HIGH PERFORMANCE COGNITION: INFORMATION-PROCESSING IN COMPLEX SKILLS, EXPERT PERFORMANCE, AND FLOW

EDITED BY: Benjamin Cowley, Frederic Dehais, Stephen Fairclough, Alexander John Karran, Otto Lappi and Jussi Palomäki
PUBLISHED IN: Frontiers in Psychology and Frontiers in Human Neuroscience

frontiers Research Topics

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

HIGH PERFORMANCE COGNITION: INFORMATION-PROCESSING IN COMPLEX SKILLS, EXPERT PERFORMANCE, AND FLOW

Topic Editors:
Benjamin Cowley, University of Helsinki, Finland
Frederic Dehais, Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-SUPAERO), France
Stephen Fairclough, Liverpool John Moores University, United Kingdom
Alexander John Karran, Université de Montréal, Canada
Otto Lappi, University of Helsinki, Finland
Jussi Palomäki, University of Helsinki, Finland

Table of Contents

Check for updates

Editorial: High Performance Cognition: Information-Processing in Complex Skills, Expert Performance, and Flow

Benjamin Ultan Cowley [1,2]*, Frederic Dehais [3], Stephen Fairclough [4], Alexander John Karran [5], Jussi Palomäki [2] and Otto Lappi [2]

[1] Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland, [2] Department of Digital Humanities, Faculty of Arts, University of Helsinki, Helsinki, Finland, [3] Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-SUPAERO), Toulouse, France, [4] Liverpool John Moores University, Liverpool, United Kingdom, [5] Hautes études Commerciales Montréal, Université de Montréal, Montreal, QC, Canada

Editorial on the Research Topic

High Performance Cognition: Information-Processing in Complex Skills, Expert Performance, and Flow

In this Research Topic, High Performance Cognition (HPC) emerges as a generic term for the study of human performance and skill acquisition, from novices on an unfamiliar task (as most psychological experiments) to experts displaying superior levels of domain-specific skill. We received 11 pieces of empirical work across several experimental tasks and three articles describing theoretical/conceptual work on flow, cognitive training, and Deliberate Practice framework for expert performance. The papers received in this Research Topic encompassed a range of methodologies including reports of phenomenology, measurement of behavior, and psychological traits, and measures based on psychophysiology and neurophysiology.

What were the commonalities underlying the various contributions received for the topic? Most empirical papers report that the cognitive process under investigation showed interdepencies with other processes: from the longitudinal relationship between performance anticipation and Flow shown in Cowley et al., to the proactive control of executive function shown by "open"-skill athletes (i.e. trained in dynamic, externally-paced sports) in Yu et al.. This theme also reflects in theoretical contributions which call for revision of the longstanding constructs of deliberate practice and Flow, to either account for interdependency in a more holistic framework (Hambrick et al.), or control for it (Abuhamdeh), respectively. Clearly, even to make such revisions is challenging; beyond which is the yet more-challenging task to build a unified framework on the revised constructs, coherent with existing evidence. Yet such a framework will refine the generic term HPC to a definition of cognitive performance, and thereby provide grounds for empirical predictions and a direction for future work for many years.

Thus, while we cannot yet answer the question: "do cognitive processes that are specific to contexts of high performance also generalize across task domains?," this Topic clarifies somewhat how we might start to address the question. A summary of the contributions follows to provide an overview of the specific articles received as part of this Research Topic.

EMPIRICAL STUDIES

Flow Focus

Chin and Kales investigate whether autonomic arousal (heart-rate variability, HRV) influences self-reported flow and performance in the Stroop task. To increase between-subjects HRV, participants first performed techniques in either nasal respiration, arm-muscle contraction, both combined, or read emotionally-neutral articles. The results showed an inverted-U pattern: performance and flow were maximal at moderate arousal levels. Optimal performance was also associated with predominantly sympathetic autonomic activity.

Cowley et al. ask: How does subjective feeling of high performance relate to development of (personally) optimal performance? They report a longitudinal (40 trials over 8 sessions) study, recording Flow and behavior as participants learned to perform a visuomotor steering task. Participants did not experience more Flow over sessions; instead, their trial-wise subjective anticipation of performance (estimated from a power-law model over trials) was strongly related to Flow.

Murch et al. report on three experiments which aim to identify objective markers of flow while gambling. The authors collected self-reported gambling flow and physiological data as measured by the cardiac pre-ejection period (PEP). The authors did not find evidence of changes in PEP when interacting with the electronic gambling machines but only interactions between subjective and objective measures during the first block of each experiment.

In a study examining the effect of positive feedback on perceived self-efficacy, flow and performance, Peifer et al. hypothesize that self-efficacy acts as a mediator, whereby positive feedback on one task imparts a positive effect upon performance and flow in subsequent tasks. They report evidence to support the hypothesis and propose multiple methods of intervention to further increase these positive effects.

Sinnett et al. ask: can increased flow experience can improve perception? The authors implemented a within-subjects design involving groups of athletes and musicians facing a temporal order judgement task paradigm. This task encompasses a measure of temporal processing measure and of spatial attention. Their results disclosed a positive relationship between the self-reported value of experienced flow and the efficiency of spatiotemporal information processing.

Cognitive-Performance Focus

Here we order the studies by participant expertise, starting with naïve-subject designs.

Kee et al. investigate the effects of non-striving states, induced by a repetitive water-pouring task, on a word-length comparison task. The authors reported that the experimental group tended to exhibit lower performance and was significantly faster to perform the word comparison task than the control group. Though the authors did not use self-report questionnaires, they suggest their experimental design could help to improve mindfulness practice.

Prolonged task execution brings mental fatigue, characterized by "indolence, reduced motivation, and impaired performance." Liu et al. investigate the motivating effect of monetary rewards (given during this low-vigilance state) on performance in a flanker task, and concomitant electrophysiological measures linked to selective visual attention, primarily P300. Monetary reward did improve performance during low vigilance, but neural measures recovered selectively, indicating that some constraints on higher cognitive performance are non-volitional.

Gong et al. investigate how EEG neurofeedback training (NFT) affects shooting performance and neuroplasticity among police students. They found that pre-post shooting performance increased for participants using sensorimotor-rhythm NFT, but slightly decreased for participants using alpha-rhythm NFT. Participants learned to alter their EEG patterns in both NFT groups. Pre-post measures also showed that neuroplasticity was affected by NFT. Thus, NFT may facilitate training for activities requiring sensorimotor accuracy.

Muñoz et al. present work toward a psychophysiological model of firearm training for real police officers using a combination of head-mounted virtual reality and non-intrusive physiological sensors. The authors demonstrate how changes in frontal theta activation and heart rate variability were affected by changes in task difficulty levels, which they articulate as metrics of "concentration" and "calmness."

The study by Savostyanov et al. utilizes ERPs to examine the effect of long-term meditation on comprehension of statements about oneself and others. Three groups (control, short- and long-term meditators) completed a reading task containing self or non-self-related evocative statements and grammatical errors. They demonstrate that meditation increases negative affect processing and an ERP analysis that potentially shows an increase of voluntary control over emotional states for meditators.

Work presented by Yu et al. investigates how motor-skills experience modulates proactive and reactive control of executive function, using a cued task-switching protocol with "open"- and "closed"-skilled athletes and non-athlete controls. Their findings highlight that open-skilled participants showed significantly less positive-going parietal cue-locked P3 and better predictive task-switching performance. They conclude that proactive control may be enhanced in open-skilled participants when compared to the other groups.

THEORETICAL PAPERS

Aidman proposes a framework for Cognitive Fitness: trainable cognitive abilities (as distinct from domain-specific knowledge, skills and attitudes) that can be improved and maintained by specific cognitive fitness training, analogous to physical-fitness training for strength, endurance, and flexibility. Aidman proposes his framework would be valuable for training performance in sport, arts, emergency services—indeed, any domain where individuals must deliver high performance on demand under stressful conditions.

Hambrick et al. and Abuhamdeh present critical opinion pieces on two highly-influential theoretical constructs in study of HPC: Deliberate Practice (DP, Ericsson et al., 1993), and Flow (Csikszentmihalyi, 1975). The Hambrick et al. paper

continues their ongoing interrogation of the internal consistency of existing DP research, building on the authors' earlier meta-analysis (Macnamara et al., 2014). The Abuhamdeh paper looks critically at how the original definition of flow has since been operationally defined.

Both contributions share an underlying message: after decades of research, both lines of investigation remain pre-paradigmatic, in the sense of Kuhn (1962). Each is identified with one researcher's seminal account, but subsequently has been operationalized in an unsystematic way. This inconsistency, alongside lack of agreement on theoretical attributes, threatens the ability of both theoretical perspectives to develop into mature paradigms. Both papers challenge researchers to unify methodology and theory, and develop "progressive research programs" (Lakatos, 1978).

SUMMARY

The studies presented in this Research Topic illustrate both current diversity, and a bright future, for research in performance cognition. Researchers investigating flow states posit new experiments to model the conditions and relationships involved and call for a consistent operationalization of the flow state. Neurofeedback researchers propose to move into real-time adaptation of training protocols; those investigating meditation propose novel interventions to further increase the benefits of the practice. Extension and improvement of human cognition through augmented and directed training comprises a common vision for the future. To achieve it, we must unify this diverse field around a common and parsimonious framework, such that we can move toward measuring and understanding high performance cognition as it develops through learning and training across various domains.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*. San Francisco, CA: Jossey-Bass.

Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychol. Rev.* 100, 363–406.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Lakatos, I. (1978). *The Methodology of Scientific Research Programmes: Volume 1: Philosophical Papers*. Cambridge: Cambridge University Press.

Macnamara, B. N., Hambrick, D. Z., and Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, professions, and education: a meta-analysis. *Psychol. Sci.* 25, 1608–1618. doi: 10.1177/0956797614535810

Flow Experiences During Visuomotor Skill Acquisition Reflect Deviation From a Power-Law Learning Curve, but Not Overall Level of Skill

Benjamin Ultan Cowley [1,2*†], Jussi Palomäki [1,2†], Tuisku Tammi [1,3†], Roosa Frantsi [1,3], Ville-Pekka Inkilä [1], Noora Lehtonen [1], Pasi Pölönen [1], Juha Vepsäläinen [1] and Otto Lappi [1,2,3]

[1] Cognitive Science, Department of Digital Humanities, Faculty of Arts, University of Helsinki, Helsinki, Finland, [2] Helsinki Centre for Digital Humanities, Helsinki, Finland, [3] Traffic Research Unit, TRUlab, University of Helsinki, Helsinki, Finland

Flow is a state of "optimal experience" that arises when skill and task demands match. Flow has been well studied in psychology using a range of self-report and experimental methods; with most research typically focusing on how Flow is elicited by a particular task. Here, we focus on how the experience of Flow changes during task skill development. We present a longitudinal experimental study of learning, wherein participants ($N = 9$) play a novel steering-game task designed to elicit Flow by matching skill and demand, and providing clear goals and feedback. Experimental design involves extensive in-depth measurement of behavior, physiology, and Flow self-reports over 2 weeks of 40 game trials in eight sessions. Here we report behavioral results, which are both strikingly similar and strong within each participant. We find that the game induces a near-constant state of elevated Flow. We further find that the variation in Flow across all trials is less affected by overall performance improvement than by deviation of performance from the expected value predicted by a power law model of learning.

Keywords: Flow, skill acquisition, power law of practice, visuomotor performance, steering, high performance cognition

1. INTRODUCTION

In many fields of human endeavor—such as music, art and sports—the skilful performance of a demanding task can elicit a state of "optimal experience" called Flow (Csikszentmihalyi, 1975). The Flow state is thought to be dependent on several pre-conditions for the eliciting task, and characterized by several phenomenological features (Nakamura and Csikszentmihalyi, 2002; Engeser and Schiepe-Tiska, 2012; Keller and Landhäußer, 2012). *Conditions* for Flow to occur define certain characteristics of the task and one's skill in the task: C1: challenge should match skill in a demanding task. C2: the task setting should present clear and personally significant goals. And C3 the setting should provide unambiguous feedback on goal achievement. When these conditions are met and the individual enters a mode of high performance, they may experience a set of phenomenological features characterizing "the Flow experience": F1 total focus in the present moment, and concentration on what one is doing; F2 merging of action and awareness ("being one with the task"); F3 loss of reflective self-consciousness, a sense of effortlessness; F4 a sense of personal control and confidence in one's skill; F5 positive affect, the activity is experienced as

highly enjoyable; F6 a distortion of temporal experience (time may seem to go slower or faster than normal). These *consistently co-occurring* features of the experience imply that Flow is *autotelic*, i.e., Flow-producing activities are intrinsically rewarding, people want to do them for their own sake regardless of external reward.

The antecedent conditions (C1-3) and phenomenological features (F1-6) of Flow have been investigated for several decades, mainly using analysis of self-report data from people engaged in natural everyday or expert performance (Csikszentmihalyi and Bennett, 1971; Moneta, 2012). Despite this, debate continues around the precise definition of the pre-conditions, especially C1 (challenge–skill balance).

At least three different models have described how Flow depends on challenge–skill ratio. Each model makes different assumptions for how this dependence is affected when skill and challenge change during learning. The original Flow model (Csikszentmihalyi, 1975) assumed that levels of challenge and skill can vary independently from low to high, and that Flow can happen when skill and challenge match at any level. This is the seminal Flow "channel" model.

The now-classic octant/quadrant models (Massimini et al., 1988) instead suggested that Flow only happens when skill and challenge exceed a certain threshold level. If the challenge or skill of task performance are too low, even if matched, then the task will not elicit Flow. At a certain point, the level of matched skills and challenges will become enough to elicit Flow; we will refer to this point as the *reference level*. The octant/quadrant models have been criticized because it is unclear how to determine the reference level. Does challenge need to exceed the challenge of most typical everyday tasks? Does the task need to be challenging for the individual, or relative to typical skills of a reference population? Or, does the reference level come from some universal benchmark of physical effort or information-processing complexity[1]? Challenge (and skill) could also be task-specific, with its reference level continually recalibrating to the reference level of past performance in the specific task. Each of these issues beg the question: what happens during task learning? Does the reference level need to be recalibrated to track the levels of skill and challenge as they increase?

In their critical examination of these models, Keller and Landhäußer (2012) argue that the concept of reference level is problematic, as it might not be empirically determinable for researchers, or even psychologically available to performing individuals (see also Moneta, 2012). Keller and Landhäußer (2012, p. 56) proposed the Flow intensity model, which has the dimensions "perceived challenge–skill balance" and "subjective value of the task" to define conditions that elicit Flow. This model has no reference levels. On the other hand, the intensity model takes no account of the *direction* of a challenge–skill imbalance, thus losing information which may be important to understand the Flow experience.

To sum up, the state of the art is unclear on three main points. First, it is unclear whether reference levels of challenge

and/or skill govern the emergence of Flow. If they do, it is unclear how the levels are defined. If defined relative to a given task or task episode, we must ask: should the level change with skill acquisition? If defined relative to other tasks, we must ask: should the task demand be compared to the skill sets of the individual, or a reference population, or an absolute standard? Second, it is unclear if the direction of challenge–skill ratio is important, i.e., does it make a difference that skill exceeds challenge vs. challenge exceeds skill. Third and finally, these models capture a static snapshot of Flow. Thus, Flow research must still deal with the effects of *learning* on C1 (challenge–skill balance).

The matter is of importance because understanding how the Flow conditions behave across different levels of skill is relevant to any field interested in the development of performance (e.g., development of coaching practices or concentration techniques in sport Jackson and Marsh, 1996). Also, an understanding of the mechanisms that mediate Flow across a learning process could help to enhance enjoyment or performance through better design, e.g., of recreational tasks such as games (Chen, 2007). However, these aims call for studies of Flow elicited across different stages of learning, with a more controlled and quantitative approach. Such an approach can build on recent studies of Flow from fields of experimental psychology (Keller and Blomann, 2008; Harris et al., 2017) and psychophysiology (Peifer, 2012; Peifer et al., 2014; Harmat et al., 2015; Wolf et al., 2015; Labonté-LeMoyne et al., 2016).

Here we report an experimental skill-acquisition study on the connections between performance and the self-reported phenomenology of Flow. We introduce a novel, demanding visuomotor task. With a longitudinal design, we are afforded more power to examine the connections of Flow and performance, within- and between-subjects. We also model the learning shown by participants, finding good fit of the data to a power-law curve, which has been shown to closely approximate a very wide range of skill acquisition datasets (Newell and Rosenbloom, 1982; Logan, 1988; Palmeri, 1999).

Results show that our task successfully elicits an elevated level of Flow across all self-reports. Yet when we examine Flow with fine granularity, we see that variation in Flow responses relates less to overall performance improvement, than it does to deviation of performance from the expected value predicted by a power law model of learning, with higher performance associated with higher Flow and lower performance with lower Flow.

1.1. Protocol and Research Questions

Participants learned to play a custom-made high-speed steering game (**Figure 1**[2]). The game was specifically designed to elicit Flow through balancing task demand with the skill level of the participant, and providing clear immediate feedback. The aim in the game was to steer a blue cube through a course with randomly placed red obstacles at the highest possible speed. The cube started each game at a fixed forward velocity, which increased at a constant rate. The lateral position of the cube was controlled by the steering wheel. Collision with obstacles reduced speed by

---

[1]As an example of a universal benchmark, the maximum propagation speed of neural signals along an axon sets a limit on response times.

[2]For game video see https://doi.org/10.6084/m9.figshare.7269395.v1
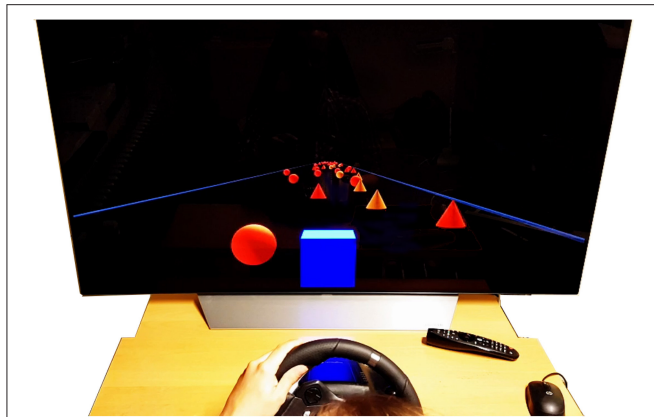
FIGURE 1 | The high-speed steering task. The participant steers the blue cube to avoid conical/spherical obstacles on the track, which is bounded to each side by dark blue parallel lines. The game was designed to continually adapt the difficulty level (speed) to the participant's skill (obstacle collisions). Such balance is considered one of the key antecedents of Flow.

a fixed amount, also indicated by a flashing of the screen (see section 2 for units).

This design, inspired by psychophysical staircase methods (Cornsweet, 1962), ensured constant match between skill and demand at the participant's level of performance. The performance was measured by duration of the trial (shorter duration = faster average speed = better), displayed as a time score at the end of each trial.

Participants played the game for forty trials across eight sessions, over a period of 2–3 weeks, which was sufficient to achieve good proficiency in this task with no ceiling effect. The 10 item Flow Short Scale (Engeser and Rheinberg, 2008) was filled after each trial to probe self-reported Flow in the task. Physiological data were recorded (skin conductance, heart rate, and eye tracking), during the task and a 5 min baseline, in sessions one and five-to-eight. This data introduces considerable further research questions and so falls outside the scope of this report.

This design allowed us to explore the following Research Questions:

1. RQ1. How does performance change over time, i.e., what is the shape of the "learning curve" (LC)? Specifically, does performance in the game improve, and does improvement follow a power law of practice (as found in much previous work on visuomotor skill acquisition Newell and Rosenbloom, 1982)?
2. RQ2. How is trial-wise self-reported Flow (from here on simply "Flow") related to performance? Specifically, is performance improvement across the whole experiment (i.e., learning) accompanied by higher levels of Flow?

2. METHODS
2.1. Participants
A convenience sample ($N = 9$, 6 males, 3 females) was recruited via student mailing lists at the University of Helsinki. The

participants were between 22 and 38 years of age (mean 27, SD 3) with normal or corrected-to-normal visual acuity and no history of neurological or psychiatric disease.

Eight of the participants had a driving license; two participants reported <10,000 km lifetime kilometrage, three participants 10,000–30,000 km, two 30,000–100,000 km, and one participant >100,000 km. Two had no or very little previous gaming experience, two participants played 1–3 h a month, and five participants stated they play over 1 h a week. **Table 1** shows the details of each participant.

All participants were naive about the specific hypotheses and purpose of the study, other than that the time of recruiting they were informed that the experiment was about game experience and learning. Participants were given 11 cultural vouchers (1 voucher is worth 5 euro) in compensation for their time. They were told that they would get 9 vouchers for participating in all sessions and 2 extra vouchers if they improved their performance in the game. The criteria for sufficient improvement were not stated explicitly, and in fact all participants were given the two extra vouchers.

Participants were briefed and provided written informed consent before entering the study, and were aware of their legal rights. The study followed guidelines of the Declaration of Helsinki and was approved by the University of Helsinki Ethical review board in humanities and social and behavioral sciences (statement 31/2017; study title MulSimCoLab).

2.2. Design
The experiment was divided into eight sessions, on eight different days over a period of 2–3 weeks scheduled at each participant's convenience. In each session, the participant played five trials of the driving game, each trial lasting 2–4 min depending on their performance, for approximately 15 min of driving time per session. The judgement of how much total playtime (here, 2 h) would be sufficient to develop good task proficiency was based on extensive informal piloting, including prior observations with other convenience samples. **Figure 2** illustrates the protocol.

After each trial, the participant was shown the trial duration and the number of collisions, after which they filled in a self-report questionnaire (FSS). In sessions 1 and 5–8 (lasting approx. an hour), eye-tracking and physiological signals (electrodermal activity and heart rate) were recorded in a 5 min baseline recording before playing, and during gameplay. In sessions 2–4 (lasting 20–30 min), no physiological measurements were taken.

2.3. Materials
2.3.1. Game
The experimental task was a custom-made high-speed steering game *CogCarSim* designed specifically for the study of Flow and coded in Python[3].

