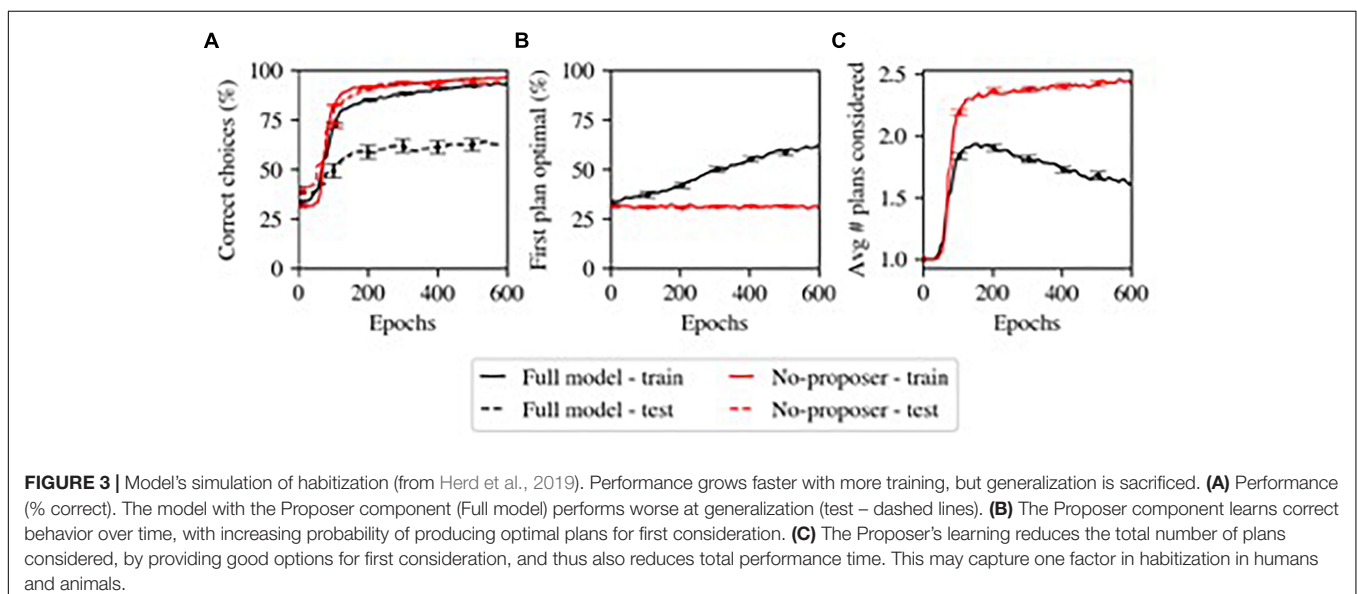
above-chance level, but learned slowly (Figure 3). This matches the slow transition from controlled to automatic processing (Gray et al., 1997) (note that we did not optimize parameters for Proposer learning in this task; some other parameterizations did produce better and faster learning).

Thus, our model illustrates one case in which goal-directed behavior results from thoroughly model-free computations.

Serial Processing Enables Coherent Predictions

A key advantage of the serial evaluation of different proposed plans in our model is that it allows many different brain areas to contribute to the evaluation process, without suffering from the *binding problem* that would otherwise arise from an attempt to evaluate multiple options in parallel. For example, if two options are considered together, and another brain area generates an activation associated with a prediction of difficulty, while another activates a prediction of relative ease, how do we know which prediction goes with which option?

This is analogous to the binding problem in visual search, where serial processing has also been implicated as a solution (Treisman and Gelade, 1980; Wolfe, 2003). For example, people cannot identify in parallel whether a display contains a particular *conjunction* of features (e.g., a red X among green Xs and red Os), whereas they can identify separable features in parallel (just Xs among Os, or just red among green). Likewise, the conjunction of options and their predicted consequences at many different levels in the brain, which likely depends on the current internal and external state, can be much more coherently evaluated by considering options one at a time. Furthermore, this serialization of the processing enables the *same* predictive and evaluative neural representations to be re-used across different situations and contexts, thus facilitating the transfer of knowledge to novel situations. In short, more complex model-based, predictive forms of control must involve serial processing mechanisms.



However, there are costs associated with serial processing, not only in terms of time, but also in terms of the coordination and control required to organize the serial processing itself. In addition, evaluating any one option relative to the predicted properties of other options requires some form of maintenance and comparison operations across these predictions, placing demands on working memory and other limited cognitive resources. Nevertheless, there are strong serial-order effects on decision-making, which such a serial model can naturally account for, so future modeling work will need to address these challenges in order to better address the complexities of these phenomena.

In summary, our sequential, integrated, systems-based approach provides some potentially novel perspectives on central questions about the nature of controlled/goal-driven vs. automatic/habitual behavior.

DISCUSSION

We have presented a computational systems-neuroscience approach to understanding the dynamics of decision making and action selection, which suggests that the classical dichotomy between habitual/automatic vs. goal-directed/controlled processing can be understood as different modes of functioning within a unitary system, operating fundamentally in a serial manner. The serial nature of the processing affords a natural incrementality to the continuum between these modes of processing – as the system iterates longer and engages more elaborated predictive and evaluative forms of processing, it shades more toward the goal-driven, controlled-processing end of the spectrum. By contrast, there is a fast track through the system where a proposed plan of action is derived rapidly through parallel constraint-satisfaction processing, which is then quickly approved by the basal ganglia Go/NoGo system – this corresponds to the habitual end of the spectrum. However, even this habitual level of behavior is contingent on an outer-loop of decision making that has established relevant thresholds and control parameters to enable the fast-track to be taken in the first place. Thus, habitual behavior still operates within an at-least minimally controlled context, in situations where the overall benefits of so behaving make sense compared to investing greater levels of control.

This framework contrasts with the dual-pathway model proposed by Miller et al. (2019) and similar models which suggest that habitual and controlled, goal-driven processing are subserved by parallel pathways that compete via an Arbiter system for control over behavior. It also contrasts with other models having a similar overall structure, but which use model-free and model-based components that likewise require an Arbiter system (e.g., Daw and Dayan, 2014). The framing of the interrelationship of habitual and controlled processing provided by Cushman and Morris (2015) is much more consistent with our framework, but further work is required to establish more detailed comparisons between their implemented models and our model. Likewise, the Pezzulo et al. (2013) model shares the central idea that model-based predictive mechanisms are only engaged when they yield additional value, and we will be working to relate

their computational-level mechanisms to the more biologically based framework we have developed.

Behaviorally, there are several important predictions that our model makes, which can be tested empirically. For example, consistent with a great deal of theory as well as folk psychology, we argue that habitual control is only enabled in either low-stakes or highly urgent situations. How does this outer loop of control interact with the various behavioral paradigms that have established the relative value-insensitivity of habitual behavior (Wood and R  nger, 2016)? Can our model account for both this value-insensitivity but also the cases where relevant expected reward values shift the system to more controlled, goal-driven behavior? What behavioral paradigms can effectively test such dynamics? One recent result provides a nice confirmation of one of our model's core predictions: that habitization is primarily about rapid activation of a good proposed plan of action (i.e., the Proposer in our model), but there remains a final "goal-directed" process (the Actor in our model) responsible for actual action initiation (Hardwick et al., 2019).

Another fertile ground for testing the model is in the domain of serial order effects on decision-making. For example, the balloon analog risk task (Lejuez et al., 2003; White et al., 2008; Van Ravenzwaaij et al., 2011; Fukunaga et al., 2012; Fairley et al., 2019) involves making a long sequence of decisions about whether to keep pumping a simulated balloon, or cash out with a potentially sub-optimal level of reward, and it seems uniquely capable of capturing real-world individual differences in propensity toward risky behaviors (Lejuez et al., 2003; White et al., 2008). Various sources of evidence suggest that there is something about the sequential nature of this task that is critical for its real-world validity. Thus, we are actively exploring this question in terms of the serial processing present in our model. In addition, there are other well-established serial-order effects in decision making, including framing effects (Tversky and Kahneman, 1981; De Martino et al., 2006), and the widely-studied foot-in-the-door and door-in-the-face strategies (Pascual and Gu  guen, 2005), which our serial model is particularly well-suited to explain.

AUTHOR CONTRIBUTIONS

All authors contributed to the development of the ideas and computational models described herein, and to the writing of the manuscript.

FUNDING

This work was supported by the ONR N00014-18-C-2067, N00014-18-1-2116, N00014-14-1-0670/N00014-16-1-2128, and NIH R01GM109996.

ACKNOWLEDGMENTS

We thank the other members of the CCN Lab for comments and discussion.