The participant steered a cube "avatar" moving forward along a straight track bounded by edges that could not be crossed. The cube's side length was 2 units, and the track was 25 units wide. The horizontal field of view angle of the virtual

---

[3]The game code as used herein is permanently available under open source license at https://doi.org/10.6084/m9.figshare.7269467

Participant	Gender	Age	Driving license	Driving experience (km)	Gaming experience
1	M	≤ 27	Yes	1,000–10,000	At least 1 h a week
2	M	≤ 27	Yes	10,000–30,000	At least 1 h a week
3	F	≤ 27	No	0–1,000	None or very little
4	F	≥ 28	Yes	0–1,000	At least 1 h a week
5	M	≤ 27	Yes	30,000–100,000	At least 1 h a week
6	M	≤ 27	Yes	30,000–100,000	1–3 h a month
7	F	≥ 28	Yes	10,000–30,000	None or very little
8	M	≥ 28	Yes	100,000+	1–3 h a month
9	M	≤ 27	Yes	10,000–30,000	At least 1 h a week

FIGURE 2 | The game was played in eight sessions on eight different days. In sessions 1 and 5–8, physiological signals were recorded during task performance; in sessions 2–4(*) no physiology was recorded. Each session consisted of five trials (2–4 min) followed by a self-report questionnaire (FSS, Flow Short Scale) about the latest trial.
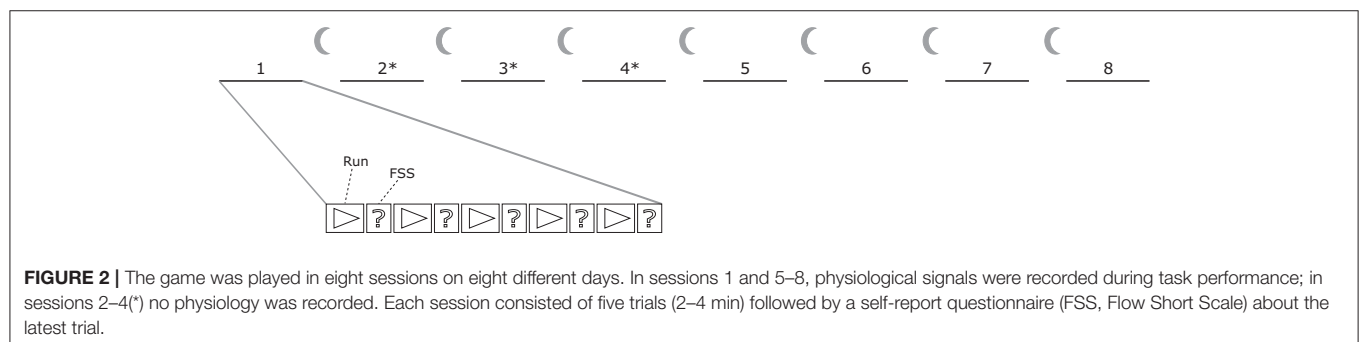
camera was 60° and vertical 32°. The camera was positioned behind the cube at 4 units height, pointing forward along the track.

Stationary obstacles (red cones, red or yellow spheres with a height/diameter of 2 units) on the track had to be avoided. For each trial, a total of 2,000 obstacles were placed randomly on the track, with placement constrained to always allow a path through. Track length varied between 24196.4 and 24199.7 units (mean 24197.8, sd 0.8). The speed of the cube was initially set to 1.6 units per step (96 units per second); increased at a constant rate (0.0012 units/step at every step); and slowed down if obstacles were hit (0.102 units/step at each collision). When a collision caused a speed drop, the screen flashed to indicate a collision; there followed an immunity period of 100 steps during which additional collisions did not cause further speed drops. Participants could only affect speed indirectly, by avoiding collisions. Participants were instructed to avoid as many obstacles they could in order to complete the trial as fast as possible.

The game had maximally simple one degree-of-freedom linear and holonomic dynamics: the horizontal position of the cube was directly proportional to steering wheel angle. Extensive self-piloting was done to adjust the graphics, e.g., virtual eye height; plus starting and increment speeds, rate of change of speed during collisions, and steering wheel sensitivity (steering ratio and damping).

The participants started each trial by pressing a button on the steering wheel when they felt ready. At the end of each trial, the elapsed time and number of collisions were displayed, along with a high score of the participant's ten best trials so far.

Data collected by CogCarSim included the positions, shape, and color of obstacles on the track; trial-level aggregated performance data (trial duration, number of collisions, average velocity); and within-trial time series data (steering wheel and cube position, speed, registered collisions).

2.3.2. Equipment

The game was run on a Corsair Anne Bonny with Intel i7 7700k processor and an Nvidia GTX 1080 graphics card, running Windows 10.

The participant was seated in a Playseat Evolution Alcantara playseat (Playseats B.V., The Netherlands) aligned with the mid point of the 55″ display screen (LG 55UF85). The screen resolution was 1,920 × 1,080 pixels, the frame rate was 60 and the refresh rate 60 Hz. The viewing distance was adjusted for each participant (so that they could place their hands on the steering wheel comfortably) and was approximately between 90 and 120 cm from the eye to the screen. The game was controlled with a Logitech G920 Driving Force steering wheel (Logitech, Fremont, CA). Steering wheel settings in Logitech Gaming Software 8.96.88 were: sensitivity 100%, centering spring strength 4 percent, and wheel operating range 900°.

Eye-tracking and physiological signals were collected and stored on an Asus UX303L laptop with Debian GNU/Linux 9 OS. Electrodermal activity (EDA) and blood volume pulse (BVP) were recorded at 128 Hz sampling rate using NeXus-10 (Mind Media B.V, Roermond-Herten, The Netherlands). For EDA, silver-silver chloride (Ag-AgCl) electrodes with 0.5% saline paste were attached to the medial side of the left foot with

adhesive skin tape and gauze. The BVP (heart rate) sensor was attached to the left index toe of the participant. Eye tracking was measured with a Pupil Labs Binocular 120 Hz headset with a custom-built headband[4].

2.3.3. Flow Short Scale

To measure self-reported Flow, participants were asked to fill in the Flow Short Scale (FSS) after each trial (Rheinberg et al., 2003; Engeser and Rheinberg, 2008). FSS has 10 core items which load the subfactors *fluency of performance* (6 items) and *absorption by activity* (4 items); plus 3 items for *perceived importance*. The response format of FSS is a 7-point Likert scale ranging from *Not at all* to *Very much*. Higher scores on the scales indicate higher experienced Flow and perceived importance. Example items include "My thoughts/activities run fluidly and smoothly" (*fluency of performance*), "I do not notice time passing" (*absorption by activity*), and "I must not make any mistakes here" (*perceived importance*). See **Supplementary Information** for full English text and Finnish translation.

Cronbach's alpha for a 10-item scale including the *fluency of performance* and *absorption by activity* items was 0.92; Cronbach's alpha was 0.87 for the 13-item FSS scale including *perceived importance* (Rheinberg et al., 2003). FSS authors (Rheinberg et al., 2003) suggest using the 10-item scale (excluding *perceived importance* subfactor) as a measure of experienced Flow. For our data also, Cronbach's alpha was higher for the core 10- than for 13-item scale. Thus, the Flow scale used in our analyses was formed by averaging the items in the *fluency of performance* and *absorption by activity* subfactors. The *perceived importance* subfactor was used separately in some analyses (see section 3).

In addition to the 13 main items asked after every trial, participants were asked at the end of every session to report 3 more items measuring the fit of skills and demands of the task (Rheinberg et al., 2003). These items also had 7-point scales, e.g.,: "For me personally, the current demands are... (too low—just right—too high)."

There was no Finnish translation of the scale available, so it was translated into Finnish by the authors. Two of the authors (native speakers of Finnish, no formal qualifications for English-Finnish translation) first made translations independently; these translations were compared and revised, then reviewed by other Finnish-native authors, and revised.

2.4. Procedure

After recruiting, participants selected eight suitable dates within a 3-week period. All sessions took place between 8 a.m. and 7 p.m. at Traffic Research Unit, Department of Digital Humanities, University of Helsinki. In the first session, participants were informed about the procedure of the study and asked to fill in a background information questionnaire, including information on health, driving experience and gaming experience, and an informed consent form.

The sessions were managed by two research assistants at a time, who observed the measurement, out of participants' line of sight behind a partition wall, and took notes about

---

[4]For technical details, see https://zenodo.org/record/1246953#.XJT-Ki10eqA

possible confounding factors and problems within the session. In the beginning of each session participants filled in a session-wise questionnaire on the use of contact lenses, restedness, and medication, caffeine, and nicotine intake.

In sessions with physiological measurements (1 and 5 to 8), participants were dressed in physiological sensors and an eye-tracking headset, seated in the driving seat in quiet, low-light conditions for baseline measurement. They were asked to sit still for 5 min, looking at a dark blue screen, while baseline was recorded. After baseline recording, participants played five game trials, filling FSS after each trial. Eye-tracking and physiological signals were recorded during trials. In sessions 2–4, participants played five trials straight after filling in the session-wise questionnaire, without a baseline period. The FSS was filled after each trial. At the end of Session 8, the participants were debriefed and given the reward of culture vouchers.

2.5. Statistical Methods

All statistical data processing reported herein was implemented with R platform for statistical computing (R Core Team, 2014). Where possible, exact corrected *p*-values are reported; inequalities are reported where exact values were not available. All *p*-values were corrected for multiple comparisons using Bonferroni-Holm. For all simple correlations we calculated Pearson's correlation coefficient, because all data in these tests were shown to be normally distributed by Shapiro-wilk tests and associated Q-Q plots.

For RQ1, participant-wise linear regression models were fitted using *lm* function in R, which also supplies R^2 values. The same approach was used to fit the "grand model" to group-wise data (i.e., pooled participants).

For RQ2, we obtained the independent variable as follows. For each participant and for each trial (40 trials in total), we subtracted predicted trial duration (*y*-value of power-law performance line) from observed trial duration, thus obtaining power-law model residuals, in units of log(sec). We refer to these within-participant trial-duration residuals as *deviation scores*, because they represent how much each observed trial duration deviates from the duration predicted by the model. Note, residuals are in the space of log-transformed trial durations in seconds and are therefore equivalent to ratio of performance in seconds. So for similar deviation scores from two trials, the later deviation represents a larger (or equal) effect in seconds.

Specifically, we first fit a linear mixed model with non-standardized Flow scores as the dependent variable, deviation scores as the predictor, and participant (numerical participant identifier ranging from 1 to 9) as a random factor with both random intercept and slope. This approach was chosen to handle the non-independence of data points within participants (see Bates et al., 2015).

Note that that there is no consensus on the best way to obtain *p*-values or estimates of effect sizes from linear mixed models. We have treated the *t* statistic as a *z* statistic using a standard normal distribution as a reference, and followed the method by Nakagawa and Schielzeth (2013) to obtain pseudo-R^2 values. Another way to statistically evaluate the significance of these results is via the binomial distribution: The (two-tailed)
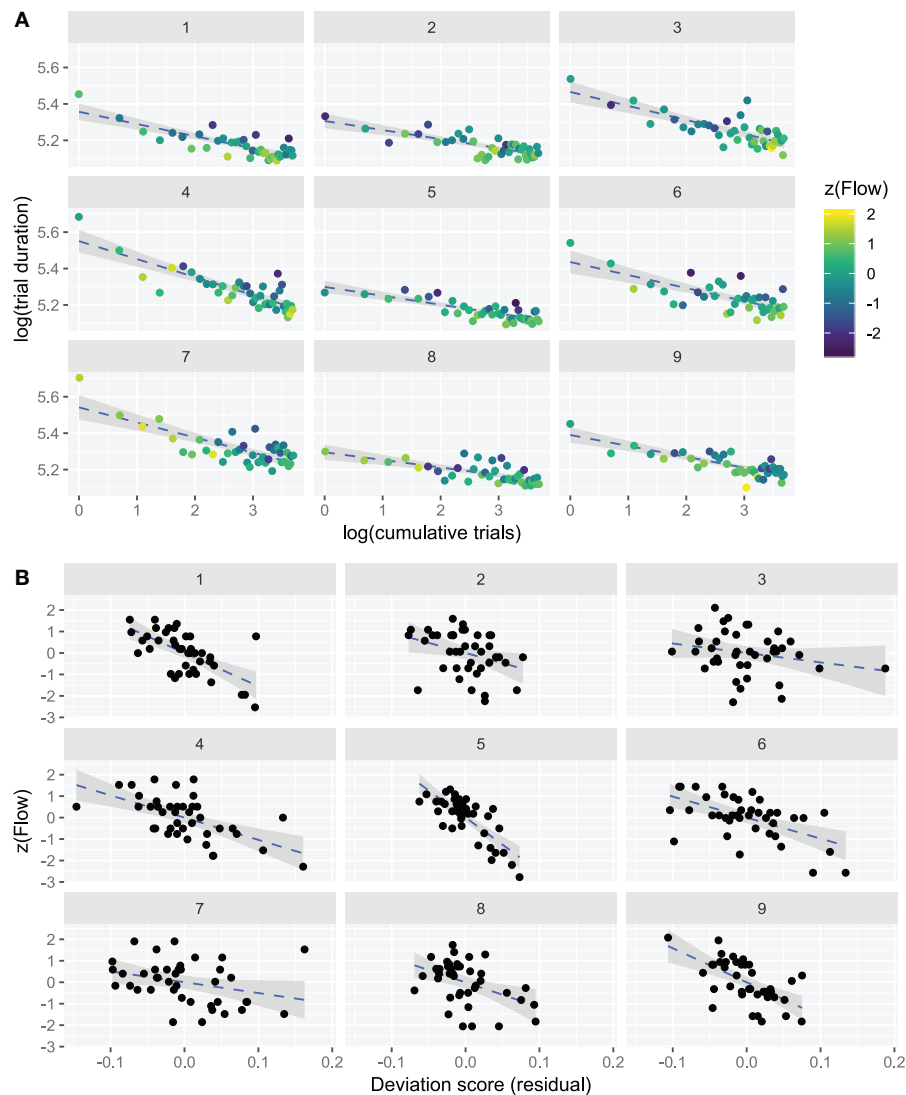
FIGURE 3 | (A) Participant-wise data showing logarithm-transformed performance and Flow self-reports in the speeded steering task. Ordinate shows log-duration of trials, abscissa shows log-cumulative trial count. Dashed blue lines fitted to the data are power-law learning curves, which transform to linear in log-log space. 95% confidence intervals around the slope are grayed. Flow scores (z-scored) are indicated by color. **(B)** Participant-wise deviation scores (observed trial duration minus predicted trial duration) plotted against Flow scores for each participant, and fitted by linear models. 95% confidence intervals around the slope are grayed.

probability of 9 negative slopes (should the probability of a negative slope per participant be 0.5, that is, fully random) is $p = 0.007$.

3. RESULTS

All participants completed the task (40 trials in total). Average trial duration was 186s (SD 18.2 s, min 162.2 s, max 300.1s). Average number of collisions was 17.8 (SD 4.9, min 5, max 40). Average trial velocity ranged between 1.37 and 2.54 units per step (mean 2.23, sd 0.19). Maximum instantaneous speed was 3.6 and minimum 1.06. The **Supplementary Information** provides comprehensive data on

performance-related features, such as trial duration, along with correlations between them; also it includes further details on participant self-report and background, plus validation of our main result for RQ2.

3.1. RQ1: How Does Performance Change Over Time?

What is the form of the learning curve, does it consistently improve e.g., as a power law of practice (Newell and Rosenbloom, 1982)? A power-law curve transformed to log-space will be linear. Thus, to investigate whether participant behavior follows a power law, we fitted a linear model in log-log space (log-transformed dependent and independent variable) of trial durations as a

function of cumulative number of trials, for each participant separately. **Figure 3A** shows this log-log performance data for each participant in each trial. Blue dashed lines indicate the power-law LC. Distance of points from the line (residuals) indicate the deviation of each trial from predicted learning: points above the line indicate longer duration (worse performance) than predicted by the LC, and vice versa.

All participant-specific log-log models had negative slopes, which indicates that with experience each participant learned to play better (obtained faster trial times). The variation in intercepts reflects disparity in participants' initial skill levels, and the variation in slopes the different learning rates. The individual intercepts and slopes of the models are presented in **Table 2**. A grand model was also fitted for all participants, and cumulative number of trials explained 39.6% of variance in trial durations. As the performance generally improves with cumulative trials, in agreement with a power-law of learning model, the explained variance can be ascribed to learning.

To confirm that a power law model gives a good approximation of learning, we compared its model-fit criterion against the fit of an exponential curve model (see **Supplementary Information** for details). While both models had good fit, the power law model was slightly better.

RQ1 can thus be answered: *the task was learned and the LC fit well to a power law model*. Given these positive answers, we may assume that the model provides a useful statistical estimate of performance expectation, i.e., how well the participants expect to perform can be estimated from the model.

3.2. RQ2: How Is Flow Related to Performance?

Participant-wise mean Flow and LC slope were related but not significantly correlated (Pearson's correlation coefficient $r = 0.6$, $p = 0.6$, $N = 9$). Since we have established that performance improves over sessions, we also used session number as a simple proxy of performance improvement. **Figure 4** shows the group-wise distribution of Flow scores plotted against sessions: clearly, there is no effect of session on group-wise median Flow (Pearson's $r = -0.12$, $p = 1.0$, $N = 8$). We found no significant effect of session type (sessions 2–4 vs. sessions 1, 5–8) on Flow scores [$F_{(1, 8)} = 3.18$, $p = 0.11$], by repeated measures ANOVA.

Next, we calculated the group-wise correlation of median duration and median Flow, separately for each session. The relationship between duration and Flow was intermittently significant *before correction for multiple comparisons*, but not after, and with no particular trend (range of Pearson's $r = [-0.05 \ldots -0.74]$, $p = [0.2 \ldots 1.0]$, $N = 9$ for all). These results suggest that higher Flow was sometimes associated with lower trial durations (i.e., better performance), but not strongly and not systematically. If we group sessions by condition (introduction = 1, practice = 2–4, main test = 5–8), we can visualize the evolution of performance against Flow more clearly than by plotting each session individually, see **Figure 5**.

The relationships between global (over all responses) Flow and performance appear weak, but we also wish to examine local
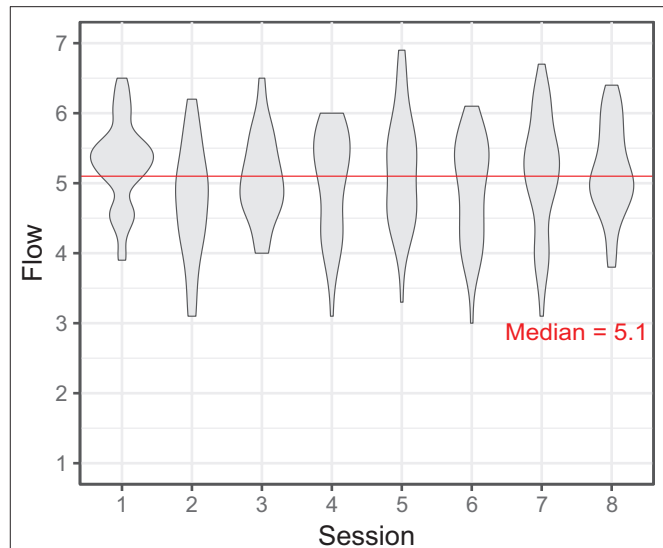
FIGURE 4 | Violin plot representing participants' self-reported Flow in sessions 1–8 (per-session Flow = mean of five trials. The self-report items are given in **Supplementary Information**, the scale was 1–7).

Flow for each trial separately. The points in each subplot of **Figure 3A** are colored according to Flow self-reports made after each trial, in a standardized range (original scores transformed to *z*-scores). The highest Flow scores are yellow, the lowest are navy-blue. Interestingly, this figure reveals at a glance that the points lie above and below the log-transformed power-law line in good agreement with the level of experienced Flow: worse performing trials (data-point above the line) tend to be more blue (Flow scores below the participant-wise mean), and better performing trials tend to be more yellow (scores above the mean). In other words, it seems that whenever participants were performing *better than predicted by the power-law line*, they were experiencing more Flow, and *vice versa*.

We evaluated whether this effect was robust and statistically significant. For each participant, we correlated their *deviation scores* (signed residuals from the power-law model) with their Flow scores, using a linear mixed model (see section 2). This model was statistically significant (deviation score $\beta = -8$, $t = -4.36$, $p = 0.002$) and the relationship is shown in **Figure 3B** (Flow scores are standardized). The conditional pseudo-R^2 value for this model was 0.47, corresponding to a correlation of 0.68, so that the model explains ∼47% of Flow score variability.

Thus, *high Flow scores are associated with better than predicted results* (trial durations below the predicted performance line), and *vice versa*. The strength of this association per participant follows from the strength of the correlation, and overall the model has large effect size.

As can be seen in **Figure 3B**, the trend was clearly negative for 7 out of 9 participants, while for two participants, 3 and 7, the trend was similar but the relationship was weaker. Notably, these two participants also reported lower scores on perceived importance: mean scores for these participants were 2.03 and

TABLE 2 | Individual learning rate parameters (cols 2–4), Flow (cols 5–6) and perceived importance (P.I., cols 7–8) scores. Last row shows group-mean values of each column.

Participant	Intercept sec	Intercept log	Slope	Flow mean	Flow SD	P.I. mean	P.I. SD
1	213	5.36	−0.067	5.10	0.51	3.53	0.57
2	202	5.31	−0.049	5.18	0.39	4.29	0.47
3	237	5.47	−0.077	5.36	0.64	2.03	0.50
4	257	5.55	−0.099	4.40	0.39	4.12	0.46
5	200	5.30	−0.049	5.44	0.88	5.16	0.64
6	230	5.44	−0.071	5.22	0.82	4.33	0.56
7	255	5.54	−0.083	4.69	0.53	2.22	0.53
8	198	5.29	−0.041	4.94	0.90	3.67	0.70
9	219	5.39	−0.059	5.25	0.79	4.62	0.55
Group mean	223	5.41	−0.066	5.06	0.65	3.77	0.55

FIGURE 5 | Duration and flow over sessions grouped: 1 (Introduction+physiological measurements), 2–4 (training), 5–8 (physiological measurements), $N = 72$.
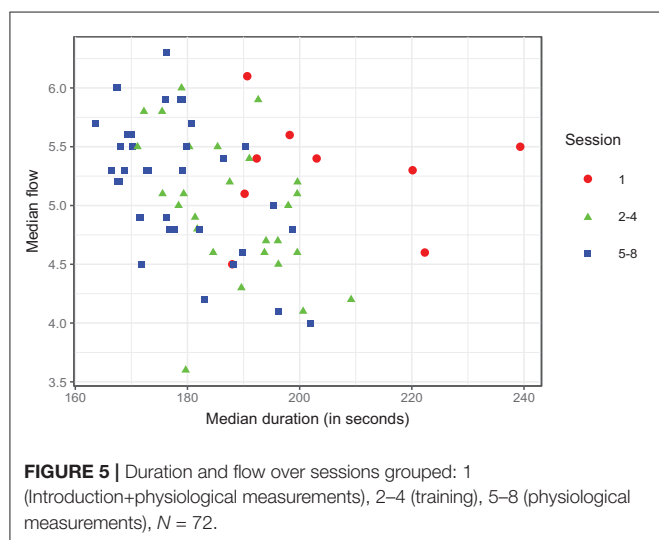
2.22, whereas the overall mean was 3.77 (see **Table 2**). However, the group-wise interaction between perceived importance scores and deviation scores was not statistically significant.

RQ2 can thus be answered: Flow was not consistently and robustly related to improvement in task performance with the skill acquisition occurring over 2 h of practice. It was, however, consistently related to whether performance was *better (or worse) than predicted* given the participant LC. Moreover, this effect might be moderated by self-reported perceived importance of the steering task (more data would be required to clarify).

4. DISCUSSION

We present a longitudinal experiment of Flow in a game-like high-speed steering task where task performance is easily parameterized and its relation to Flow analyzed. To induce Flow, the game was designed to hold the balance between skill and challenge constant: the difficulty of the game continually adapted to the skill level of the participant.

The results show the game was clearly Flow-inducing: mean Flow across sessions was reported as 5.1 (out of 7) on the FSS. This relatively high and stable mean Flow "baseline" induced by the game could be construed as reflecting the meeting of skill and challenge C1–3 by design.

We further found that Flow was not associated with gaining experience and skill in the game—our participants did not reliably report more Flow even as they learned, session by session, to complete the trials faster. This supports the theoretical position that Flow is elicited by the balance of skill and challenge, but show that Flow is less sensitive to the absolute level (within a task) of skill or challenge. This fits with the models which are more lenient regarding skill/challenge level: the original Flow Channel model and the latest Flow intensity model. The Quadrant/Octant models, which require "above-average" skills and challenges for Flow, are only supported under the assumption that this above-average reference level is task-specific and dynamically adjusted in step with the learning curve. Otherwise, if the reference level is fixed, then based on the Octant model experience of Flow would be predicted to increase along with the increase of skill level (and demand) during our longitudinal measurements of participant learning. In other words, when skills and challenges increased from a fixed reference, participants should have felt further "north-east" of the model midpoint where Flow bottoms out, and thus be more likely to report Flow and assign it greater intensity on a reporting scale. This was not observed. Therefore, our results do not fit the predictions of the Quadrant/Octant models *under the assumption of a fixed-demand reference level*. See also Keller and Landhäußer (2012) for critical discussion of these models. In absence of an independent motivation for such adjustment hypothesis it must be considered somewhat *ad hoc*. So, we suggest that—in order to incorporate the present findings—the octant model should be developed to provide a valid set of assumptions to support clear conclusions about the reference level.

4.1. Mechanisms of Learning and Flow
We also showed that higher trial-wise Flow (trials with higher self-reported Flow) was associated with trial durations shorter than expected by the power law LC model, and vice-versa.

Thus, Flow for each trial was higher-than-average or lower-than-average in agreement with task performance that was better or worse than expected (at the current level of skill). This stands in contrast to how mean Flow remained stably elevated across sessions. In other words, learning to play the game did not itself increase Flow; rather, the game induced a fairly high level of mean Flow, and trial-wise variability of Flow was correlated with better or worse than statistically-expected performance.

4.1.1. Alternative Explanations of Trial-Wise Results

It is a novel observation that trial-wise Flow relates to fluctuations of performance around the level expected from the learning curve. It is interesting because higher-than-expected performance in an individual trial can plausibly indicate either higher skill (e.g., better concentration), or lower challenge (easier-to-negotiate random placement of obstacles). Both are deviations from the average skill-challenge balance, yet may be associated with higher Flow. Either way, *our result undermines the assumption of Keller and Landhäußer (2012)'s Flow Intensity model that the direction of skill-challenge ratio can be ignored: our results show that Flow is elevated when skills exceed challenges.*

There are naturally several alternative explanations for this result. One possibility is that performance on some trials is enhanced by increased Flow during the trial—i.e., participants perform better when they "get into Flow." Another possibility is that the randomly-generated geometric layout of each trial might be more or less easy to negotiate; the "easier" trials would afford faster performance. Such random fluctuations of task difficulty could shift the skill-challenge ratio closer to one that the participant finds Flow-inducing. Alternatively, participants may be more likely to report higher Flow (after the fact) on a more successful (hence, more rewarding) trial, because they see their score before they complete the FSS. From these three alternatives, we find the first one the most convincing, because (A) comparing first to second, we believe that whatever biological and psychological mechanisms might be underpinning Flow will tend to create greater variation than the randomly-generated track layout; and (B) comparing first to third, the game design undermines the third alternative (see Limitations below). Ultimately however, the present paradigm cannot conclusively support any one alternative.

4.1.2. Cognitive Mechanisms of Flow

How can our approach and the present results be helpful to understand the mechanisms generating the Flow experience?

Methodologically, this study follows a *time series* approach rarely used in Flow research, by looking at changes in Flow over time, with relatively highly frequent and non-independent repeated measurements. This contrasts with much prior work which treats Flow as a relatively stable property, and allows us to look at on-task learning.

Learning implies skill increase, which (by game design) implies challenge increase, which (by design of Quadrant/Octant models) together imply Flow increase. As discussed, for such "state Flow" models the assumption of the fixed reference frame leads to the prediction of increased Flow with skill acquisition, which is not supported by our data. It is thus not straightforward

to reason about the evolution of Flow, or learning and Flow, based on these models, before resolving the crucial issue of the reference level for above-average task demand and skill.

For the Flow intensity model this particular problem does not arise (Keller and Landhäußer, 2012). However, the relation of learning and Flow is not entirely straightforward here, either. Increased skills should eventually increase task demands (because skill-learning increases access to the task's deeper levels of challenge), and thus perceived fit of skills and task demands. We did not assess perceived value directly, but it is plausible that time investment in and enjoyment of the game (as indicated by high mean Flow) would *also* increase the subjective value of the activity. If this is the case, again higher Flow should be elicited in step with increasing performance. This was not observed.

Overall, the lack of mechanistic hypotheses about the processes underlying the proposed dependencies in the Flow models make it difficult to make definite predictions in novel tasks, especially ones with changing task demands and skill, such as here. Prior work has provided some (neuro)cognitive, information-processing views on Flow (Marr, 2001; Cowley et al., 2008; Šimleša et al., 2018). Such work could provide an approach to make cognitive hypotheses about Flow, but these models have not been empirically tested, so it is unclear which (if any) to follow.

The aim of future work should then be to find out: what cognitive processes are *specific* to Flow-inducing task performance (in different stages of learning), but also *general* to multiple performance-domains. By so doing, we can in future attempt to clarify empirical observations by reference to a distinct cognitive theory of how Flow is generated.

4.1.3. Flow and Task Complexity

A possibly useful novel way to view Flow and learning is via task complexity. Csikszentmihalyi (1999) proposed that Flow should be possible in *any* task, complex (e.g., car driving) or simple (e.g., dish washing). But Keller and Landhäußer (2012) also proposed that Flow depends on perceived task value as well as challenge–skill balance. One way to resolve these ideas is to consider that the individual can *introduce* complexity (or value) to their activity if they appear to have exhausted that task's potential to challenge them. Nakamura and Csikszentmihalyi (2002) suggest such an exploratory mechanism to explain how individuals maintain Flow in complex tasks: "As people master challenges in an activity...to continue experiencing Flow, they must identify and engage progressively more complex challenges." The corollary for simple tasks is that individuals *create* complexity, e.g., with self-defined goals (Rauterberg, 1995). For example, a similar state to Flow, called the Zone, has been reported for machine-gambling addicts whose pastime is in fact skill-free, but who nevertheless *believe* that they are skilled (Schull, 2014).

In summary, complex tasks have deep structure to be learned, requiring non-trivial skill acquisition for any duration of learning and thus a shallow LC (learning is slow). Importantly, the skill level does not quickly peak, such as with simpler tasks like washing the dishes, where Flow might be obtained but cannot strongly interact with learning (without self-created complexity). Learning comes into play when we consider that the same

task can appear at first simple and later complex, e.g., as our experiment game.

4.2. Limitations and Future Work

Our study had a small convenience sample because (a) the recording paradigm was extensive (around 8 h of contact time), and (b) it was to some degree an exploratory study; both implying the need to constrain datasets to tractable sizes. It is worth noting that while we recruited only 9 participants, we collected a significant amount of behavioral and physiological data for each participant over a course of 2 weeks and 8 recording sessions. This amounts to a quite rich dataset allowing us to dig deep into the underpinnings of skilled high speed steering and Flow. Moreover, given the fact our experimental paradigm has not been used previously, we could not *a priori* easily estimate the statistical power required to discover significant effects. However, our results ended up being both strikingly *similar* and *strong* within each participant, suggesting that collecting a larger sample size would likely have *not* changed the pattern of the results or provided more in-depth insights. Regardless of the justifications, the sample size and recruitment method are minor limitations to be remedied in future work. For example, gender is known to affect performance visuomotor tasks (Feng et al., 2007). However, although there is a gender imbalance in our sample, the factor gender is confounded by background variables including gaming and driving experience, masking the true effect of gender. In order to properly study gender differences we would need to tailor our recruitment to that purpose. On the other hand, the result of Feng et al. (2007) suggests that gender differences might anyway balance out over the course of learning such a task as ours. Finally, because it lacks experimental manipulations the study cannot make strong causal claims, which could be improved by recording separate conditions of the game task, e.g., with varied difficulty levels.

As stated above, there are different plausible explanations of the main result linking Flow and performance. If FSS reports were indeed influenced by seeing the score beforehand, this should be considered a design limitation. However, the game gives such clear and direct feedback on performance (i.e., after collisions), that it is likely that self-assessment of performance would be similar with or without seeing the score. The score can thus be considered just a reinforcement of the perception of fit between skills and demands, which is anyway a required part of Flow reporting (Keller and Landhäußer, 2012). There is a possibility that Flow states could be affected by wearing eye-tracking and physiological equipment in some of the sessions. However, there was no difference in the level of subjective Flow reports in these sessions compared to other sessions.

Trial-by-trial analysis is limited by the Flow self-report which only has one data point per trial. It would be more powerful to analyse *inside* each trial. In our data, self-reported Flow is a point model of an entire trial, for which the participant knows their score. Self-reported Flow is thus an after-the-fact report, and could be criticized for not capturing the in-the-moment *experience of Flow*, which might fluctuate greatly during a trial. Analysing the fluctuation of Flow-experience requires us to model individual actions and/or their outcomes, and sample Flow during performance. This is a difficult challenge, because

paying attention to one's phenomenal state might easily disrupt the very processes sustaining that state, especially for Flow which is unreflective by definition. Future work should aim to model the conditions of Flow (C1–3) in real-time, while simultaneously recording participant physiology, to uncover in greater detail the relationships involved. The existing dataset will be used for this purpose, in a pending report on the biosignals recorded with high temporal resolution, primarily electrodermal activity.

Future work should also look into individual differences in learning (or cognitive) styles. For example, ample evidence suggests people can be roughly placed on a continuum of verbal as opposed to visual learners. These learning styles, in turn, have been positively linked to cognitive abilities on either verbal or visual tasks (e.g., Choi and Sardar, 2011; Knoll et al., 2017). Since the task used in our study was highly visual (there were no verbal cues nor audio), learning in it might be moderated by players cognitive learning styles. For example, perhaps those with a propensity for visual learning find it easier to get into flow and thus perform better. Therefore, an interesting avenue for further research using our driving task is first dividing people into verbalizers and visualizers, and then seeing if visualizers in particular find the game flow-inducing.

In terms of possible applications, game-induced Flow has been studied in the context of technology-enhanced learning (TEL) games (e.g., Cowley et al., 2014), but the style of activity in such games tends to be rather more complex than the driving task reported here. Use of such simple tasks for TEL games has been reported (Cowley et al., 2011), but it remains unclear what higher-learning benefit is derived from the Flow induced by the TEL game (Cowley and Bateman, 2017). In summary, there remains a large conceptual gap between what is known about task-learning and Flow, and how to make use of Flow-inducing tasks for higher learning applications.

4.3. Conclusion

We report results that self-reported Flow in a novel, challenging, and engaging high-speed steering task relates to trial-by-trial task performance relative to the learning curve: "better than expected" trials have higher Flow scores, and "worse than expected" trials have lower scores. The average level of self-reported Flow was high, as the game was specifically designed to meet the main preconditions of Flow, including balance of current skill and challenge. Perhaps surprisingly, Flow did not seem to change with global skill development or improvement in task performance.

These results show that: (1) If a reference level is important (as the octant model requires), it is so on a trial-wise scale. In other words, the reference level is task-specific and is continually adjusted during skill acquisition, following in step with the individual's own learning curve; (2) Contrary to the intensity model (Keller and Landhäußer, 2012) the direction of challenge-skill deviation cannot be ignored. Our study highlights a need for models of Flow to be developed in a way that better captures Flow dynamics, over the range of skill acquisition from novice to expert, than the state-like models of the phenomenal psychology tradition (Moneta, 2012).

Understanding how phenomenological experiences, such as Flow, relate to task performance is an important topic for

understanding human motivation and performance. This study contributes to this goal and we hope will inspire more inquiry into the dynamics of the Flow experience in different stages of learning a skill.

4.4. Context

The three senior authors conceived and guided the work. BC (the lead author) has studied computer gameplay from the perspective of performance, psychophysiology and emotional experiences. OL (the last author) has studied visuomotor skill in the domain of steering, and with JV he designed the "Flow-inducing" high-speed steering game. JP (the second author) has studied emotions and decision-making in games of skill and chance, and written on the similarities and differences between the Zone and Flow phenomena. Because these authors share a keen interest in understanding the development of expertise and the phenomenon of Flow, it was decided to join forces and put together a team of researchers to develop the present experiment on the basis of the steering game designed earlier. The other authors (all more junior graduate students) were recruited to work as part of their studies. This experiment is part of a larger effort to initiate a line of research into the neurocognitive processes underlying Flow and expert performance, combining experimental, psychophysiological, and computational methods.

DATA AVAILABILITY

The R code and data used to produce all analyses and figures is permanently available online at https://doi.org/10.6084/m9.figshare.7268387.

ETHICS STATEMENT

Participants were briefed and provided written informed consent before entering the study, and were aware of their legal rights. The study followed guidelines of the Declaration of Helsinki and was approved by the University of Helsinki Ethical review board in humanities and social and behavioral sciences (statement 31/2017; study title MulSimCoLab).

AUTHOR CONTRIBUTIONS

OL, JP, and BC conceived the study. OL and JV designed the gameplay. JV developed the game software. BC and TT designed and implemented the data collection. NL, TT, PP, RF, V-PI, and JP translated the FSS. All authors participated in decisions on the experiment specifications. RF, V-PI, NL, PP, and TT conducted the experiment. BC, V-PI, TT, RF, PP, NL, JP, and OL analyzed and interpreted the results. BC, JP, and OL drafted the paper. All authors participated in writing and reviewing, and approved the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01126/full#supplementary-material

REFERENCES

Bates, D., Mächler, 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

Chen, J. (2007). Flow in games (and everything else). *Commun. ACM* 50, 31–34. doi: 10.1145/1232743.1232769

Choi, J., and Sardar, S. (2011). An empirical investigation of the relationships among cognitive abilities, cognitive style, and learning preferences in students enrolled in specialized degree courses at a canadian college. *Can. J. Scholarship Teach. Learn.* 2:5. doi: 10.5206/cjsotl-rcacea.2011.1.5

Cornsweet, T. N. (1962). The staircase-method in psychophysics. *Am. J. Psychol.* 75, 485–491. doi: 10.2307/1419876

Cowley, B., Charles, D., Black, M., and Hickey, R. (2008). Toward an understanding of flow in video games. *Comput. Entertain.* 6, 1–27. doi: 10.1145/1371216.1371223

Cowley, B., Fantato, M., Jennett, C., Ruskov, M., and Ravaja, N. (2014). Learning when serious: psychophysiological evaluation of a technology-enhanced learning game. *J. Educ. Technol. Soc.* 17, 3–16.

Cowley, B., Moutinho, J. L., Bateman, C., and Oliveira, A. (2011). Learning principles and interaction design for green my place: a massively multiplayer serious game. *Entertain. Comput.* 2, 103–113. doi: 10.1016/j.entcom.2011.01.001

Cowley, B. U., and Bateman, C. (2017). Green my place: evaluation of a serious social online game designed to promote energy efficient behaviour change. *Int. J. Ser. Games* 4. doi: 10.17083/ijsg.v4i4.152

Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *J. Humanist. Psychol.* 15, 41–63. doi: 10.1177/002216787501500306

Csikszentmihalyi, M. (1999). If we are so rich, why aren't we happy? *Amer. Psychol.* 54, 821–827. doi: 10.1037//0003-066X.54.10.821

Csikszentmihalyi, M., and Bennett, S. (1971). An exploratory model of play. *Amer. Anthropol.* 73, 45–58. doi: 10.1525/aa.1971.73.1.02a00040

Engeser, S., and Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivat. Emot.* 32, 158–172. doi: 10.1007/s11031-008-9102-4

Engeser, S., and Schiepe-Tiska, A. (2012). *Historical Lines and an Overview of Current Research on Flow*. New York, NY: Springer.

Feng, J., Spence, I., and Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychol. Sci.* 18, 850–855. doi: 10.1111/j.1467-9280.2007.01990.x

Harmat, L., de Manzano, Ö., Theorell, T., Högman, L., Fischer, H., and Ullén, F. (2015). Physiological correlates of the flow experience during computer game playing. *Int. J. Psychophysiol.* 97, 1–7. doi: 10.1016/j.ijpsycho.2015.05.001

Harris, D. J., Vine, S. J., and Wilson, M. R. (2017). "Neurocognitive mechanisms of the flow state," in *Sport and the Brain: The Science of Preparing, Enduring and Winning, Part B*, volume 234 of *Progress in Brain Research*, eds M. R. Wilson, V. Walsh, and B. Parkin (Cambridge, MA: Elsevier), 221–243.

Jackson, S. A., and Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: the flow state scale. *J. Sport Exerc. Psychol.* 18, 17–35. doi: 10.1123/jsep.18.1.17

Keller, J., and Blomann, F. (2008). Locus of control and the flow experience: an experimental analysis. *Eur. J. Pers.* 22, 589–607. doi: 10.1002/per.692

Keller, J., and Landhäußer, A. (2012). *The Flow Model Revisited*. New York, NY: Springer.

Knoll, A. R., Otani, H., Skeel, R. L., and Van Horn, K. R. (2017). Learning style, judgements of learning, and learning of verbal and visual information. *Brit. J. Psychol.* 108, 544–563. doi: 10.1111/bjop.12214

Labonté-LeMoyne, É., Léger, P.-M., Resseguier, B., Bastarache-Roberge, M.-C., Fredette, M., Sénécal, S., et al. (2016). "Are we in flow neurophysiological correlates of flow states in a collaborative game," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16* (New York, NY: ACM Press), 1980–1988.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychol. Rev.* 95, 492–527. doi: 10.1037/0033-295X.95.4.492

Marr, A. J. (2001). Commentary: in the zone - a biobehavioral theory of the flow experience. *Athletic Insight* 3.

Massimini, F., Csikszentmihalyi, M., and Delle Fave, A. (1988). "Flow and bio-cultural evolution," in *Optimal Experience: Psychological Studies of Flow in Consciousness* (New York, NY: Cambridge University Press), 60–83.

Moneta, G. B. (2012). *On the Measurement and Conceptualization of Flow*. New York, NY: Springer.

Nakagawa, S., and Schielzeth, H. (2013). A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* 4, 133–142. doi: 10.1111/j.2041-210x.2012.00261.x

Nakamura, J., and Csikszentmihalyi, M. (2002). "The concept of flow," in *Handbook of Positive Psychology*, eds C. R. Snyder and S. J. Lopez (New York, NY: Oxford University Press), 89–105.

Newell, A., and Rosenbloom, P. S. (1982). *Mechanisms of Skill Acquisition and the Law of Practice*. Pittsburgh, PA: Design Research Center, Carnegie-Mellon University.

Palmeri, J. T. (1999). Theories of automaticity and the power law of practice. *J. Exp. Psychol.* 25, 543–551. doi: 10.1037/0278-7393.25.2.543

Peifer, C. (2012). *Psychophysiological Correlates of Flow-Experience*. New York, NY: Springer.

Peifer, C., Schulz, A., Schächinger, H., Baumann, N., and Antoni, C. H. (2014). The relation of flow-experience and physiological arousal under stress

Can u shape it? *J. Exp. Soc. Psychol.* 53, 62–69. doi: 10.1016/j.jesp.2014. 01.009

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rauterberg, M. (1995). "About a framework for information and information processing of learning systems," in *IFIP International Working Conference on Information System Concept: Toward a Consolidation of views* (Marburg: Chapman & Hall), 54–71.

Rheinberg, F., Vollmeyer, R., and Engeser, S. (2003). *Die Erfassung des Flow-Erlebens [The Assessment of Flow Experience]*. Göttingen: Hogrefe.

Schull, N. D. (2014). *Addiction by Design: Machine Gambling in Las Vegas*. Oxford; Princeton, NJ: Princeton University Press.

Šimleša, M., Guegan, J., Blanchard, E., Tarpin-Bernard, F., and Buisine, S. (2018). The flow engine framework: a cognitive model of optimal human experience. *Eur. J. Psychol.* 14, 232–253. doi: 10.5964/ejop.v14i1. 1370

Wolf, S., Brölz, E., Keune, P. M., Wesa, B., Hautzinger, M., Birbaumer, N., et al. (2015). Motor skill failure or flow-experience? Functional brain asymmetry and brain connectivity in elite and amateur table tennis players. *Biol. Psychol.* 105, 95–105. doi: 10.1016/j.biopsycho.2015.01.007

Is There an Optimal Autonomic State for Enhanced Flow and Executive Task Performance?