REFERENCES

- Aarts, H., and Dijksterhuis, A. (2000). Habits as knowledge structures: automaticity in goal-directed behavior. *J. Pers. Soc. Psychol.* 78:53. doi: 10.1037/0022-3514.78.1.53
- Alexander, G. E., DeLong, M. R., and Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.* 9, 357–381. doi: 10.1146/annurev.ne.09.030186.002041
- Balleine, B. W. (2005). Neural bases of food-seeking: affect, arousal and reward in corticostriatal limbic circuits. *Physiol. Behav.* 2005, 717–730. doi: 10.1016/j.physbeh.2005.08.061
- Balleine, B. W., Delgado, M. R., and Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* 27, 8161–8165. doi: 10.1523/jneurosci.1554-07.2007
- Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419. doi: 10.1016/s0028-3908(98)00033-1
- Bargh, J. A. (1989). Conditional automaticity: varieties of automatic influence in social perception and cognition. *Unintended thought* 3, 51–69.
- Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *Eur. J. Soc. Psychol.* 36, 147–168. doi: 10.1002/ejsp.336
- Berseth, G., Kyriazis, A., Zinin, I., Choi, W., and van de Panne, M. (2018). “Model-based action exploration for learning dynamic motion skills,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, (Piscataway, NJ: IEEE), 1540–1546.
- Cohen, J. D., Dunbar, K., and McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychol. Rev.* 97:332. doi: 10.1037/0033-295x.97.3.332
- Collins, A. G., and Frank, M. J. (2014). Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psycho. Rev.* 121:337. doi: 10.1037/a0037015
- Cushman, F., and Morris, A. (2015). Habitual control of goal selection in humans. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13817–13822. doi: 10.1073/pnas.1506367112
- Daw, N. D., and Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130478. doi: 10.1098/rstb.2013.0478
- Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 2005, 1704–1711. doi: 10.1038/nn1560
- Dayan, P. (1993). Improving generalization for temporal difference learning: the successor representation. *Neural Comput.* 5, 613–624. doi: 10.1162/neco.1993.5.4.613
- De Martino, B., Kumaran, D., Seymour, B., and Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science* 313, 684–687. doi: 10.1126/science.1128356
- Dickinson, A., and Balleine, B. (2002). “The role of learning in the operation of motivational systems,” in *Learning, Motivation, and Emotion*, eds H. Pashler, and C. R. Gallistel, (New York, NY: John Wiley & Sons), 497–533.
- Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Netw.* 12, 961–974. doi: 10.1016/s0893-6080(99)00046-5
- Doya, K., Samejima, K., Katagiri, K., and Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural Comput.* 14, 1347–1369. doi: 10.1162/089976602753712972
- Fairley, K., Parelman, J. M., Jones, M., and Carter, R. M. (2019). Risky health choices and the Balloon Economic Risk Protocol. *J. Econ. Psychol.* 73, 15–33. doi: 10.1016/j.joep.2019.04.005
- Frank, M. J., and O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. *Behav. Neurosci.* 120:497. doi: 10.1037/0735-7044.120.3.497
- Fukunaga, R., Brown, J. W., and Bogg, T. (2012). Decision making in the balloon analogue risk task (BART): anterior cingulate cortex signals loss aversion but not the infrequency of risky choices. *Cogn. Affect. Behav. Neurosci.* 12, 479–490. doi: 10.3758/s13415-012-0102-1
- Gershman, S. J. (2018). The successor representation: its computational logic and neural substrates. *J. Neurosci.* 38, 7193–7200. doi: 10.1523/JNEUROSCI.0151-18.2018
- Gray, J. A., Buhusi, C. V., and Schmajuk, N. (1997). The transition from automatic to controlled processing. *Neural Netw.* 10, 1257–1268. doi: 10.1016/s0893-6080(97)00058-0
- Haber, S. N. (2010). “Integrative networks across basal ganglia circuits,” in *Handbook of Behavioral Neuroscience*, ed M. S. Gazzaniga, (Amsterdam: Elsevier), Vol. 24, 535–552. doi: 10.1016/b978-0-12-802206-1.00027-1
- Haber, S. N. (2017). *Anatomy and Connectivity of the Reward Circuit*. In *Decision Neuroscience*, eds J.-C. Dreher and L. Tremblay (Cambridge, MA: Academic Press), 3–19.
- Hardwick, R. M., Forrence, A. D., Krakauer, J. W., and Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. *Nat. Hum. Behav.* 3, 1252–1262. doi: 10.1038/s41562-019-0725-0
- Hassin, R. R., Bargh, J. A., and Zimerman, S. (2009). Automatic and flexible: the case of nonconscious goal pursuit. *Soc. Cogn.* 27, 20–36. doi: 10.1521/soco.2009.27.1.20
- Herd, S. A., Banich, M. T., and O'Reilly, R. C. (2006). Neural mechanisms of cognitive control: an integrative model of Stroop task performance and fMRI data. *J. Cogn. Neurosci.* 18, 22–32. doi: 10.1162/089892906775250012
- Herd, S. A., Krueger, K. A., Nair, A., and Mollick, J. A. (2019). Neural mechanisms of human decision-making. *Under Revis. Cogn. Affect. & Behav. Neurosci.* (inpress).
- Hikosaka, O., and Isoda, M. (2010). Switching from automatic to controlled behavior: cortico-basal ganglia mechanisms. *Trends Cogn. Sci.* 14, 154–161. doi: 10.1016/j.tics.2010.01.006
- Huang, J. Y., and Bargh, J. A. (2014). The selfish goal: autonomously operating motivational structures as the proximate cause of human judgment and behavior. *Behav. Brain Sci.* 37, 121–135. doi: 10.1017/S0140525X13000290
- Hull, C. L. (1943). *Principles of Behavior: An Introduction to Behavior Theory*. New York, NY: Appleton-Century.
- James, W. (1890). *The principles of Psychology*. New York, NY: Henry Holt and Company.
- Killcross, S., and Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex.* 13, 400–408. doi: 10.1093/cercor/13.4.400
- Kool, W., Cushman, F. A., and Gershman, S. J. (2018). “Competition and cooperation between multiple reinforcement learning systems,” in *Goal-Directed Decision Making*, eds R. W. Morris, A. Bornstein, and A. Shenhav, (Cambridge, MA: Academic Press), 153–178. doi: 10.1016/b978-0-12-812098-9.00007-3
- Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., and Pedulla, C. M. (2003). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *J. Adolesc.* 26, 475–479. doi: 10.1016/s0140-1971(03)00036-8
- Littman, M. L., and Sutton, R. S. (2002). Predictive representations of state. *Adv. Neural Inform. Process. Syst.* 1555–1561.
- Melnikoff, D. E., and Bargh, J. A. (2018). The mythical number two. *Trends Cogn. Sci.* 22, 280–293. doi: 10.1016/j.tics.2018.02.001
- Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167
- Miller, K. J., Ludvig, E. A., Pezzulo, G., and Shenhav, A. (2018). “Realizing models of habitual and goal-directed decision-making,” in *Goal-Directed Decision Making*, eds R. W. Morris, A. Bornstein, and A. Shenhav, (Cambridge, MA: Academic Press), 407–428. doi: 10.1016/b978-0-12-812098-9.00018-8
- Miller, K. J., Shenhav, A., and Ludvig, E. A. (2019). Habits without values. *Psychol. Rev.* 126, 292–311. doi: 10.1037/rev0000120
- Miller, R. (1981). *Meaning and Purpose in the Intact Brain: a Philosophical, Psychological, and Biological Account of Conscious Processes*. Oxford: Oxford University Press.
- Mishkin, M., Malamut, B., and Bachevalier, J. (1984). Memories and habits: two neural systems. *Neurobiol. Learn. Mem.* 65–77. doi: 10.1523/JNEUROSCI.0840-16.2016
- Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., and Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nat. Hum. Behav.* 1, 680. doi: 10.1038/s41562-017-0180-8
- Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 16, 1936–1947. doi: 10.1523/jneurosci.16-05-01936.1996