Michael S. Chin[1,2]* and Stefanos N. Kales[3,4]

[1] Division of General Internal Medicine and Public Health, Vanderbilt University School of Medicine, Nashville, TN, United States, [2] Vanderbilt Occupational Health, Vanderbilt University Medical Center, Nashville, TN, United States, [3] Environmental and Occupational Medicine and Epidemiology Program, Harvard T.H. Chan School of Public Health, Boston, MA, United States, [4] Department of Occupational Medicine, Cambridge Health Alliance, Harvard Medical School, Boston, MA, United States

Introduction: Flow describes a state of optimal experience that can promote a positive adaptation to increasing stress. The aim of the current study is to identify the ideal autonomic state for peak cognitive performance by correlating sympathovagal balance during cognitive stress with (1) perceived flow immersion and (2) executive task performance.

Materials and Methods: Autonomic states were varied in healthy male participants (*n* = 48) using combinations of patterned breathing and skeletal muscle contraction that are known to induce differing levels of autonomic response. After autonomic variation, a Stroop test was performed on participants to induce a mild stress response, and autonomic arousal was assessed using heart rate variability. Subjective experience of flow was measured by standardized self-report, and executive task performance was measured by reaction time on the Stroop test.

Results: There were significant associations between autonomic state and flow engagement with an inverted U-shaped function for parasympathetic stimulation, sympathetic response, and overall sympathovagal balance. There were also significant associations between autonomic states and reaction times. Combining sympathetic and parasympathetic responses to evaluate overall sympathovagal balance, there was a significant U-shaped relationship with reaction time.

Discussion: Our results support the flow theory of human performance in which the ideal autonomic state lies at the peak of an inverted-U function, and extremes at either end lead to both suboptimal flow experience. Similarly, cognitive task performance was maximized at the bottom of the U-function. Our findings suggest that optimal performance may be associated with predominant, but not total, sympathetic response.

Keywords: flow, heart rate variability, cognitive performance, parasympathetic and sympathetic reactivity, sympathovagal balance

INTRODUCTION

Recent public health focus on occupational burnout and stress resiliency has prompted further investigation into the role of the autonomic system in cognitive performance. Increasing number of recent studies have suggested that burnout is related to autonomic dysfunction during excessive stress (Lennartsson et al., 2016; Kanthak et al., 2017; May et al., 2018; Zhang et al., 2018; Traunmuller et al., 2019). However, despite these concerning observations, the effects of autonomic state on cognitive performance have not been fully defined, and some degree of increased stress may actually be desired for task performance. Increased autonomic arousal correlating with improved task performance is supported by several studies (Luft et al., 2009; Mathewson et al., 2010; Murray and Russoniello, 2012).

As a possible explanation resolving these seemingly conflicting findings, flow theory, as proposed by psychologist Csikszentmihalyi (1975), describes a state of optimal experience characterized by complete task immersion, effortless intention, intrinsic reward, and increased perception of control. This experience can promote a positive adaptation to increasing stress by a matching growth between challenges and skills through immediate task feedback.

The ideal physiological conditions to facilitate the flow experience have not been established. According to Csikszentmihalyi's (1975) theory, the flow state lies on the continuum between boredom and anxiety. Csikszentmihalyi (1975) proposed that there were physiological changes associated with the flow experience, but recognized this association between physiology and psychology were not readily established. Previous studies have identified a possible association between increased sympathetic enhancement during the flow experience, but none have adequately demonstrated an optimal level of autonomic arousal for both task performance and subjective flow experience (Keller et al., 2011; Gaggioli et al., 2013; Peifer et al., 2014; Bian et al., 2016).

Many popular mind–body disciplines such as Yoga, Tai Chi, and Qigong, have been shown to activate the parasympathetic nervous system to varying degrees (Goyal et al., 2014; Sullivan et al., 2018; Walther et al., 2018). As part of a related project, the authors previously studied healthy male subjects' responses to preconditioning using various rhythmic breathing and skeletal muscle contraction methods to vary their baseline autonomic states (Chin and Kales, 2019). Our previous study assessed paced respiration and dynamic tension through rhythmic skeletal muscle contraction as two core components common to Yoga, Tai Chi, and Qigong to better understand their interaction in activating the parasympathetic nervous system. The activation of the body's parasympathetic nervous system has been demonstrated to occur through respiratory entrainment effects (Jerath et al., 2006; Nijjar et al., 2014) as well as voluntary rhythmic muscle contraction (Lehrer et al., 2009; Vaschillo et al., 2011).

The aim of the current study is to utilize the autonomic variability resulting from these different patterns of preconditioning, when faced with a cognitive stressor, to identify an autonomic state for maximized cognitive performance. This study examines the relationship between sympathovagal balance during cognitive stress with (1) perceived flow immersion, and (2) task performance. Sympathovagal balance was assessed by measurement of heart rate variability (HRV), which can be considered an indicator of autonomic activity (McCraty and Shaffer, 2015). A recent meta-analysis of 37 studies by Kim et al. (2018) concluded that HRV is impacted by stress and can be used as an objective assessment of psychological stress.

MATERIALS AND METHODS

Participants

Forty-eight healthy male participants, ages 18–55 years were recruited from Harvard T.H. Chan School of Public Health- and Harvard Medical School-affiliated student programs, fellowships, and training residencies. Males were enrolled to minimize HRV due to hormonal variation (Sato and Miyake, 2004; Thayer et al., 2012; Koenig and Thayer, 2016). Any individuals with history of restrictive or obstructive lung disease, hypertension, or taking any medications that could affect heart rate were excluded. Caffeine consumption was not specifically restricted since withdrawal effects on HRV may exist for habitual caffeine users (Zimmermann-Viehoff et al., 2016).

Procedure

The study protocol was reviewed and approved by the IRB of the Harvard T.H. Chan School of Public Health. Testing occurred over a single 30-min session between the daytime hours of 09:00 and 16:00. A Polar H7 heart rate monitor (Polar Electro Oy, Kempele, Finland) was used as a validated research device to measure R–R intervals with accuracy comparable to electrocardiograms (Barbosa et al., 2016; Giles et al., 2016). A logging application on iPad (Apple Inc., Cupertino, CA, United States) recorded the R–R interval signals from the chest strap which were further analyzed.

Participants sat upright quietly for 5 min while reading the instructions for the study, and then, heart rate, cuff blood pressure, and respiration rate were measured. To generate varying baseline autonomic states among participants, subjects were randomized to one of four preconditioning groups: (1) nasal respiration at 0.1 Hz (inhale nose 5 s, exhale mouth 5 s) for 5 min, (2) contracting arm muscles by grasping a tennis ball at 0.1 Hz for 5 min (alternating contractions in left and right arms every 5 s), (3) performing contraction and nasal respiration in synchrony at 0.1 Hz for 5 min, and (4) reading consecutively four articles rated as emotionally neutral for 5 min. The contraction tasks have been demonstrated to vary cardiac reactivity through differing levels of resonance (Lehrer et al., 2000, 2009; Vaschillo et al., 2011). A graphical timer application on iPad was used to visually cue the breathing/contraction. The articles were *Scientific American* excerpts that were previously validated as emotionally neutral (van den Broek et al., 2001).

To assess executive task function, a computerized version of the Stroop test[1] was run for 5 min. The Stroop test has been demonstrated to produce a mild sympathetic response through

---

[1] http://cognitivefun.net

dissonant executive task function (Salahuddin et al., 2007; Visnovcova et al., 2014). Participants were asked to indicate the color of the word (and not its meaning) by keystroke, as quickly as possible, while minimizing their errors. For congruent trials, the displayed word and the color described by the word were the same. For incongruent trials, the displayed word and color presented were not the same. The reaction time for each word pair is recorded by the computer program with the premise of the Stroop test that incongruent pairs have longer reaction times when compared to congruent pairs (Dyer, 1973). As a marker for performance, a reaction time gap was calculated for each participant from the difference between congruent and incongruent pair reaction times. A shorter reaction time gap was considered to indicate higher performance.

At the end of this task period, a 5-min questionnaire Short Flow State Scale-2 (SFSS-2) was then administered to assess degree of flow engagement (Jackson et al., 2008). Flow engagement is the degree of perceived task immersion, and the SFSS-2 has been validated for evaluation of performance engagement. Respiration rates were monitored during all phases of testing to ensure they were within the 9–24 cycles/min range required to correspond accurately to vagal tone (Laborde et al., 2017).

Heart Rate Variability Analyses

Recorded R–R intervals were analyzed offline using Kubios HRV Premium software (Kubios Oy, Kuopio, Finland) using 2-min intervals based on recommendations from published standards (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996). Frequency domain measures analyze the power distribution of HRV as a function of high frequency (HF) and low frequency (LF). HF and LF components are reflective of parasympathetic and sympathetic activation, respectively. LF/HF can be regarded as the overall sympathovagal balance and degree of autonomic arousal (Pagani et al., 1986; Shaffer and Ginsberg, 2017).

Heart rate variability recordings during Stroop test were processed by Kubios HRV Premium. Automated artifact correction was performed for all recordings prior to analysis. One-hundred twenty second sampling periods were utilized to derive LF, HF, and LF/HF using Fast Fourier transformation spectrum method (**Figure 1**). LF and HF bands were standardly defined as 0.04–0.15 and 0.15–0.4 Hz, respectively, and absolute power for each band was analyzed in normalized units, LF or HF divided by total power (Malliani et al., 1994).

Statistical Analyses

All statistical analyses were performed using Prism 8 (GraphPad Software, San Diego, CA, United States). To analyze the overall effects of varying HRV on reaction time, HRV measures (LF, HF, LF/HF) for all subjects during the Stroop test were regressed on measured reaction time using the least squares regression method, with no weighting. Both linear and quadratic models were generated for each comparison. For all regressions, assumption of homoscedasticity was made. To assess the effects of varying HRV on flow engagement, HRV measures were similarly regressed on SFSS-2 scores, testing both linear and quadratic models. R^2-values for each relationship were evaluated at $p < 0.05$

significance level. Outliers were identified for each analysis using robust regression and outlier removal, which is a validated method of outlier detection included in Prism 8 (Motulsky and Brown, 2006). The outlier false discovery rate was set at <1%. Any reported outliers reported were excluded from the analysis.

RESULTS

All enrolled participants ($n = 48$) completed the study sessions without any adverse events. Participants' average age was 29.9 years ($SD \pm 5.96$). As a result of randomization, 12 subjects were each assigned to one of the four preconditioning groups.

HRV Measurements Between Preconditioning Groups

A previous study compared LF, HF, and LF/HF responses between the preconditioning groups measured during application of the Stroop test (Chin and Kales, 2019). As a summary of the findings of this analysis, the alternating contraction group had 71.7% higher activation of parasympathetic signal over respiration alone ($p < 0.001$). Alternating contractions synchronized with breathing demonstrated 150% higher parasympathetic activation than control ($p < 0.0001$). Between contraction alone and synchronized contraction groups, the synchronized group demonstrated 45.9% higher parasympathetic response during the cognitive stressor ($p < 0.001$).

HRV and Maximal Flow Engagement

Heart rate variability was regressed onto SFSS-2 scores to analyze autonomic association with flow engagement. Representative examples of HRV time varying frequency domain measures and flow scores are in **Figure 1**. In all cases, quadratic functions were statistically significant, whereas linear functions were not. HF (**Figure 2A**) indicated an inverted U-shaped relationship between parasympathetic stimulation and SFSS-2 scores ($DF = 44$, $R^2 = 0.110$, $p < 0.0001$); one outlier was identified and excluded from this analysis. Similarly, when analyzing the sympathetic response (**Figure 2B**), a reciprocal relationship was found with an inverted U-shaped relationship between increasing and SFSS-2 scores ($DF = 45$, $R^2 = 0.070$, $p < 0.001$). When combining sympathetic and parasympathetic responses to measure overall sympathovagal balance, LF/HF (**Figure 2C**) demonstrated a significant inverted-U relationship when regressed on SFSS-2 ($DF = 45$, $R^2 = 0.187$, $p < 0.0001$). The interpolated value of the vertex was LF/HF 6.822 with a maximal SFSS-2 score of 38.85.

HRV and Cognitive Performance

Reaction time gaps were used as a marker for cognitive performance. When HRV was regressed onto reaction time gaps, there were significant second-order relationships for all measures; linear regression analysis demonstrated a significant relationship only for LF. HF (**Figure 3A**) indicated a positive curvilinear relationship between parasympathetic stimulation and reaction time ($DF = 45$, $R^2 = 0.053$, $p < 0.05$). When analyzing the sympathetic response (**Figure 3B**), a reciprocal
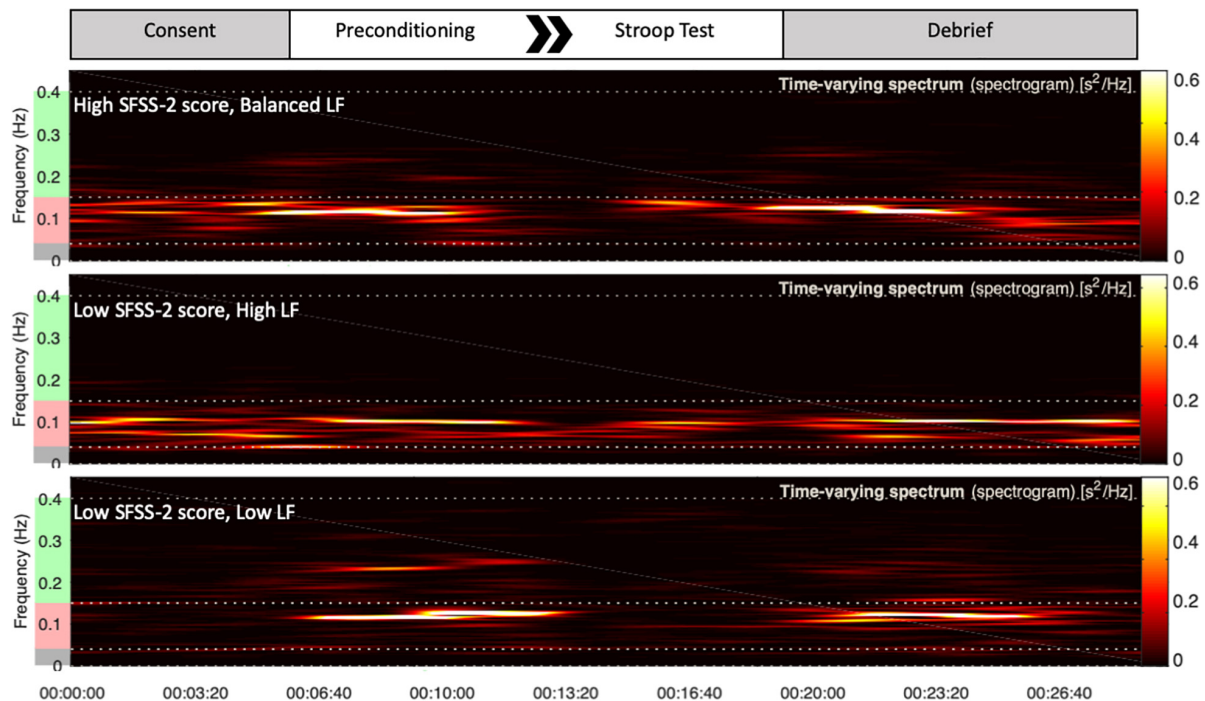
FIGURE 1 | Top portion of the figure is a schematic of each study session. Underneath are representative examples of HRV frequency domain measures during the session for three different subjects. Increased power in the green frequency band represents parasympathetic HF response. Increased power in the red frequency band represents sympathetic LF response. During the Stroop test, the top subject exhibited high flow with a balanced LF response. The bottom two subjects, at extremes of either high or low sympathetic response, reported low flow.
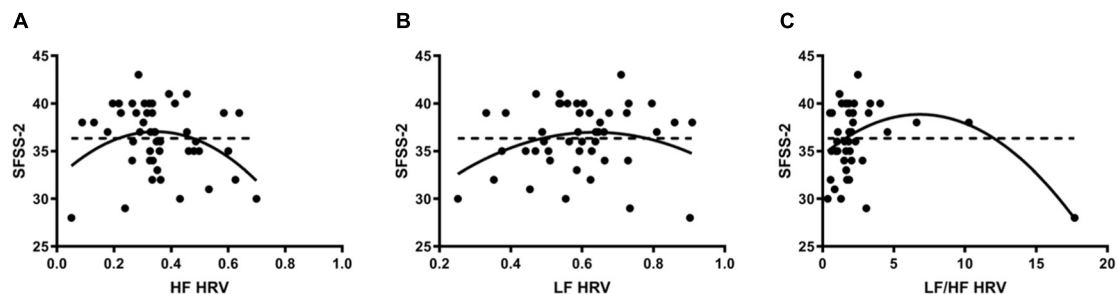
FIGURE 2 | (A) Inverted U-shaped relationship between parasympathetic stimulation and SFSS-2 scores ($DF = 44$, $R^2 = 0.110$, $p < 0.0001$); one outlier was identified and excluded from this analysis. Dotted line represents the linear model ($DF = 46$, $R^2 = 0.026$, $p > 0.05$). Each point represents a single participant's HF measured during the cognitive stressor. **(B)** A reciprocal relationship was found with an inverted U-shaped relationship between increasing sympathetic response (LF) and SFSS-2 scores ($DF = 45$, $R^2 = 0.070$, $p < 0.001$). Dotted line represents the linear model ($DF = 46$, $R^2 = 0.00087$, $p > 0.05$). Each point represents a single participant's LF measured during the cognitive stressor. **(C)** When combining sympathetic and parasympathetic responses to measure overall sympathovagal balance, LF/HF demonstrated a significant inverted-U relationship when regressed on SFSS-2 ($DF = 45$, $R^2 = 0.187$, $p < 0.0001$). The interpolated value of the vertex was LF/HF 6.822 with a maximal SFSS-2 score of 38.85. Dotted line represents the linear model ($DF = 46$, $R^2 = 0.031$, $p > 0.05$). Each point represents a single participant's LF/HF measured during the cognitive stressor.

relationship was found with increasing LF resulting in decreasing reaction time gaps ($DF = 45$, $R^2 = 0.117$, $p < 0.05$). After combining sympathetic and parasympathetic responses to evaluate overall sympathovagal balance, LF/HF (**Figure 3C**) demonstrated a significant U-shaped relationship when regressed on reaction time ($DF = 45$, $R^2 = 0.046$, $p < 0.0001$). The interpolated value of the vertex was LF/HF 11.61 with a minimal reaction time gap of 143.5 ms.

DISCUSSION

After autonomic markers were regressed on flow scores, we demonstrated a possible inverted U-relationship between overall sympathovagal balance and self-reported experience of flow. With the exception of the LF relationship to reaction time, none of the linear models were statistically significant. Our finding supports Csikszentmihalyi's (1975) theory of human

FIGURE 3 | (A) When HRV was regressed onto reaction time gaps, HF indicated a positive curvilinear relationship between parasympathetic stimulation and reaction time ($DF = 45$, $R^2 = 0.053$, $p < 0.05$). Dotted line represents the linear model ($DF = 46$, $R^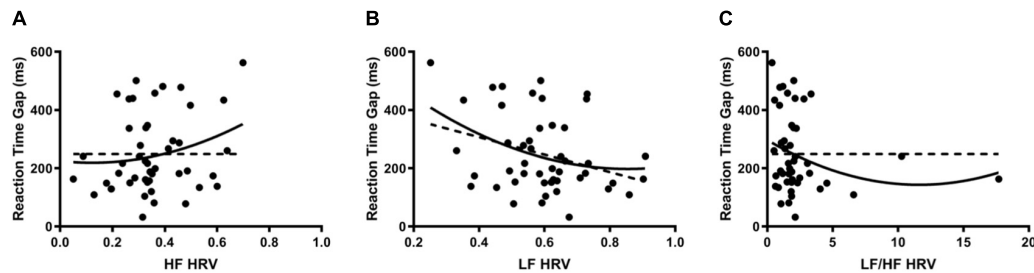2 = 0.046$, $p > 0.05$). Each point represents a single participant's HF measured during the cognitive stressor. **(B)** When analyzing the sympathetic response, a reciprocal relationship was found with increasing LF resulting in decreasing reaction time gaps ($DF = 45$, $R^2 = 0.117$, $p < 0.05$). Dotted line represents the linear model ($DF = 46$, $R^2 = 0.10$, $p < 0.05$). Each point represents a single participant's LF measured during the cognitive stressor. **(C)** After combining sympathetic and parasympathetic responses to measure overall sympathovagal balance, LF/HF demonstrated a significant U-shaped relationship when regressed on reaction time ($DF = 45$, $R^2 = 0.046$, $p < 0.0001$). The interpolated value of the vertex was LF/HF 11.61 with a minimal reaction time gap of 143.5 ms. Dotted line represents the linear model ($DF = 46$, $R^2 = 0.0083$, $p > 0.05$). Each point represents a single participant's LF/HF measured during the cognitive stressor.

performance in which the optimal autonomic state lies at the peak of an inverted-U, in which extremes at either end lead to suboptimal experience.

Studies on leisure activities, such as playing piano (de Manzano et al., 2010) and video games (Kozhevnikov et al., 2018), have effectively demonstrated the left side of the inverted-U curve, in which increasing sympathetic activation leads to increased flow experience. Paradoxically, other studies on job stress illustrate that increasing sympathetic arousal can also lead to declining performance and burnout. A recent systematic review identified 13 studies which confirmed a negative association between parasympathetic response and job stress or burnout (de Looff et al., 2018). When taken together, these seemingly opposing responses to arousal may actually just represent different sides of the inverted-U relationship demonstrated in our results.

A significant limitation in previous studies remained that the full spectrum of sympathetic arousal had not been represented due to insufficiently varied autonomic states. Peifer et al. (2014) addressed this shortcoming in study design by varying the arousal stimulus and inducing differing levels of social stress prior to task testing. Interestingly, using this design to create further autonomic variability, the authors did successfully demonstrate a quadratic inverted-U relationship for sympathetic LF HRV. However, they demonstrated only a linear association between increasing HF parasympathetic response and flow experience, but it should be noted that the study power was limited by the small study sample of only 22 participants.

Similar to Peifer et al. (2014)'s study design, we varied our baseline autonomic states. However, instead of priming subjects using varying social stressors, our methodology used differing combinations of patterned breathing and skeletal muscle contraction that are known to induce varying levels of autonomic response. With this preconditioning, we were able to prime enough baseline autonomic variation to demontrate the inverted-U relationship between increasing sympathetic arousal and flow experience. Improving on Peifer et al. (2014), we did successfully demonstrate an inverted-U

relationship between increasing parasympathetic response and flow. Furthermore, when analyzing combined sympathic and parasympathetic balance as LF/HF, we found that the inverted-U curve remained significant with the interpolated maximum flow occuring when LF/HF balance was 6.822. Stated another way, the optimal flow experience occurred when the autonomic state comprised of 87% sympathetic and 13% parasympathetic response.

Our results also support the potential existence of an optimal autonomic state for high cognitive performance. Reaction time during the Stroop test is often referenced as a marker of executive function (Egner and Hirsch, 2005). When interpreting the effect of HRV on reaction times, the overall sympathovagal balance suggested the minimum reaction time at a sympathetically dominated balance. With the estimated minimum reaction time at LF/HF 11.61, this could be interpreted as the fastest reaction times occurring when the autonomic state is at predominantly (92%) sympathetic response. It should also be noted that while our results were generally statistically significant, autonomic state only explained approximately 10–20% of the overall variability observed in flow scores. While not able to fully account for changes in flow, our results suggest that autonomic state at least partially influences the flow experience. Other unaccounted factors, such as mindset or task familiarity, may play important roles in further determining the flow relationship.

The finding of increased arousal in autonomic states correlating with improved task performance is supported by several studies. In a study on executive function using a similar Stroop color test, increased sympathetic tone was associated with faster response times in color naming (Mathewson et al., 2010). Other studies have demonstrated that invoking a higher stress response can also increase cognitive performance. In a study on perceived threat of electric shock, subjects with increased anxiety demonstrated faster reaction time in executive tasks under the threat of unpredictable shocks to the hand (Cantelon et al., 2018). In studies on physical exertion in athletes, higher stress responses from

exertion were associated with faster reaction times (Luft et al., 2009; Murray and Russoniello, 2012).

Limitations

There are two main limitations to the external validity of our pilot study. Since estrogen levels can affect HRV during the stress response, we chose to limit our participant enrollment to male participants in order to minimize hormonal variability. Therefore, we may have limited generalizability to females. Future studies should validate our findings in a female population. Secondly, our findings are limited based on a relatively small pilot study. Replication of these findings should be attempted in a future on a larger sample size.

Use of LF/HF as a measure of sympathovagal balance has been highly debated (Eckberg, 1997; Billman, 2013), but despite this challenge, recent studies on stress response still use LF/HF as a marker for sympathovagal balance (Lennartsson et al., 2016; Cao et al., 2019). von Rosenberg et al. (2017) has proposed a new two-dimensional method of LF/HF analysis to categorize mental and physical stresses. This new methodology might be promising for a future analysis in a larger confirmatory study.

CONCLUSION

When reviewing both executive task function and flow experience, both indices appeared to be maximized at approximately 90% sympathetic state. To our knowledge, this is the first study to suggest U-shaped relationships existing simultaneously for both flow experience and executive task function, suggesting that optimal performance may be associated with predominant, but not total, sympathetic response. Our findings are based on a small pilot so the results should be approached with caution. However, they do provide a preliminary foundation to understand the practical applications of autonomic modulation to potentially enhance performance during high-stress situations.

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 IRB of the Harvard T.H. Chan School of Public Health 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 IRB of the Harvard T.H. Chan School of Public Health.

AUTHOR CONTRIBUTIONS

MC originated the study design, conducted the experiments, and performed all study analyses. SK contributed to the study design. Both authors drafted, revised, and approved the manuscript.

REFERENCES

Barbosa, M. P., da Silva, N. T., de Azevedo, F. M., Pastre, C. M., and Vanderlei, L. C. (2016). Comparison of polar(R) RS800G3 heart rate monitor with Polar(R) S810i and electrocardiogram to obtain the series of RR intervals and analysis of heart rate variability at rest. *Clin. Physiol. Funct. Imaging* 36, 112–117. doi: 10.1111/cpf.12203

Bian, Y., Yang, C., Gao, F., Li, H., Zhou, S., Li, H., et al. (2016). A framework for physiological indicators of flow in VR games: construction and preliminary evaluation. *Pers. Ubiquitous Comput.* 20, 821–832. doi: 10.1007/s00779-016-0953-5

Billman, G. E. (2013). The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Front. Physiol.* 4:26. doi: 10.3389/fphys.2013.00026

Cantelon, J. A., Giles, G. E., Eddy, M. D., Haga, Z., Mahoney, C. R., Taylor, H. A., et al. (2018). Exerting cognitive control under threat: interactive effects of physical and emotional stress. *Emotion* 15. doi: 10.1037/emo0000509 [Epub ahead of print].

Cao, X., MacNaughton, P., Cadet, L. R., Cedeno-Laurent, J. G., Flanigan, S., Vallarino, J., et al. (2019). Heart rate variability and performance of commercial airline pilots during flight simulations. *Int. J. Environ. Res. Public Health* 16:237. doi: 10.3390/ijerph16020237

Chin, M. S., and Kales, S. N. (2019). Understanding mind–body disciplines: a pilot study of breathing and dynamic muscle contraction on autonomic nervous system reactivity. *Stress and Health* 1–7. doi: 10.1002/smi.2887

Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*, 1st Edn. San Francisco, CA: Jossey-Bass Publishers, 231.

de Looff, P. C., Cornet, L. J. M., Embregts, P., Nijman, H. L. I., and Didden, H. C. M. (2018). Associations of sympathetic and parasympathetic activity in job stress and burnout: a systematic review. *PLoS One* 13:e0205741. doi: 10.1371/journal.pone.0205741

de Manzano, O., Theorell, T., Harmat, L., and Ullen, F. (2010). The psychophysiology of flow during piano playing. *Emotion* 10, 301–311. doi: 10.1037/a0018432

Dyer, F. N. (1973). The stroop phenomenon and its use in the stlldy of perceptual, cognitive, and response processes. *Mem. Cogn.* 1, 106–120. doi: 10.3758/bf03198078

Eckberg, D. L. (1997). Sympathovagal balance: a critical appraisal. *Circulation* 96, 3224–3232. doi: 10.1161/01.cir.96.9.3224

Egner, T., and Hirsch, J. (2005). The neural correlates and functional integration of cognitive control in a Stroop task. *Neuroimage* 24, 539–547. doi: 10.1016/j.neuroimage.2004.09.007

Gaggioli, A., Cipresso, P., Serino, S., and Riva, G. (2013). Psychophysiological correlates of flow during daily activities. *Stud. Health Technol. Inform.* 191, 65–69.

Giles, D., Draper, N., and Neil, W. (2016). Validity of the polar V800 heart rate monitor to measure RR intervals at rest. *Eur. J. Appl. Physiol.* 116, 563–571. doi: 10.1007/s00421-015-3303-9

Goyal, M., Singh, S., Sibinga, E. M., Gould, N. F., Rowland-Seymour, A., Sharma, R., et al. (2014). Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. *JAMA Intern. Med.* 174, 357–368. doi: 10.1001/jamainternmed.2013.13018

Jackson, S. A., Martin, A. J., and Eklund, R. C. (2008). Long and short measures of flow: the construct validity of the FSS-2, DFS-2, and new brief counterparts. *J. Sport Exerc. Psychol.* 30, 561–587. doi: 10.1123/jsep.30.5.561

Jerath, R., Edry, J. W., Barnes, V. A., and Jerath, V. (2006). Physiology of long pranayamic breathing: neural respiratory elements may provide a mechanism that explains how slow deep breathing shifts the autonomic nervous system. *Med. Hypotheses* 67, 566–571. doi: 10.1016/j.mehy.2006.02.042

Kanthak, M. K., Stalder, T., Hill, L. K., Thayer, J. F., Penz, M., and Kirschbaum, C. (2017). Autonomic dysregulation in burnout and depression: evidence for the central role of exhaustion. *Scand. J. Work Environ. Health* 43, 475–484. doi: 10.5271/sjweh.3647

Keller, J., Bless, H., Blomann, F., and Kleinböhl, D. (2011). Physiological aspects of flow experiences: skills-demand-compatibility effects on heart rate variability and salivary cortisol. *J. Exp. Soc. Psychol.* 47, 849–852. doi: 10.1016/j.jesp.2011.02.004

Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., and Koo, B. H. (2018). Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry Investig.* 15, 235–245. doi: 10.30773/pi.2017.08.17

Koenig, J., and Thayer, J. F. (2016). Sex differences in healthy human heart rate variability: a meta-analysis. *Neurosci. Biobehav. Rev.* 64, 288–310. doi: 10.1016/j.neubiorev.2016.03.007

Kozhevnikov, M., Li, Y., Wong, S., Obana, T., and Amihai, I. (2018). Do enhanced states exist? Boosting cognitive capacities through an action video-game. *Cognition* 173, 93–105. doi: 10.1016/j.cognition.2018.01.006

Laborde, S., Mosley, E., and Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research - recommendations for experiment planning, data analysis, and data reporting. *Front. Psychol.* 8:213. doi: 10.3389/fpsyg.2017.00213

Lehrer, P., Vaschillo, E., Trost, Z., and France, C. R. (2009). Effects of rhythmical muscle tension at 0.1Hz on cardiovascular resonance and the baroreflex. *Biol. Psychol.* 81, 24–30. doi: 10.1016/j.biopsycho.2009.01.003

Lehrer, P. M., Vaschillo, E., and Vaschillo, B. (2000). Resonant frequency biofeedback training to increase cardiac variability: rationale and manual for training. *Appl. Psychophysiol. Biofeedback* 25, 177–191.

Lennartsson, A. K., Jonsdottir, I., and Sjors, A. (2016). Low heart rate variability in patients with clinical burnout. *Int. J. Psychophysiol.* 110, 171–178. doi: 10.1016/j.ijpsycho.2016.08.005

Luft, C. D., Takase, E., and Darby, D. (2009). Heart rate variability and cognitive function: effects of physical effort. *Biol. Psychol.* 82, 164–168. doi: 10.1016/j.biopsycho.2009.07.007

Malliani, A., Lombardi, F., and Pagani, M. (1994). Power spectrum analysis of heart rate variability: a tool to explore neural regulatory mechanisms. *Br. Heart J.* 71, 1–2. doi: 10.1136/hrt.71.1.1

Mathewson, K. J., Jetha, M. K., Drmic, I. E., Bryson, S. E., Goldberg, J. O., Hall, G. B., et al. (2010). Autonomic predictors of Stroop performance in young and middle-aged adults. *Int. J. Psychophysiol.* 76, 123–129. doi: 10.1016/j.ijpsycho.2010.02.007

May, R. W., Seibert, G. S., Sanchez-Gonzalez, M. A., and Fincham, F. D. (2018). School burnout and heart rate variability: risk of cardiovascular disease and hypertension in young adult females. *Stress* 21, 211–216. doi: 10.1080/10253890.2018.1433161

McCraty, R., and Shaffer, F. (2015). Heart rate variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk. *Glob. Adv. Health Med.* 4, 46–61. doi: 10.7453/gahmj.2014.073

Motulsky, H. J., and Brown, R. E. (2006). Detecting outliers when fitting data with nonlinear regression - a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinform.* 7:123. doi: 10.1186/1471-2105-7-123

Murray, N. P., and Russoniello, C. (2012). Acute physical activity on cognitive function: a heart rate variability examination. *Appl. Psychophysiol. Biofeedback* 37, 219–227. doi: 10.1007/s10484-012-9196-z

Nijjar, P. S., Puppala, V. K., Dickinson, O., Duval, S., Duprez, D., Kreitzer, M. J., et al. (2014). Modulation of the autonomic nervous system assessed through heart rate variability by a mindfulness based stress reduction program. *Int. J. Cardiol.* 177, 557–559. doi: 10.1016/j.ijcard.2014.08.116

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996). Heart rate variability. Standards of measurement physiological. interpretation, and clinical use. *Eur. Heart J.* 17, 354–381. doi: 10.1093/oxfordjournals.eurheartj.a014868

Pagani, M., Lombardi, F., Guzzetti, S., Rimoldi, O., Furlan, R., Pizzinelli, P., et al. (1986). Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circ. Res.* 59, 178–193. doi: 10.1161/01.res.59.2.178

Peifer, C., Schulz, A., Schächinger, H., Baumann, N., and Antoni, C. H. (2014). The relation of flow-experience and physiological arousal under stress — Can u shape it? *J. Exp. Soc. Psychol.* 53, 62–69. doi: 10.1016/j.jesp.2014.01.009

Salahuddin, L., Cho, J., Jeong, M. G., and Kim, D. (2007). Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2007, 4656–4659. doi: 10.1109/IEMBS.2007.4353378

Sato, N., and Miyake, S. (2004). Cardiovascular reactivity to mental stress: relationship with menstrual cycle and gender. *J. Physiol. Anthropol. Appl. Hum. Sci.* 23, 215–223. doi: 10.2114/jpa.23.215

Shaffer, F., and Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Front. Public Health* 5:258. doi: 10.3389/fpubh.2017.00258

Sullivan, M. B., Erb, M., Schmalzl, L., Moonaz, S., Noggle Taylor, J., and Porges, S. W. (2018). Yoga therapy and polyvagal theory: the convergence of traditional wisdom and contemporary neuroscience for self-regulation and resilience. *Front. Hum. Neurosci.* 12:67. doi: 10.3389/fnhum.2018.00067

Thayer, J. F., Ahs, F., Fredrikson, M., Sollers, J. J. III, and Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neurosci. Biobehav. Rev.* 36, 747–756. doi: 10.1016/j.neubiorev.2011.11.009

Traunmuller, C., Stefitz, R., Gaisbachgrabner, K., Hofmann, P., Roessler, A., and Schwerdtfeger, A. R. (2019). Psychophysiological concomitants of burnout: evidence for different subtypes. *J. Psychosom. Res.* 118, 41–48. doi: 10.1016/j.jpsychores.2019.01.009

van den Broek, P., Lorch, R. F. Jr., Linderholm, T., and Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Mem. Cogn.* 29, 1081–1087. doi: 10.3758/bf03206376

Vaschillo, E. G., Vaschillo, B., Pandina, R. J., and Bates, M. E. (2011). Resonances in the cardiovascular system caused by rhythmical muscle tension. *Psychophysiology* 48, 927–936. doi: 10.1111/j.1469-8986.2010.01156.x

Visnovcova, Z., Mestanik, M., Javorka, M., Mokra, D., Gala, M., Jurko, A., et al. (2014). Complexity and time asymmetry of heart rate variability are altered in acute mental stress. *Physiol. Meas.* 35, 1319–1334. doi: 10.1088/0967-3334/35/7/1319

von Rosenberg, W., Chanwimalueang, T., Adjei, T., Jaffer, U., Goverdovsky, V., and Mandic, D. P. (2017). Resolving ambiguities in the LF/HF ratio: LF-HF scatter plots for the categorization of mental and physical stress from HRV. *Front. Physiol.* 8:360. doi: 10.3389/fphys.2017.00360

Walther, A., Lacker, T. J., and Ehlert, U. (2018). Everybody was Kung-Fu fighting-The beneficial effects of Tai Chi Qigong and self-defense Kung-Fu training on psychological and endocrine health in middle aged and older men. *Complement Ther. Med.* 36, 68–72. doi: 10.1016/j.ctim.2017.11.021

Zhang, M., Liu, L., Shi, Y., Yang, Y., Yu, X., Angerer, P., et al. (2018). Longitudinal associations of burnout with heart rate variability in patients following acute coronary syndrome: a one-year follow-up study. *Gen. Hosp. Psychiatry* 53, 59–64. doi: 10.1016/j.genhosppsych.2018. 05.008

Zimmermann-Viehoff, F., Thayer, J., Koenig, J., Herrmann, C., Weber, C. S., and Deter, H. C. (2016). Short-term effects of espresso coffee on heart rate variability and blood pressure in habitual and non-habitual coffee consumers– a randomized crossover study. *Nutr. Neurosci.* 19, 169–175. doi: 10.1179/ 1476830515Y.0000000018

Effects of a Brief Strange Loop Task on Immediate Word Length Comparison: A Mindfulness Study on Non-striving

Ying Hwa Kee*, Khin Maung Aye, Raisyad Ferozd and Chunxiao Li

National Institute of Education, Nanyang Technological University, Singapore, Singapore

Non-striving is an important aspect of mindfulness practice, but it has not been sufficiently researched. This study examines whether a strange loop-based task – Infinite Water Scooping Task – performed for 10 min, has an effect on non-striving behavior and performance in a subsequent word length comparison task. Results showed that performance (number of correct trials) did not differ significantly between the two groups, though the experimental group tended to perform worse. However, participants in the experimental group took a significantly shorter time to respond to the word length comparison task than those in the control group. It is inferred that shorter time taken reflects response without investing much effort to count with care, i.e., non-striving. The present study demonstrates that the brief strange loop task implemented in this study elicited non-striving behavior compared to the effects of the control task, and this adds to the understanding of non-striving in the context of mindfulness. The Infinite Water Scooping Task may be useful for illustrating and teaching non-striving within mindfulness practice.

Keywords: mindfulness, paradox and ambiguity, reaction time, motivation, non-striving, goal setting

Mindfulness can be generally described as repetitive and sustained attentional efforts toward ongoing moments in an accepting and non-judgmental fashion (Kabat-Zinn, 1990). The consensus from previous operationalization of mindfulness points to attentional regulation and non-judgmental acceptance as two key aspects of mindfulness (Bishop et al., 2004; Lindsay and Creswell, 2017), consistent with the general definition above. The notion of acceptance and being non-judgmental underpins the uniqueness of mindfulness practice relative to other coping strategies, which are often focused on problem identification and intended eradication of problems, such as those found in traditional cognitive behavioral therapy approaches (Hofmann and Asmundson, 2008). In contract, the practice of acceptance and non-judgmental awareness within the mindfulness approach downplays problem eradication as mindfulness practice typically involves an observation mode, mere labeling of emotions, and distancing from mental content (Hinton et al., 2013). Suffice to say, the notion of "letting go" captures the sense of non-judgmental acceptance sufficiently well, as seen in previous literature (e.g., Frewen et al., 2008; Bergeron et al., 2016; Ruskin et al., 2017; Blackie and Kocovski, 2018). Thus, it can be assumed that implicit within the mindfulness approach is a sense of "letting go" or non-striving orientation which serves to obliterate the usual sense of resistance and judgmental stance toward undesired (or even desirable) circumstances and emotions.

Non-striving is an associated quality of mindfulness that is relatively less researched upon compared to other better known correlates such as compassion (Shonin et al., 2015), attention (Mak et al., 2018), and self-esteem (Randal et al., 2015). In English, the definition of striving is "to devote serious effort or energy" (Striving, n.d.). Accordingly, non-striving would mean the absence of devotion of serious effort or energy toward a task. In terms of mindfulness practice, an attitude of non-striving is about non-doing while undertaking the practice, trying less, and simply experiencing the moment (Kabat-Zinn, 1990). Kabat-Zinn (1990) elucidates that adopting a non-striving mind-set means that one is not desiring to change anything related to the moment through mindfulness practice, but instead being simply aware of the ongoing circumstances. This notion of non-striving discussed in secular mindfulness literature possibly roots from works in Eastern and Buddhist philosophy. Specifically, within the Chinese philosophy, the term *wu-wei* or effortless action has the connotation of non-striving as it refers to the harmonized state of mind while one performs actions spontaneously, with freedom from "the need for extended deliberation of inner struggle" (Slingerland, 2003, p. 7). This lack of inner struggle can be construed as non-striving. While discussion on non-striving remains scarce in the secular academic space today, mindfulness scholars such as Baer (2006) and Shapiro et al. (2018), echoing Kabat-Zinn's view, too noted that mindfulness meditation should be practiced with no specific goal in mind, operationalizing the practical notion of non-striving for mindfulness practice to some extent. To further contribute to understanding of non-striving in the context of mindfulness practice, we conducted an experiment to investigate the effects of a brief strange loop task, purported to elicit sense of futility and paradoxity, on subsequent non-striving behavior and performance. We posit that this is an important endeavor that can help further operationalize the notion of non-striving for secular mindfulness practice.

The notion of non-striving in mindfulness practice is a paradox. Goal-oriented instructions for mindfulness practice, such as paying attention to the target object, are typically prescribed; and repeated attempts to adhere to the instruction are undertaken by the practitioner (Baer, 2006). But yet, mindfulness practice necessitates non-judgmental awareness, which could also, in theory, include openness and acceptance of whether there was success in adhering to the goal-oriented instruction. To illustrate, one may begin a mindfulness practice session with the goal of sitting for 30 minutes, with attention purposefully fixated on breathing and adopting a strategy to be aware of passing thoughts without reacting to them. Clearly, this takes some discipline and effort. Paradoxically, to adhere to this instruction well, it is also important to adopt a non-striving attitude toward the task. That is, wanting and trying too hard to achieve a certain "mindful state" can be counter-productive (Shapiro et al., 2018). Taking non-striving to another level, even the desire for relaxation, reduction of pain, or alteration of thoughts and emotions should be downplayed even if those are the reasons for initiating mindfulness practice (Baer, 2006). In sum, Shapiro et al. (p. 1697) describe non-striving in mindfulness practice as "an alert, relaxed attention (which requires effort to develop) without pursuing any specific goal,"

and acknowledge that this attitude is often elusive. For example, the instruction not to strive hard in mindfulness practice may lead one to give up the practice too easily (Hopkins and Proeve, 2013). Suffice to say, non-striving is a nontrivial issue in the study of mindfulness that deserves more research as its coverage is currently limited in the extant literature.

Despite the centrality of non-striving in mindfulness, published academic work discussing non-striving in the context of mindfulness is surprisingly rare. By far, Shapiro et al. (2018) provided the most extensive discussion of non-striving in a narrative review which described paradoxes of mindfulness. In essence, Shapiro et al. (p. 1697) highlighted that practitioners often find it difficult to "… simultaneously allow what arise to arise – not strive to cultivate a particular state of mind – while trying to focus the mind on a particular object of attention…" They further explained that a way out of this paradoxical dilemma is to adopt the middle path, by striving at times and letting go in other instances. A more accurate interpretation would be to strive or to non-strive for different aspects of practice. For example, making the necessary effort to apply the mindfulness technique, while letting go of expectations. The difficulty of actualizing non-striving was documented in a research conducted by Solhaug et al. (2016), where they found that participants who underwent a mindfulness intervention program expressed lesser ease in appreciating the non-striving aspects of mindfulness compared to the attentional aspects. The former is somewhat counterintuitive while the latter is more concrete and less abstract. In essence, such empirical work on non-striving remains scarce. Further research is needed to better understand the operationalization of non-striving within mindfulness practices, particularly for future possible meaningful incorporations in clinical (Roemer and Orsillo, 2003) and other areas of interventions. Particularly, mindfulness practice tasks that can concretely introduce or elicit the experience of non-striving is especially needed.

Given the lack of research on non-striving and practice tasks that can introduce this concept well, the present study investigated the effects of a mindfulness-like practice task aimed at priming non-striving, futility, and process focus, on consequent non-striving behavior and performance. To this end, we created a brief intervention task – "Infinite Water Scooping Task," which we speculate could be useful as a mindfulness practice for introducing non-striving in the future. This task involves using a small scoop to continuously transfer water over a string dividing a filled water container. Essentially, the task is perceptually futile in nature, as water would flow freely beneath the string after pouring, making the effort made in scooping and pouring seem purposeless. We speculate that performing this task would induce non-striving attitude temporarily even when practiced briefly for 10 min, compared to performing the control task of scooping and pouring from one container to another. A person performing the control task would see water level reducing in one container and increasing in another, whereas one performing the Infinite Water Scooping Task would not see incremental changes in water level over time. While both tasks involve essentially performing the same scooping and pouring action, it is intended the Infinite Water Scooping

Task would elicit a mental state that departs from the usual outcome-oriented mental state that would be primed through the control task. The experience of performing the control task is deemed to be consistent with the *modus operandi* of most human as goal driven cognition is a norm in modern societies (e.g., Kenrick et al., 2010). On the other hand, the brief repetition of Infinite Water Scooping Task attempts to elicit non-striving orientation without explicitly instructing and emphasizing non-striving, as would be communicated during typical mindfulness practice.

The creation of the Infinite Water Scooping Task was inspired by the idea of strange loop proposed by Hofstadter (2007). Briefly, strange loop is characterized by self-referentiality and paradoxity. One example of strange loop offered by Hofstadter (2007) is M. C. Escher's lithograph Drawing Hands, which depicts the right hand drawing the left hand, which in turn draws the right hand, forming an infinite loop. Another example is the case of Penrose stairs or impossible stairs created by Lionel Penrose and Roger Penrose, which is a continuously looping staircase that creates the impossible perception of climbing higher (Penrose and Penrose, 1958). In Hofstadter's (2007, p. 101–102) words, strange loop is "... an abstract loop in which, in the series of stages that constitute the cycling-around, there is a shift from one level of abstraction (or structure) to another, which feels like an upwards movement in a hierarchy, and yet somehow the successive 'upward' shifts turn out to give rise to a closed cycle. That is, despite one's sense of departing ever further from one's origin, one winds up, to one's shock, exactly where one had started out. In short, a strange loop is a paradoxical level-crossing feedback loop."

While those above examples of strange loop comprised of optical illusions, the Infinite Water Scooping Task is aimed at eliciting a sense of paradox through a physical experience. In this task, when water is scooped and lifted across the string, there is an upward shift from one level of abstraction to another (where the abstraction refers to the state of the actions, such as where the scooped water is). The shift advances further when the water is poured out. Paradoxically, with this further advancement, water from the scoop merges with the original water source in the container, and the original state is revisited. In theory, this task can continue infinitely, like the continuously looping staircase in Penrose stairs. Hofstadter (2007) noted in the earlier description that some kind of "shock" is experienced in realizing the paradox of returning to the original state despite advancing. We liken this "shock" as the realization of the futility or the paradox of "physically exerting effort to scoop and pour water and yet effecting no change." By continuously performing the Infinite Water Scooping Task for a few minutes, we expect participants to experience a temporary departure from the usual orientation of associating effort with outcome, which is typically ingrained in modern cultures.

We expect the experience of dissociation of effort with outcome through the Infinite Water Scooping Task to have some effects on subsequent non-striving behavior as performing this task continuously for a brief period could elicit psychological effects similar to mindfulness practice. The similarity lies in its repetitive nature, akin to that of, say, continuously paying attention to one's breathing in mindfulness breathing exercises or repetitively watching how one walks in walking meditations. Typically, in mindfulness practice, one simply observes the continuous process unfolds cycle after cycle without expectation of any advancement. In other words, watching the in- and out-breaths, and appreciating that these are the only two states in the cycle, and that there is no need for striving to advance in breathing stages. In the Infinite Water Scooping Task, it is perceptually clear that there is no necessity of striving as there will be no visible change in water level. As the instruction to focus on the process is given concurrently, we therefore speculate that performing the task for 10 min is akin to mindful movement practice. Thus, upon completing this task, a mental state resembling mindfulness state could be activated. Here, we are primarily interested in examining whether there will be indications of weaker willingness to strive in a secondary task as a result of performing the Infinite Water Scooping Task.

In summary, we argued that non-striving is an important aspect of mindfulness practice, but it has not been sufficiently researched. To fill this gap in research, we conducted a randomized experiment to examine whether an essentially futile task designed based on the ideas of strange loop – Infinite Water Scoping Task – performed for 10 min has an effect on subsequent non-striving behavior and performance. Given the likelihood that the psychological experience of performing the Infinite Water Scooping Task is different from transferring water from one container to another (control condition), the main hypothesis tested is that the experimental group participants would spend less time in performing the subsequent word length comparison task due to lower efforts invested in counting carefully (i.e., showing lesser degrees of striving), relative to the control group.

MATERIALS AND METHODS

Participants

Sixty participants comprised of 38 males (*M* age = 24.61, SD = 2.62) and 22 females (*M* age = 24.36, SD = 4.85) took part in the study. Participants were recruited from the university community *via* social media, posters, and word of mouth. The study was approved by the university's institutional review board where the study was conducted. All participants provided informed consent. A SGD 10 (approx. USD 7.5) shopping voucher was given to each participant for his or her involvement in this study. They were randomly assigned to either the experimental or control condition in equal distribution. The methods were carried out in accordance to the guidelines stated by the university's institutional review board.

Brief Intervention Tasks

Common in the manipulation tasks of the experimental and control conditions was the task of scooping and pouring water at one's self-selected pace for 10 min continuously using a small plastic scoop (15 cc in capacity with a 18-cm handle) in a standing position. The difference in conditions was in

terms of the way the water container was set up on the table (i.e., using one or two containers). The video depicting the intervention tasks is provided in the **Supplementary Material** section. Before participants started the water scooping task, a brief condition-specific instruction was delivered to them *via* an Android tablet app in text and in audio script. When the time was up, the app played a chime to indicate the end of the intervention task.

Experimental Condition (Infinite Water Scooping Task)

In the experimental condition, a string was tied across the top of a container with the dimension of 42 (length) × 34.5 (width) × 17 (height) in cm, dividing the left and right of the container equally. The container was half-filled with tap water. Participants were tasked to scoop water from one side of the container, lifting the scoop over the string before pouring it back into the container. The task was meant to be perceptually futile and purposeless, because no drastic change to the water level would be observed as water flows across freely under the string within the same container. The instruction given to them *via* the app was to pour water using the scoop across the string at their own pace and to focus on the process during the task.

Control Condition

In the control condition, two containers, each with dimensions of 27 (length) × 20 (width) × 16.5 (height) cm were placed side by side. One of them was half-filled with tap water. Participants were tasked to scoop water from one container, and then pour it into the other container. The instruction given to them by the app was to pour water using the scoop across two containers at their own pace and that the task is practiced as a means of developing wrist control and strength. The control task was an appropriate match for the experimental task in terms of motor execution, effort, and timing. The effort applied was not futile and changes in water level can be observed. In contrast with the experimental task, the control task can be deemed as a common task.

Secondary Task: Word Length Comparison Task

To detect the extent to which participants display non-striving behavior after undergoing the respective manipulations, we used a secondary task adapted from the task outlined by Touré-Tillery and Fishbach (2012) in their Experiment 5 and implemented it using OpenSesame (Mathôt et al., 2012) on a laptop computer. In this task, within each trial, two English words were shown separately on left and right sides of the screen. The task requires participants to choose the word with fewer letters by responding accordingly on the keyboard. There was no emphasis on timing and accuracy. The task is challenging as each trial had a pair of words that were different in length only by a letter. For example, "outstretches" vs. "repolarized," "fishburger" vs. "assignation," "birthweight" vs. "utterances," etc. Words used were between 9 and 12 letters long, and were generated randomly using a computer script

beforehand. Trials comprised of longer words would present more difficulty relative to trials with shorter words, by virtue of effort needed to count them.

The entire task comprised of two practice trials and seven blocks of six trials each (42 actual trials). We derived two types of measures from this task for each participant. First, the sum of correctly performed trials was calculated to serve as a measure of task performance. Second, the mean time (in ms) spent on each trial was calculated for all trials performed by each participant. We further derived the individualized mean time spent on trials that featured (1) 9 and 10 letters, (2) 10 and 11 letters, and (3) 11 and 12 letters, such that between-group comparisons can be made based on performances for trials of similar difficulty. This allows us to verify if the observed results were consistent regardless of task difficulty. The time taken to respond to the task was used as a proxy for striving as it reflects one's effort to count the letters accurately before making a decision. Conversely, lesser time spent reflects non-striving as decisions were likely made without careful counting. As we did not emphasize on the speed or accuracy requirements, participants who strove to perform the task correctly might be more likely to be counting the words rather than estimating to get the correct answer. They were also told to move at their own pace and to feel free to take short breaks between blocks of trials if needed. In a nutshell, the salient instruction given to the participants was to pick the shorter word without time pressure. With that, we expected that the general speed-accuracy trade-off phenomenon predicted by Fitts' law would naturally occur (Fitts, 1954), i.e., those who responded faster would tend to sacrifice accuracy, and vice versa. The experimental group, hypothesized to be non-striving, should respond faster but would have less correct trials, indicative of their lack of effort in getting the task right, if they simply adhered to the instruction of picking the shorter word without trying to be fast in responding.

Procedure

Participants were individually tested in a quiet room and each session lasted for approximately 30 min. When participants arrived for the study, they were randomly assigned to one of the two conditions. After providing informed consent and given the shopping voucher, they proceeded to receive the task instructions and performed the water scooping task for 10 min according to the condition assigned, following which, they proceeded to complete the word length comparison task. Finally, participants were debriefed and thanked for their participation.

Data Analysis

The initial step involved identifying and removing the outliers among the participants by detecting peculiar performances, such as those resulting from misinterpretation of task instructions. Next, we tabulated the descriptive statistics for sample-wise mean count of correctly identified trials and mean completion time spent. Data normality was examined using Shapiro-Wilk

tests. Inferential statistical tests were performed using the robust Yuen's t test for trimmed means, with bootstrapping set at 2,000, and trim level for the mean set as 0.20, as recommended in Field and Wilcox, 2017. Tests for group differences were conducted for the following dependent variables: number of correct trials (proxy for task performance), and mean time taken for completing each trial based on all 42 trials as well as that of trials with 9 and 10 letters, 10 and 11 letters, and 11 and 12 letters (proxies for non-striving to test the main hypothesis). The reason for analyzing the trials of similar word length separately is that shorter sets of words could take lesser striving/effort to perform while longer sets of words could require more effort, by virtue of total length presented. The additional analyses serve to determine whether group differences observed are present regardless of the word length used. Additionally, we also repeated the analyses for mean completion time of trials that were correctly performed to countercheck if the result was consistent with the earlier analyses based on all trials (correct and incorrect trials). The *yuenbt* function from R package WRS2 (Mair and Wilcox, 2015) was used for the robust statistical analysis. In this package, akp effect size which is a robust version of Cohen's d, proposed by Algina et al. (2005), was used. The same rules of thumb as for Cohen's d can be used to interpret the effect sizes; that is, 0.2, 0.5, and 0.8 correspond to small, medium, and large effects, respectively.

RESULTS

Descriptive Statistics

Among the 60 participants, data from three participants were removed from further analysis as they achieved less than two correct trials in the word length comparison task. They failed to adhere to the given task instructions. This resulted in 27 participants in the control group and 30 participants in the experimental group. Of these remaining 57 participants, the mean count of correctly identified trials is 39.02 out of 42 (92.90%), and ranged from 34 to 42 trials. The mean trial completion time is 3,876 ms, and ranged from 786 to 10,770 ms. Results of Shapiro–Wilk test of normality for counts of correct response ($p = 0.004$) and time spent on each correct answer ($p < 0.001$) suggest that the distribution of the data is significantly different from normal distribution, i.e., normal distribution cannot be assumed. The decision to rely on robust approach for inferential statistical analysis based on Yuen's modified t test for independent trimmed means with bootstrapping (Field and Wilcox, 2017) was thus made.

Inferential Statistics

Before testing the main hypothesis, we compared the number of correct trials between the two groups to ascertain if there was a difference in task performance (based on the number of correct trials). The robust Yuen's t test for trimmed means (bootstrapping set at 2,000, and trim level for the mean as 0.20) detected no significant difference between trimmed means in

number of correct trials for the control group and the experimental group, $M_{diff} = 1.27$ (-0.03, 2.56), $Y_t = 2.02$, $p = 0.055$. However, it can be interpreted that the value of p is close to the critical value for significance difference. **Figure 1** suggests that the number of correct trials in the control group tended to be higher than that of the experimental group. The effect size was found to be 0.57, which was slightly larger than a medium effect size.

The main hypothesis that degree of non-striving differed between the two conditions was tested based on mean time spent on trials as a proxy measure for non-striving. When all 42 trials were considered, the robust Yuen's t test for trimmed means (bootstrapping set at 2,000, and trim level for the mean as 0.20) showed that there was a significant group difference, $M_{diff} = 2358.35$ (573.48, 4143.21), $Y_t = 2.65$, $p = 0.016$. **Figure 2**
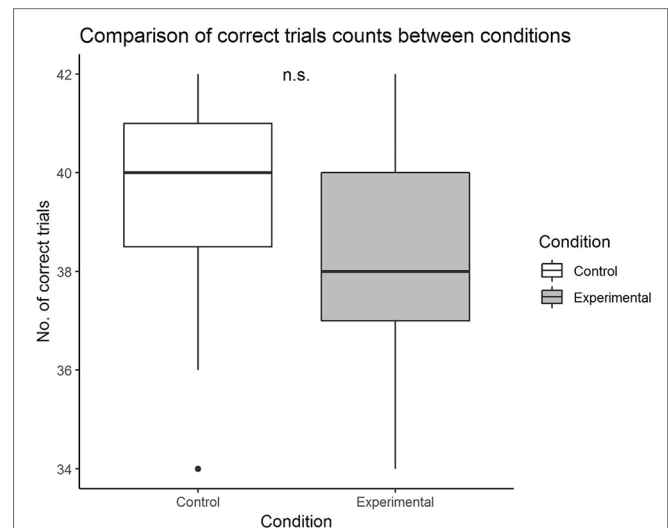

FIGURE 1 | Comparison of number of correct trials between control and experimental groups.
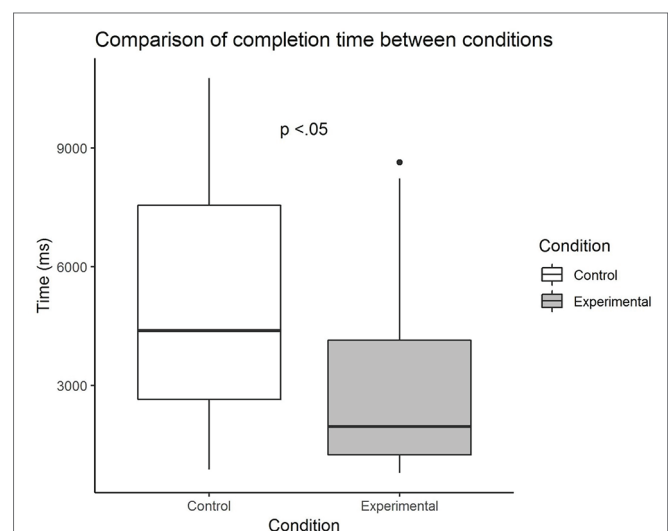
FIGURE 2 | Comparison of mean time spent on each trial between control and experimental groups.

shows that participants in the experimental group spent significantly less time working on each trial compared to those in the control group. The effect size was found to be 0.78, which was close to a large effect size. The combination of results above shows that the control group took a longer time and yielded more correct trials, compared to the experimental group which took a shorter time but registered less correct trials. This suggests that the expected speed-accuracy trade-off occurred with the mere instruction to pick the shorter word without time pressure.

Further tests were undertaken to ascertain that the observed differences in time spent were also present for trials of similar difficulty (i.e., trials with words that had the same total number of letters). When trials with words of 9 and 10 letters are considered, the Yuen's t test for trimmed means (bootstrapping set at 2,000, and trim level for the mean as 0.20) showed that there was a significant group difference, M_{diff} = 1871.91 (171.56, 3572.26), Y_t = 2.39, p = 0.031. The effect size of 0.70 was observed. For trials with words of 10 and 11 letters, the result again revealed a significant difference between groups, M_{diff} = 2378.98 (414.04, 4343.91), Y_t = 2.53, p = 0.017. The effect size was found to be 0.74. Lastly, for trials with words of 11 and 12 letters, the result is similar in that a significant difference between groups was observed, M_{diff} = 2693.13 (712.90, 4673.36), Y_t = 2.76, p = 0.010. The effect size value was large at 0.80. In all cases, the experimental group spent lesser time than the control group as depicted in **Figure 3**.

Lastly, as the analyses above were based on response timings of all trials, regardless of whether they were correctly or incorrectly performed, we repeated these analyses based on response timings of trials that were correctly performed to remove the possible noise in the data due to incorrect trials. Similar findings were observed. This additional observation suggests that the experimental effects are present even when we removed trials that were performed incorrectly due to the participants giving up or responding carelessly on certain trials. The effects of incorrect trials were negligible.

FIGURE 3 | Comparison of mean time spent on each trial between control and experimental groups for three difficulty levels.
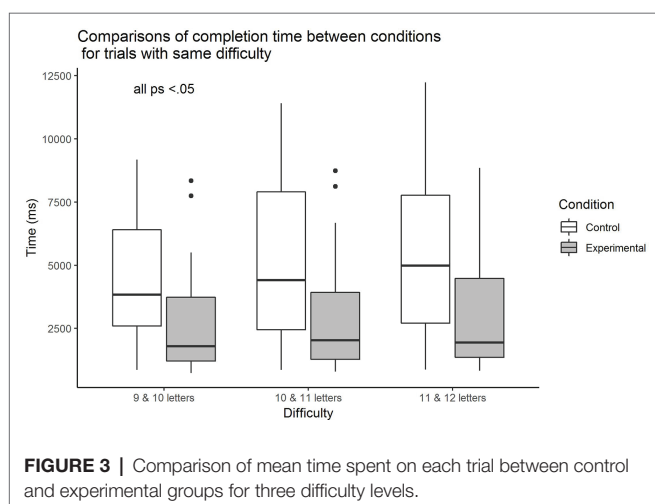
DISCUSSION

As there is a dearth of empirical research on the non-striving aspect of mindfulness, we tested the effects of a presumably futile strange loop task that has the potential of being developed as a mindfulness task, on subsequent non-striving behavior and performance. Although there was no significant group difference observed in terms of performance (based on total number of correct trials achieved), the overall results (close to significant p value and medium effect size) suggest that the control group showed tendencies of better performance than the experimental group. It is likely that those in the control group applied more effort and adopted the counting strategy to achieve the accuracy while the experimental group tended not to. Findings from the test of the main hypothesis suggest that participants who underwent the Infinite Water Scooping Task took significantly shorter time to perform the task compared to the control group. The effect was observed when all trials were considered, and when trials of similar difficulty level were considered. Collectively, our hypothesis that performing the futile strange loop task would prime subsequent non-striving behavior is supported.

The current outcome of eliciting non-striving behavior is noteworthy as non-striving is an important aspect of mindfulness practice that is rarely examined in research. Here, we observed that 10 min of repetitively scooping and pouring water over a string tied across a container, peppered only with a brief initial reminder to focus on the process of the actions (without mention of it being a mindfulness practice), led to non-striving consequently when compared to the control task. One speculation could be that the implicit realization of task's futility or paradoxity arising from repetitive performance of the strange loop task (Hofstadter, 2007) is having an effect on one's tolerance of ambiguity. Having to repeatedly scoop and pour water for 10 min without perceiving any sign of advancement may have gradually activated a sense of acceptance toward the ambiguities of the task. The inner dialogue could be something like: "this task is paradoxical in nature and does not make sense, but I continue with it nevertheless." A sense of acceptance may have been resulted through repeated actions required of the task, akin to acceptance of unfoldment of in- and out-breaths in mindful breathing practices. Following the completion of this task, this tolerance for ambiguity or paradox, in turn, may have lowered one's efforts to judge critically (or count carefully) during the subsequent task. Previously, Carson and Langer (2006) noted that actively thinking about paradoxes could increase one's ability to tolerate ambiguity, which they viewed as a hallmark of mindfulness. In a way, performing the Infinite Water Scooping Task for a few minutes possibly succeeded in exposing the participants to a sense of paradox, akin to the sense of perceiving paradoxes when contemplating Zen koans (Christopher, 2003; Maex, 2011), which has been linked to psychological awakening.

In the present context, by psychological awakening, we mean leading participants out of their fixed ways of thinking or habits of striving, at least temporarily. Since the control group performed an objective-oriented task (i.e., clearly perceiving changes in water levels as a result of their action, and being told that the task is a means of developing wrist control and strength), performing

the task could have heightened the sense of purpose-driven thinking and certainty that is implicitly habitual in most people (Covington, 2000; Monteiro et al., 2018; Ronkainen et al., 2018). Upon completing the control task that is perceptually more normal and non-ambiguous, the tendency to count the letters purposefully to achieve better accuracy in the subsequent word length comparison task may have been more readily evoked, as shown by the longer trial durations. Based on the shorter trial duration observed, it appears that those in the experimental group tended not to count before making a response. If they were striving to achieve task accuracy, they would have spent more time in the task by counting the letters carefully. As earlier alluded to, shorter time spent can be viewed as a sign of their willingness to tolerate ambiguity and errors in their task performance. They are striving less toward achieving accuracy as per required by the task. In a way, those in the experimental group may be less judgmental about themselves in terms of whether they performed the task accurately, while those in the control group seem more objective-oriented about their task performance. Taken together, the Infinite Water Scooping Task seems to be instrumental in leading participants out of their habits of striving, at least temporarily, when compared to the control task.

There are some practical implications arising from the current findings worthy of mention. First, given the current results, the Infinite Water Scooping Task could potentially be used to introduce the notion of non-striving within mindfulness practice in various contexts, such as when teaching mindfulness in clinical interventions (Roemer and Orsillo, 2003), in education (McKeering and Hwang, 2019), and in sport as part of mental skills training (Ortega and Wang, 2018). For instance, the Infinite Water Scooping Task can be used as an introductory task to teach what is meant by notions of repetition, focus on process, and purposelessness within mindfulness practice, relative to a task that is done with a purpose. Beginning mindfulness practitioners could especially benefit by experiencing the nuts and bolts of mindfulness practice through the physical nature of the task (Kee, in press), alleviating some of the difficulty raised by Solhaug et al. (2016). Second, the possibility of eliciting one's tolerance for ambiguity and non-striving through the Infinite Water Scooping Task may be useful when it comes to helping one prepare for performance or creative situations when a creative, relaxed, and open mind-set is needed. For example, in sports, some athletes described their best performance during a flow experience (a peak psychological state in which one is fully immersed in an activity experience) as one of intense focus without making effort in keeping focused (Jackson and Roberts, 1992). The relaxed and non-striving experience accompanied by the Infinite Water Scooping Task or other similar tasks may help athletes navigate through the nuances accompanying such peak psychological states when they use it as part of their mental skills repertoires. For example, an archer can use this task to get him/herself mentally prepared to focus on the process and not the outcome for his/her shoot. Likewise, this task could add to the list of other known strategies for enhancing creativity such as one described and tested by Chirico et al. (2018), in that this task may help one to suspend judgment momentarily to allow creativity to flow. Lastly, beyond these two specific implications, there is potential for the mindful repetition

of the Infinite Water Scooping task to be used as a mindfulness practice task for those who find difficulty in performing seated meditation, given that it is an overt task.

Although the findings and implications seem promising, there are some limitations that must be noted. First, as no manipulation check was conducted, we cannot be certain that the aforementioned non-striving or paradoxical effects were indeed evoked. One main reason for not conducting a manipulation check immediately after participants completed the brief intervention task is the concern that such check (e.g., using questions) may result in unexpected priming of psychological effects that affect the performance of secondary task. While we maintain that the paradoxity explanation is plausible for the experimental task given that the design principle of the Infinite Water Scooping Task took reference from concepts of strange loop (Hofstadter, 2007), we acknowledge that a better approach would be to conduct a manipulation check after the secondary task. Secondly, although we argued that the control task is aligned with participants' *modus operandi* of being goal driven and performing the control task could be treated as typical experiences, the control condition can potentially prime goal-focused orientation, and thus the possibility of participant striving as a result cannot be ruled out. It would be worth considering including a neutral control task that does not prime striving and non-striving in future studies. Thirdly, although the main difference between the experimental and control condition is essentially pouring across a string within the same container compared to pouring across containers, the brief instruction of focusing on the process was given to the experimental group but not to the control group. This raises the question whether the observed effects came from the instruction to focus on the process. Although the likelihood is slim as the instruction was only given at the start in a brief fashion, future studies could consider keeping the instructions to focus on the process consistent for both groups to rule out this possibility. Fourthly, there could be an alternative interpretation in that those who underwent the Infinite Water Scooping Task actually strove harder than the control group, since they could have completed the word length comparison task faster due to them putting in more efforts to respond sooner without compromising on overall performance. It is a limitation in that there was no measure of perceived efforts. Nevertheless, since there was no explicit instruction to complete the task fast, we expected participants to direct their effort toward responding correctly as required by the task. The observed speed-accuracy trade-off supports this view. We maintain that the control group indeed made more efforts to count carefully compared to the experimental group, judging by the seemingly better task performance in the control group. In the future, potential mediators such as a perceived effort, drowsiness, sleepiness, and apathy can also be examined.

In summarizing the lessons learned in terms of designing and conducting this research on non-striving, we conclude that non-striving *per se* is very difficult to operationalize, which perhaps explains the lack of such works in the literature. The difficulty roots from the fact that, by definition, non-striving is the antithesis of striving. To understand non-striving without making reference to striving is not possible in some sense. Experimentally, it is

difficult to create a neutral control task that is completely devoid of any form of striving to compare non-striving with. Any form of active control task would necessarily involve some level of striving, simply because there would be something to be done. That inevitably primes striving. On the other hand, adopting a passive task of sitting still as a control task is also not suitable as a control condition, because the non-striving condition such as the Infinite Water Scooping task would then be construed as requiring more striving than the control task. To this end, we propose a possible approach to make further inroads into understanding non-striving by focusing on the nature of the secondary task rather than on the manipulations. That is, by examining whether non-striving can manifest even in a seemingly effortless task. Putting it in the context of the present experimental setup, we can get closer to know if we effected non-striving if performance on word length comparison task involving very easy tasks (say, obviously long versus short words) is different for those undergoing control and experimental treatments. The assumption is that easy tasks take very minimal striving to perform. If the Infinite Water Scooping task resulted in even lesser striving for easy tasks, that could be a clearer indication of non-striving. Future studies could explore this approach to advance the operationalization of non-striving.

As mindfulness research advances over the years, many constructs related to mindfulness were examined to better understand mindfulness (Kee et al., 2019). Since works on non-striving have been especially scarce, the current study presents an initial foray into the empirical examination of non-striving in the context of mindfulness practice. The main takeaway message is that the Infinite Water Scooping Task seems instrumental in eliciting non-striving (relative to the control task), which may be useful for teaching mindfulness practitioners about the notion of non-striving. Clearly, more works in examining the nature and practical value of non-striving is warranted. Beyond investigating non-striving in the context of mindfulness, non-striving as a research topic is also relevant for line of research questioning the values of objectives and goals setting, such as works championed by Ordóñez et al. (2009) and Swann et al. (in press). There is a possibility that the line of research questioning the values of goals and non-striving through mindfulness may converge, hopefully illuminating further lessons beyond the duality

of striving and non-striving. The conversations arising from the testing of Infinite Water Scooping Task may also contribute to discussions on altered sense of consciousness-related issues such as *wu-wei*, flow, mystical experiences, and awakening, since there had been discussion of "letting go," non-striving, and such states in neuroscience literature (Austin, 2006, p. 275).

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 Nanyang Technological University – Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02314/full#supplementary-material

REFERENCES

Algina, J., Keselman, H. J., and Penfield, R. D. (2005). An alternative to Cohen's standardized mean difference effect size: a robust parameter and confidence interval in the two independent groups case. *Psychol. Methods* 10, 317–328. doi: 10.1037/1082-989X.10.3.317

Austin, J. H. (2006). *Zen-brain reflections*. Cambridge, MA: MIT Press.

Baer, R. A. (2006). Mindfulness training as a clinical intervention: a conceptual and empirical review. *Clin. Psychol. Sci. Pract.* 10, 125–143. doi: 10.1093/clipsy/bpg015

Bergeron, C. M., Almgren-Doré, I., and Dandeneau, S. (2016). "Letting go" (implicitly): priming mindfulness mitigates the effects of a moderate social stressor. *Front. Psychol.* 7:872. doi: 10.3389/fpsyg.2016.00872

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., et al. (2004). Mindfulness: a proposed operational definition. *Clin. Psychol. Sci. Pract.* 11, 230–241. doi: 10.1093/clipsy.bph077

Blackie, R. A., and Kocovski, N. L. (2018). Forgive and let go: effect of self-compassion on post-event processing in social anxiety. *Mindfulness* 9, 654–663. doi: 10.1007/s12671-017-0808-9

Carson, S. H., and Langer, E. J. (2006). Mindfulness and self-acceptance. *J. Ration. Emot. Cogn. Behav. Ther.* 24, 29–43. doi: 10.1007/s10942-006-0022-5

Chirico, A., Glaveanu, V. P., Cipresso, P., Riva, G., and Gaggioli, A. (2018). Awe enhances creative thinking: an experimental study. *Creat. Res. J.* 30, 123–131. doi: 10.1080/10400419.2018.1446491

Christopher, M. S. (2003). Albert Ellis and the Buddha: rational soul mates? A comparison of rational emotive behaviour therapy (REBT) and Zen Buddhism. *Mental Health, Religion & Culture* 6, 283–293. doi: 10.1080/1367467031000100975

Covington, M. V. (2000). Goal theory, motivation, and school achievement: an integrative review. *Annu. Rev. Psychol.* 51, 171–200. doi: 10.1146/annurev.psych.51.1.171

Field, A. P., and Wilcox, R. R. (2017). Robust statistical methods: a primer for clinical psychology and experimental psychopathology researchers. *Behav. Res. Ther.* 98, 19–38. doi: 10.1016/j.brat.2017.05.013

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.* 47, 381–391. doi: 10.1037/h0055392

Frewen, P. A., Evans, E. M., Maraj, N., Dozois, D. J. A., and Partridge, K. (2008). Letting go: mindfulness and negative automatic thinking. *Cogn. Ther. Res.* 32, 758–774. doi: 10.1007/s10608-007-9142-1

Hinton, D. E., Pich, V., Hofmann, S. G., and Otto, M. W. (2013). Acceptance and mindfulness techniques as applied to refugee and ethnic minority populations with PTSD: examples from "culturally adapted CBT". *Cogn. Behav. Pract.* 20, 33–46. doi: 10.1016/j.cbpra.2011.09.001

Hofmann, S. G., and Asmundson, G. J. G. (2008). Acceptance and mindfulness-based therapy: new wave or old hat? *Clin. Psychol. Rev.* 28, 1–16. doi: 10.1016/j.cpr.2007.09.003

Hofstadter, D. R. (2007). *I am a strange loop*. New York, NY: Basic Books.

Hopkins, A., and Proeve, M. (2013). Teaching mindfulness-based cognitive therapy to trainee psychologists: qualitative and quantitative effects. *Couns. Psychol. Q.* 26, 115–130. doi: 10.1080/09515070.2013.792998

Jackson, S. A., and Roberts, G. C. (1992). Positive performance states of athletes: toward a conceptual understanding of peak performance. *Sport Psychol.* 6, 156–171. doi: 10.1123/tsp.6.2.156

Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your mind to face stress, pain and illness*. New York: Dell Publishing.

Kee, Y. H. (in press). Reflections on athletes' mindfulness skills development: Fitts and Posner's (1967) three stages of learning. *Journal of Sport Psychology in Action*, 1–6. doi: 10.1080/21520704.2018.1549640

Kee, Y. H., Li, C., Kong, L. C., Tang, C. J., and Chuang, K. L. (2019). Scoping review of mindfulness research: a topic modelling approach. *Mindfulness* 10, 1474–1488. doi: 10.1007/s12671-019-01136-4

Kenrick, D. T., Neuberg, S. L., Griskevicius, V., Becker, D. V., and Schaller, M. (2010). Goal-driven cognition and functional behavior: the fundamental-motives framework. *Curr. Dir. Psychol. Sci.* 19, 63–67. doi: 10.1177/0963721409359281

Lindsay, E. K., and Creswell, J. D. (2017). Mechanisms of mindfulness training: monitor and acceptance theory (MAT). *Clin. Psychol. Rev.* 51, 48–59. doi: 10.1016/j.cpr.2016.10.011

Maex, E. (2011). The Buddhist roots of mindfulness training: a practitioners view. *Contemporary Buddhism* 12, 165–175. doi: 10.1080/14639947.2011.564835

Mair, P., and Wilcox, R. (2015). Robust statistical methods using WRS2. Available at: https://cran.r-project.org/web/packages/WRS2/vignettes/WRS2.pdf (Accessed February 2, 2019).

Mak, C., Whittingham, K., Cunnington, R., and Boyd, R. N. (2018). Efficacy of mindfulness-based interventions for attention and executive function in children and adolescents: a systematic review. *Mindfulness* 9, 59–78. doi: 10.1007/s12671-017-0770-6

Mathôt, S., Schreij, D., and Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav. Res. Methods* 44, 314–324. doi: 10.3758/s13428-011-0168-7

McKeering, P., and Hwang, Y.-S. (2019). A systematic review of mindfulness-based school interventions with early adolescents. *Mindfulness* 10, 593–610. doi: 10.1007/s12671-018-0998-9

Monteiro, D., Pelletier, L. G., Moutão, J., and Cid, L. (2018). Examining the motivational determinants of enjoyment and the intention to continue of persistent competitive swimmers. *Int. J. Sport Psychol.* 49, 484–504. doi: 10.7352/IJSP.2018.49.484

Ordóñez, L. D., Schweitzer, M. E., Galinsky, A. D., and Bazerman, M. H. (2009). Goals gone wild: the systematic side effects of overprescribing goal setting. *Acad. Manag. Perspect.* 23, 6–16. doi: 10.5465/amp.2009.37007999

Ortega, E., and Wang, J. C. K. (2018). Effectiveness of an integrated mental skills and biofeedback training program on sport shooters. *Int. J. Sport Psychol.* 49, 35–54. doi: 10.7352/IJSP2018.49.035

Penrose, L. S., and Penrose, R. (1958). Impossible objects: a special type of visual illusion. *Br. J. Psychol.* 49, 31–33. doi: 10.1111/j.2044-8295.1958.tb00634.x

Randal, C., Pratt, D., and Bucci, S. (2015). Mindfulness and self-esteem: a systematic review. *Mindfulness* 6, 1366–1378. doi: 10.1007/s12671-015-0407-6

Roemer, L., and Orsillo, S. M. (2003). Mindfulness: a promising intervention strategy in need of further study. *Clin. Psychol. Sci. Pract.* 10, 172–178. doi: 10.1093/clipsy.bpg020

Ronkainen, N. J., Tikkanen, O. M., and Nesti, M. S. (2018). Vocation: a concept for studying meaningful lives and careers in sport. *Int. J. Sport Psychol.* 49, 291–310. doi: 10.7352/IJSP.2018.49.291

Ruskin, D., Harris, L., Stinson, J., Kohut, S. A., Walker, K., and McCarthy, E. (2017). "I learned to let go of my pain". The effects of mindfulness meditation on adolescents with chronic pain: an analysis of participants' treatment experience. *Children* 4:110. doi: 10.3390/children4120110

Shapiro, S., Siegel, R., and Neff, K. D. (2018). Paradoxes of mindfulness. *Mindfulness* 9, 1693–1701. doi: 10.1007/s12671-018-0957-5

Shonin, E., Van Gordon, W., Compare, A., Zangeneh, M., and Griffiths, M. D. (2015). Buddhist-derived loving-kindness and compassion meditation for the treatment of psychopathology: a systematic review. *Mindfulness* 6, 1161–1180. doi: 10.1007/s12671-014-0368-1

Slingerland, E. (2003). *Effortless action: Wu-wei as conceptual metaphor and spiritual ideal in early China*. New York, NY: Oxford University Press.

Solhaug, I., Eriksen, T. E., de Vibe, M., Haavind, H., Friborg, O., Sørlie, T., et al. (2016). Medical and psychology student's experiences in learning mindfulness: benefits, paradoxes, and pitfalls. *Mindfulness* 7, 838–850. doi: 10.1007/s12671-016-0521-0

Striving (n.d.). In Merriam-Webster's Collegiate Dictionary. Available at: https://www.merriam-webster.com/dictionary/striving (Accessed September 1, 2019).

Swann, C., Hooper, A., Schweickle, M. J., Peoples, G., Mullan, J., Hutto, D., et al. (in press). Comparing the effects of goal types in a walking session with healthy adults: preliminary evidence for open goals in physical activity. *Psychol. Sport Exerc.* doi: 10.1016/j.psychsport.2019.01.003

Touré-Tillery, M., and Fishbach, A. (2012). The end justifies the means, but only in the middle. *J. Exp. Psychol. Gen.* 141, 570–583. doi: 10.1037/a0025928

Neural Processes of Proactive and Reactive Controls Modulated by Motor-Skill Experiences

Qiuhua Yu[1,2], Bolton K. H. Chau[2], Bess Y. H. Lam[2], Alex W. K. Wong[3,4], Jiaxin Peng[2,5] and Chetwyn C. H. Chan[2,6]*

[1] Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [2] Applied Cognitive Neuroscience Laboratory, Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, China, [3] Program in Occupational Therapy, Washington University School of Medicine, St. Louis, MO, United States, [4] Department of Neurology, Washington University School of Medicine, St. Louis, MO, United States, [5] Department of Education, Shaoguan University, Shaoguan, China, [6] University Research Facility in Behavioral and Systems Neuroscience, The Hong Kong Polytechnic University, Hong Kong, China

This study investigated the experience of open and closed motor skills on modulating proactive and reactive control processes in task switching. Fifty-four participants who were open-skilled ($n = 18$) or closed-skilled athletes ($n = 18$) or non-athletic adults ($n = 18$) completed a cued task-switching paradigm task. This task tapped into proactive or reactive controls of executive functions under different validity conditions. Electroencephalograms of the participants were captured during the task. In the 100% validity condition, the open-skilled participants showed significantly lower switch cost of response time than the closed-skilled and control participants. Results showed that the open-skilled participants had less positive-going parietal cue-locked P3 in the switch than repeat trials. Participants in the control group showed more positive-going cue-locked P3 in the switch than repeat trials, whereas the closed-skilled participants had no significant differences between the two types of trials. In the 50% validity condition, the open- and closed-skilled participants had less switch cost of response time than the control participants. Participants in the open- and closed-skilled groups showed less positive-going parietal stimulus-locked P3 in the switch than repeat trials, which was not the case for those in the control group. Our findings confirm the dissociation between proactive and reactive controls in relation to their modulations by the different motor-skill experiences. Both proactive and reactive controls of executive functions could be strengthened by exposing individuals to anticipatory or non-anticipatory enriched environments, suggesting proactive and reactive controls involved in motor-skill development seem to be transferable to domain-general executive functions.

Keywords: proactive control, reactive control, task switching, ERP, motor skills

INTRODUCTION

A high level of motor skills has been associated with improved executive functions. Sports are physical activities that require high levels of motor skills. Individuals who engaged in fencing (Chan et al., 2011), baseball (Kida et al., 2005), and soccer were reported to have better executive functions than the controls (Verburgh et al., 2014). Researchers have suggested that the enhancement in

executive functions could be related to the neural plasticity brought about by the long-term aerobic fitness and cognitive trainings involved in these sports (Voss et al., 2010; Chan et al., 2011). Common features of the physical activities described above are playing facing opponents and in changing external environments, which demand open motor skills. Open motor skills involve generation of physical responses to dynamically and externally paced environment (Wang et al., 2013a; Yu et al., 2017). Contrary to open motor skills is closed motor skills, which requires participants to generate responses that are relatively consistent, stationary, and self-paced (Wang et al., 2013a; Yu et al., 2017). The typical physical activities involving closed motor skills are swimming, and track and field. In view of the differences between open and closed motor skills, it is intuitive that participants of "closed" physical activities would have gained lower level of executive functions than those of "open" physical activities. For instance, open-skilled participants were revealed to have higher levels of inhibitory control than the closed-skilled counterparts (Yamashiro et al., 2015). Nevertheless, in an earlier behavioral study, we reported that, by employing a dual cognitive control model for executive functions (Chevalier et al., 2015; Tarantino et al., 2016), both the open- and closed-skilled participants showed significantly higher levels of reactive control of executive functions than the controls (Yu et al., 2017). The differences between the two experimental groups are in the proactive versus reactive control. In this paper, we aimed to employ electroencephalogram to understand the neural mechanisms underlying how different types of motor skills would associate with proactive or reactive controls among participants engaging in physical activities.

Proactive control is an early selection process, which can optimally bias attention, perception, and action systems in a goal-driven manner (Braver, 2012). Information selected early in the process is to be deployed before the response-demanding event occurs. In contrast, reactive control is to resolve the interference imperatively after the response-demanding event appears (Braver, 2012). Physical activities dominated by open motor skills involve predictions of outcomes upon the actions of opponents and teammates for producing rapid and accurate responses (Jin et al., 2011; Abreu et al., 2012; Yu et al., 2016). These cognitive processes are comparable to those described in proactive control. In contrast, the physical activities dominated by closed motor skills less involve the prediction of actions of others before giving the responses. In addition, open motor skills require the participants to rapidly inhibit inappropriate actions, and switch from an intended movement to an appropriate one in an unpredictable environment (Taddei et al., 2012; Yu et al., 2017). The imperative inhibition and switching process are comparable to those described in reactive control. Tsai and Wang (2015) explored the effects of open- (e.g., badminton and table tennis) or closed-skilled (e.g., jogging and swimming) training in the reactive control of task switching for elderly subjects. Their results showed that open-skilled group had larger P3 amplitude in switch condition than closed-skilled and control groups, but had comparable P3 amplitude in repeat condition with the other two groups. All the studies revealed employed a unidimensional perspective of executive control. The results are

that open-skilled groups had higher level of executive functions in reactive control associated with enhanced P3 than closed-skilled groups. No study has been conducted to explore and explain the possible gains in different time processes of executive functions differentially in the open- and closed-skilled groups. The dual cognitive control model of executive function offers a theoretical basis for addressing the potential differentiation in the cognitive gains due to the engagement in these physical activities.

Neural processes related to proactive and reactive controls can be examined by using a cued task-switching paradigm with electroencephalogram (EEG). The cued task-switching paradigm involves participants predicting a switch of task rule based on the information embedded in the cue and subsequently giving a response according to the new rule. The task cue can be fully predictive (100% validity) or fully non-predictive (50% validity), which makes possible elicitation of the event-related potentials (ERPs) for reflecting the proactive or reactive control processes, respectively. The common ERPs associated with proactive control in the task-switching paradigm reported are the parietally distributed P3 elicited by the cue (called cue-locked P3) (Gajewski and Falkenstein, 2011; Tarantino et al., 2016) and the frontocentrally distributed contingent negative variant (CNV) (Tarantino et al., 2016). In contrast, the frontally distributed N2 (Hsieh and Wu, 2011) and parietally distributed P3 (called stimulus-locked P3) (Scisco et al., 2008; Tarantino et al., 2016) elicited by the stimulus were reported to associate with the reactive control in task switching. The cue-locked P3 can be identified within the 300–600 ms time-window after the appearance of a predictive cue (West et al., 2012; Tarantino et al., 2016). This component was suggested to reflect task reconfiguration – anticipatory updating of task goals and/or action rules in working memory (Nicholson et al., 2006; Gajewski and Falkenstein, 2011). The CNV is a slow wave elicited prior to the onset of the target stimulus (Funderud et al., 2013). It reflects anticipatory attention and motor preparation for the upcoming target stimulus (Funderud et al., 2013; Grane et al., 2016). Thus, the amplitude of CNV could be modulated by the response-related parameters embedded in the task cue (Scheibe et al., 2009; Linssen et al., 2013), e.g., the cue validity. Previous studies also showed that Bereitschaftspotential (BP), comparable to CNV, was more negative-going in athletes than non-athletes (Bianco et al., 2017a,b). This finding suggested that the athletes would have better motor preparation than the non-athletes. However, Wang et al. (2013b) did not reveal significant differences in motor preparation between the open- and closed-skilled players.

The N2 component has been employed as a marker reflecting reactive control – suppression of conflict responses (Rushworth et al., 2002; Hsieh and Wu, 2011). More negative-going frontal N2 was shown to associate with the switching to a new response set. The stimulus-locked P3 can be identified after the appearance of a response-demanding stimulus (Swainson et al., 2006; Tarantino et al., 2016). In the predictive cue condition, the stimulus-locked P3 accounts for stimulus-response set implementation (Jamadar et al., 2010; Gajewski and Falkenstein, 2011; Tarantino et al., 2016) or task-specific evaluation of a target stimulus (Swainson et al., 2006). The amplitude difference of stimulus-locked P3 between switch and repeat trials was negatively related to the switch cost

of the response time when the cue was predictive (Li et al., 2012), suggesting that stimulus-locked P3 in reactive control was related to the performance in task switching. In the non-predictive cue condition, stimulus-locked P3 is associated with updating of the task goal or task rules (Hillman et al., 2006; Scisco et al., 2008). It was more positive-going in the switch than repeat trials, because more attentional resources required for subsequent memory updating in reactive control (Hillman et al., 2006; Scisco et al., 2008). However, Kamijo and Takeda (2010) reported intensive experience of physical training, regardless of type of sport, showed less positive stimulus-locked P3 in the switch than repeat trials in an alternating-runs switching paradigm. The reason was likely that these studies showed differences in the task difficulties, which P3 component was sensitive to Kamijo and Takeda (2010).

In this study, a cued task-switching task was employed for eliciting the proactive and reactive control processes modulated by the participants with experience of open- or closed-skilled physical activities. As the participants had received intensive training of two types of motor skills, we hypothesized that in the trials with predictive cues (100% validity), the open-skilled participants would show fewer positive differences in the between-trial (switch verse repeat) cue-locked P3 compared with the closed-skilled participants due to employing more anticipation in the open-skilled training. It was also hypothesized that in the trials with non-predictive cues (50% validity), the open-skilled participants would show fewer positive differences in the between-trial stimulus-locked P3 than the closed-skilled and control ones due to the imperative switch in the unpredictable environment. More negative-going CNV in the open- and closed-skilled than controls participants were anticipated due to better preparation in the former two groups. Comparable N2 amplitudes were anticipated in the open- and closed-skilled participants. The performances in the neural processes of the 75% validity condition would be between those of the 100 and 50% validity conditions.

MATERIALS AND METHODS

Participants

Fifty-four university students were recruited via convenience sampling. Among them, 18 (8 females and 10 males) were members of the university badminton team (open-skilled group), 18 (7 females and 11 males) were members of the university track and field team (closed-skilled group), and 18 (9 females and 9 males) declared they had not engaged in any professional or amateur sport (control group). The selection of badminton and track and field athletes as the open- and closed-skilled participants was made reference to those recruited in Wang and Tu (2017) and Wang et al. (2017). The results obtained would have more meaningful comparisons with those reported by Wang and Tu (2017) and Wang et al. (2017). Participants in each group had matched age and education levels (**Table 1**). The participants of the open- and closed-skilled groups had five or more years for professional motor skill practices. Each athlete had won prizes in open competitions and had no regular training in other sports. The levels of skill competences (in terms

of winning international/local sport awards) were comparable between two groups ($n = 2/16$ for open-skilled versus $n = 3/15$ for closed-skilled). All of the participants were right-handed with normal or corrected-to-normal visual acuity, and had no history of neurological or cardiovascular disorders. Participants were not on regular medication. The participant's cardiorespiratory fitness was assessed by the Queen's College step test, which had been introduced in Yu et al.'s (2017). Ethical approval of this study was granted by the Departmental Research Committee of The Hong Kong Polytechnic University. Written informed consent was obtained from the participants before commencing the experiment for data collection.

Experimental Task

This study used a cued task-switching paradigm to manipulate the proactive and reactive controls. Details of the task design were described in Yu et al. (2017). The time course of one typical trial is summarized in **Figure 1**. A trial began with presentation of a task cue (4 cm × 4 cm) at the center of the screen for 1500 ms. Then the task cue was replaced by a target stimulus (4 cm × 4 cm). Upon presentation of the target stimulus, the participant was asked to give a two-key sequential response correctly as soon as possible within 3000 ms. The next trial began once the response from the participant was registered. Next, a blank interval of 1000 ms appeared before the onset of the next trial.

In this paradigm, proactive and reactive controls were manipulated by means of cue validities. Three cue validities were used: 100, 75, and 50%. A 100% valid cue appeared as a solid square ▣ (or diamond ◈), a 75% valid cue appeared as a hollow square ▣ (or diamond ◈), and a 50% valid cue appeared (as a solid star ★). The participants were asked to prepare the response selection rules based on the cue (except for a 50% valid cue). A 100% valid cue meant the rule sets conveyed in the cue would be the same as those that appeared in the target stimulus. In this case, the participant could prepare to repeat the same task rule as the previous trial or switch to a new task rule based on the rule set conveyed in the cue. Thus, a 100% valid cue would elicit more proactive but fewer reactive control processes. A 50% valid cue meant that no information on the task sets in the subsequent response was provided. In such an ambivalent situation, the participant could not give a specific preparation for the response. The participant could repeat the same task rule or switch to a new task rule based on the target stimulus. It was expected that a 50% valid cue would elicit reactive rather than proactive control processes. A 75% valid cue was expected to elicit mental processes that were a combination of those in the 50 and 100% validity conditions. A digit (1 or 2) that appeared inside the shape (a square or diamond) formed the target stimulus (**Figure 1**). Two sets of response selection rules were used. Each rule involved two sets of two-key sequential responses delivered by the target stimulus with the same shape (square or diamond). For instance, one response selection rule was that a square with digit "1" was for the participant to press the "z" and then "n" keys on the keyboard, and a square with digit "2" was "x" and then "m"; the other response selection rule was that a diamond with digit "1" was "x" and then "n", and a diamond with digit "2" was "z" and then "m." The mappings

TABLE 1 | Demographic characteristics of the open-skilled, closed-skilled and control participants.

	Open-skilled	Closed-skilled	Control	F, p
Age, M (SD)	21.1 (2.2)	21.1 (2.0)	21.8 (2.1)	$F = 0.77, p = 0.466$
Weight kg, M (SD)	67.2 (11.6)	58.3 (9.4)	57.1 (10.2)	$F = 5.00, p = 0.010$
Height cm, M (SD)	170.3 (8.4)	169.9 (7.4)	165.1 (7.5)	$F = 2.55, p = 0.088$
BMI kg/m^2, M (SD)	22.9 (2.4)	20.1 (2.0)	20.9 (2.6)	$F = 7.06, p = 0.002$
Years of professional motor skill practices*, M (SD)	11.3 (2.7)	7.9 (1.6)	N.A.	$F = 179.35, p < 0.001$
Hours of professional motor skill practices each week, M (SD)	8.3 (1.8)	8.6 (1.6)	N.A.	$F = 216.93\ p < 0.001$
VO2max mL*kg^{-1}*min^{-1}, M (SD)	54.9 (9.3)	55.0 (10.2)	47.4 (10.4)	$F = 3.42, p = 0.040$
Other expertise outside of sports	Two participants engaged in playing musical instrument	Two participants engaged in playing musical instrument; one player in swimming	Two participants engaged in playing musical instrument; one player in wing chun	N.A.

Denotes the regular motor skill practices under the coach guidance.

FIGURE 1 | Schematic illustration of one typical trial in the cued task-switching paradigm.
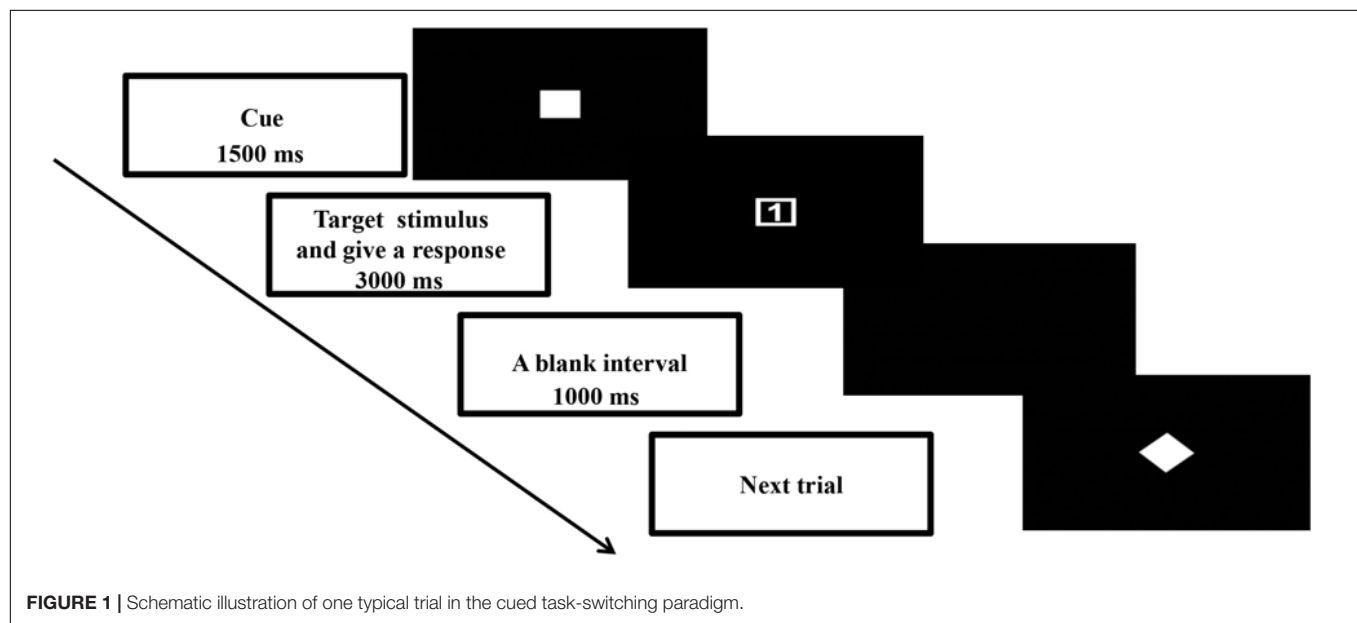
between cue stimulus (square or diamond) and two response selection rules were counterbalanced across participants. Only the 75% validity condition had congruent and incongruent trials. Congruent trials featured the same hollow square or diamond shapes appearing in the cue and target. Incongruent trials were when the shapes displayed in the cue and target differed. Only valid congruent trials in the 75% validity condition were included in the data analyses. Each trial was defined as "repeat" or "switch" depending on the response selection rule. A repeat trial was that the response selection rule was the same as that in the previous trial; whereas that of a switch trial was different from the previous trial.

The ratio of the three types of cue validities was 1:1:1. There was same ratio of repeat to switch trials in each block. Trials were organized in counterbalanced orders and grouped into nine blocks, and each block had 140 trials. It took around 9 min to complete one block followed by a 4-to-5 min break. NeuroScan Stim2 software (NeuroScan, Inc., Sterling, VA, United States) was

used for the fabrication of the trials. The time for completing the experimental task was approximately 1.5 h.

Data Collection Procedures
Preparation
Each participant was asked to complete the demographic information sheet, which included years for professional motor skill practices, hours for professional motor skill practices per week, sports categories, and other expertise outside of sports (**Table 1**). Before engaging in the experimental task, each participant sat on a comfortable chair in front of a table inside a dimly lit and electrically isolated sound-proof chamber. A 15-inch computer monitor for showing trials was placed on the table at a distance of 65–75 cm. Each participant was required to first complete 100 practice trials. Standardized instructions and feedback were given to the participant throughout the training block. This was followed by a test block in which the

participant completed 50 task trials and reached 90% accuracy before entering into the experiment. If the participant achieved an accuracy rate of less than 90%, he or she repeated the training block. The participant was also reminded to minimize eye blinks and to keep his or her eyes at the center of the monitor throughout the task.

Acquisition of ERP Data

Participants' EEG signals were captured by a 64-channel Quik-cap equipped with 90 mm Ag/AgCl sintered electrodes, SynAmps2 Digital DC EEG amplifier, and Curry 7 software (NeuroScan, Inc., Sterling, VA, United States). Vertical and horizontal electrooculograms (EOGs) were captured with two pairs of electrodes placed on the supra- and infra-orbital areas of the left eye and the left and right orbital rims of both eyes, respectively. A ground electrode was positioned on the forehead in front of the Cz electrode. All the channels were referenced to the electrodes on the left and right mastoids. The EEG and EOG signals were sampled at a rate of 1000 Hz/channel. All EEG/EOG electrode impedances were set to below 5 kΩ. EEG signals were recorded from the beginning of each block of experimental tasks. The timing of all stimuli was recorded by Curry 7 software.

Offline signal preprocessing also employed Curry 7 software. EEG signals were digitally filtered with a band pass from 0.01 to 30 Hz. The covariance analysis algorithm was used when eye movement was detected. Then the EEG signals were segmented into the cue- and stimulus-locked epochs. Cue-locked epochs were defined as −200 ms before the cue to 1,500 ms after the cue, and stimulus-locked epochs were defined as −200 ms before the target to 1,000 ms after the target. Baseline corrections were referenced to the pre-stimulus interval. Epochs with an amplitude exceeding ± 80 μv and trials with incorrect responses were excluded from the subsequent averaging procedure. The cue- and stimulus-locked waveforms of each electrode were averaged separately for three cue validities (100, 75 versus 50%) and two task conditions (repeat versus switch). The number of epochs extracted for data analysis for each cue validity in each of the repeat or switch trials was around 140 for each group.

Data Analysis

As the behavioral data of this study shared the same data set of a previous study conducted by the same research team, the detailed methods of analyzing the behavioral results of the participants can be found in Yu et al. (2017) and will not be repeated here. Analyses of the ERP data included the cue-locked P3, CNV, N2 and the stimulus-locked P3 elicited when participants engaged in the behavioral task. Independent component analysis (ICA) was conducted to confirm the time-windows set for extracting signals related to the cue-locked P3 (350–550 ms post-cue), CNV (1200–1500 ms post-cue), N2 (200–300 ms post-target), and stimulus-locked P3 (300–600 ms post-target). A short time-window was set for the cue-locked P3 for lowering the possible interferences to the CNV, as ICA results showed a slight overlap in the time-windows of these two components. In the analysis, the electrodes at the midline sites (Fz, Cz, and Pz) were included making reference to the

results of previous studies that switch effects were maximal at the midline electrode sites (Gajewski and Falkenstein, 2011; Hsieh and Wu, 2011; Li et al., 2012). A four-way repeated measures ANCOVA for validity (100, 75 versus 50%) × trial (repeat versus switch) × site (Fz, Cz versus Pz) × group (open-skilled, closed-skilled versus control) was conducted to test the mean amplitudes of the cue-locked P3, CNV, N2, and stimulus-locked P3. The years of participants' professional motor skill practices was the only covariate entered because of the significant between-group differences in this variable (**Table 1**). Another reason was that the years of professional motor skill practices rather than the MBI and VO2max was revealed to significantly predict the participants' performances on the behavioral tasks in Yu et al. (2017). *Post hoc* pairwise comparisons with the Bonferroni adjustment were applied when significant main or interaction effects were observed. This study only included the amplitudes rather than latencies in the ERP data analyses because previous study reported no significant differences between groups and between trials on the P3 latency (Scisco et al., 2008). As the switch cost of response time was related to amplitudes of cue- or stimulus-locked P3 components (Jamadar et al., 2010; Li et al., 2012), the present study examined the relations among cue- and stimulus-locked P3 and behavioral performance (i.e., switch cost of response time) for different motor skills by using a hierarchical, stepwise regression analysis for each of the validity conditions. For each regression equation, two regressors were the mean amplitude differences between switch and repeat trials of cue- ($cP3_S$-$cP3_R$) and stimulus-locked P3 ($sP3_S$-$sP3_R$); two other regressors were the identities of the open- and closed-skilled groups (with the control group as the reference). The two-way interaction terms for the neural processes and group identities were $cP3_S$-$cP3_R$ × open-skilled; $cP3_S$-$cP3_R$ × closed-skilled; $sP3_S$-$sP3_R$ × open-skilled; and $sP3_S$-$sP3_R$ × closed-skilled. The variance inflation factor (VIF) ≥ 10 and Pearson's correlation ≥ 0.85 were considered as indicators of strong multicollinearity between any of the two independent variables in a hierarchical regression model. The results showed no variable displayed strong multicollinearity, suggesting that none of the variables were related to each other. All analyses were performed with IBM SPSS statistics version 20.0 (IBM, Chicago, IL, United States).

RESULTS

The main behavioral variable was the switch cost of response time, which was defined as the difference in the reaction times between the switch and repeat trials. Two-way repeated measure ANOVA of validity (100, 75 versus 50%) × group (open-skilled, closed-skilled versus control) testing the effects on the switch cost of response times indicated that the validity × group effect on the switch cost of response times was marginally significant ($p = 0.053$). Participants in the open-skilled group showed significantly fewer switch cost values than the closed-skilled ($p = 0.023$) and control ($p < 0.001$) groups in the 100% validity condition (**Figure 2**). Participants in both the open- ($p < 0.001$) and closed-skilled ($p = 0.033$) groups showed
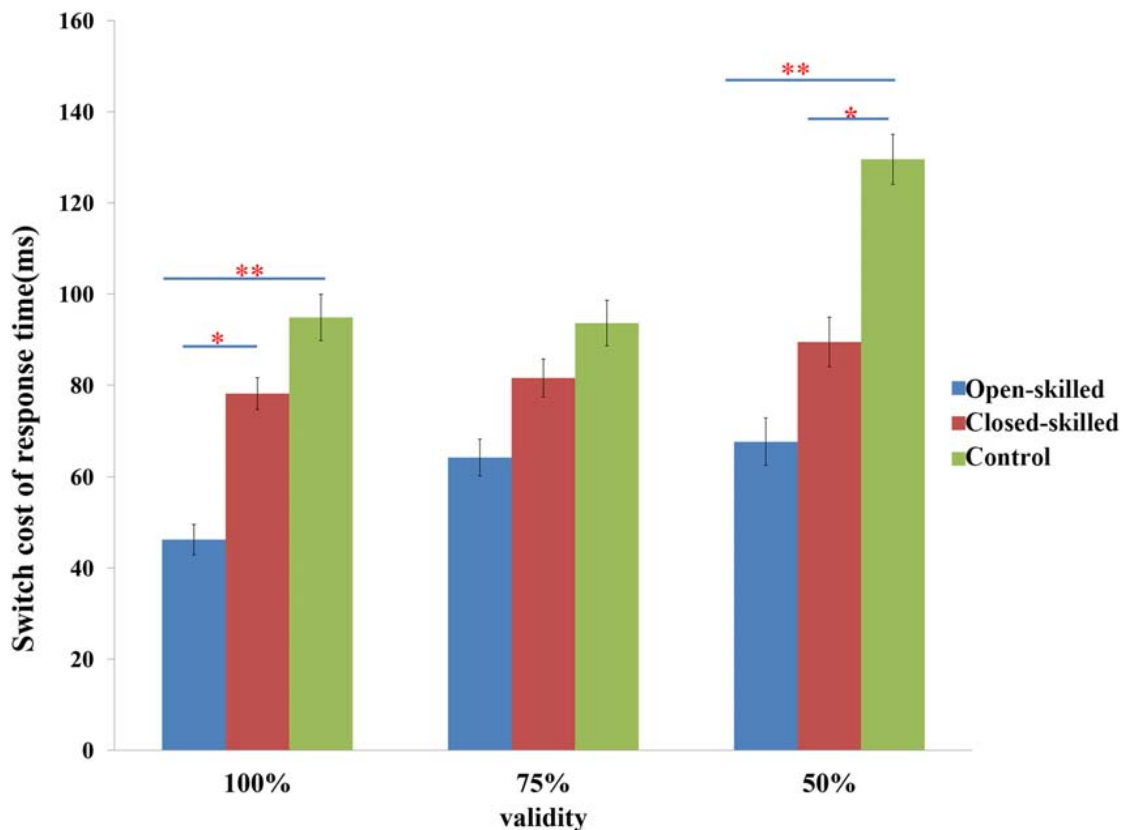
FIGURE 2 | The switch cost of response times in task-switching paradigm. Switch cost of response time = response time of switch trials − response time of repeat trials; 100% denotes 100% valid cue; 75% denotes 75% valid cue; 50% denotes 50% valid cue; *$p < 0.050$; **$p < 0.010$.

significantly fewer switch cost values than the control group in the 50% validity condition (**Figure 2**). No significant differences in switch costs were revealed between the open- and closed-skilled groups ($p = 0.473$) (**Figure 2**). Their details can be found in Yu et al. (2017).

Cue-Locked P3 (350–550 ms)

Figure 3 presents topographic maps (3A) and waveforms (3B) of the cue-locked P3 for the open-skilled, closed-skilled, and control groups. The covariate of years of professional motor skill practices [$F(1,50) = 0.54$, $p = 0.466$, $\eta_p^2 = 0.011$] was not significant. The validity × trial × site × group effect [$F(6.264,156.602) = 2.66$, $p = 0.016$, $\eta_p^2 = 0.096$] was found significant. The site main effect was also significant [$F(1.488,74.389) = 9.61$, $p = 0.001$, $\eta_p^2 = 0.161$]. However, the validity [$F(2,100) = 0.13$, $p = 0.880$, $\eta_p^2 = 0.003$], trial [$F(1,50) = 0.