- O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends Cogn. Sci.* 2, 455–462. doi: 10.1016/s1364-6613(98)01241-8
- O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition. *Science* 314, 91–94. doi: 10.1126/science.1127242
- O'Reilly, R. C., and Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput.* 18, 283–328. doi: 10.1162/089976606775093909
- O'Reilly, R. C., Hazy, T. E., and Herd, S. A. (2016). The Leabra cognitive architecture: how to play 20 principles with nature and win! *Oxford Handb. Cogn. Sci.* 91, 91–116.
- O'Reilly, R. C., Hazy, T. E., Mollick, J., Mackie, P., and Herd, S. (2014). Goal-driven cognition in the brain: a computational framework. *arXiv preprint arXiv* :[preprint]. Available at: <https://arxiv.org/abs/1404.7591> (accessed February 18, 2020).
- O'Reilly, R. C., and Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, MA: MIT Press.
- Packard, M. G., and Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.* 25, 563–593.
- Pascual, A., and Guéguen, N. (2005). Foot-in-the-door and door-in-the-face: a comparative meta-analytic study. *Psychol. Rep.* 96, 122–128. doi: 10.2466/pr0.96.1.122-128
- Pasupathy, A., and Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433:873. doi: 10.1038/nature03287
- Pauli, W. M., Clark, A. D., Guenther, H. J., O'Reilly, R. C., and Rudy, J. W. (2012). Inhibiting PKM ζ reveals dorsal lateral and dorsal medial striatum store the different memories needed to support adaptive behavior. *Learn. Mem.* 19, 307–314. doi: 10.1101/lm.025148.111
- Pezzulo, G., Rigoli, F., and Chersi, F. (2013). The mixed instrumental controller: using value of information to combine habitual choice and mental simulation. *Front. Psychol.* 4:92. doi: 10.3389/fpsyg.2013.00092
- Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., and Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Computat. Biol.* 13:e1005768. doi: 10.1371/journal.pcbi.1005768
- Sallet, J., Mars, R. B., Noonan, M. P., Neubert, F. X., Jbabdi, S., O'Reilly, J. X., et al. (2013). The organization of dorsal frontal cortex in humans and macaques. *J. Neurosci.* 33, 12255–12274. doi: 10.1523/JNEUROSCI.5108-12.2013
- Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science* 310, 1337–1340. doi: 10.1126/science.1115270
- Schultz, W. (2013). Updating dopamine reward signals. *Curr. Opin. Neurobiol.* 23, 229–238. doi: 10.1016/j.conb.2012.11.012
- Seger, C. A., and Spiering, B. J. (2011). A critical review of habit learning and the basal ganglia. *Front. Syst. Neurosci.* 5:66. doi: 10.3389/fnsys.2011.00066
- Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychol. Rev.* 84:127. doi: 10.1037/0033-295x.84.2.127
- Squire, L. R. (1992). Declarative and nondeclarative memory: multiple brain systems supporting learning and memory. *J. Cogn. Neurosci.* 4, 232–243. doi: 10.1162/jocn.1992.4.3.232
- Squire, L. R., and Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science* 253, 1380–1386. doi: 10.1126/science.1896849
- Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.
- Thorndike, E. L. (1911). *Animal Intelligence: Experimental Studies*. New York, NY: Macmillan.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychol. Rev.* 55:189. doi: 10.1037/h0061626
- Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5
- Tricomi, E., Balleine, B. W., and O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29, 2225–2232. doi: 10.1111/j.1460-9568.2009.06796.x
- Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* 211, 453–458. doi: 10.1126/science.7455683
- Van Ravenzwaaij, D., Dutilh, G., and Wagenmakers, E. J. (2011). Cognitive model decomposition of the BART: assessment and application. *J. Math. Psychol.* 55, 94–105. doi: 10.1016/j.jmp.2010.08.010
- Verplanken, B., and Aarts, H. (1999). Habit, attitude, and planned behaviour: is habit an empty construct or an interesting case of goal-directed automaticity? *Eur. Rev. Soc. Psychol.* 10, 101–134. doi: 10.1080/14792779943000035
- White, T. L., Lejuez, C. W., and de Wit, H. (2008). Test-retest characteristics of the Balloon Analogue Risk Task (BART). *Exp. Clin. Psychopharmacol.* 16:565. doi: 10.1037/a0014083
- Wickens, J. R., Horvitz, J. C., Costa, R. M., and Killcross, S. (2007). Dopaminergic mechanisms in actions and habits. *J. Neurosci.* 27, 8181–8183. doi: 10.1523/jneurosci.1671-07.2007
- Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends Cogn. Sci.* 7, 70–76. doi: 10.1016/s1364-6613(02)00024-4
- Wood, W., and Rünger, D. (2016). Psychology of habit. *Annu. Rev. Psychol.* 67, 289–314.
- Yin, H. H., and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7, 464–476. doi: 10.1038/nrn1919
- Yin, H. H., Ostlund, S. B., Knowlton, B. J., and Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.* 22, 513–523. doi: 10.1111/j.1460-9568.2005.04218.x

Conflict of Interest: RO'R, AN, and SH were employed by company eCortex, Inc.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 O'Reilly, Nair, Russin and Herd. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership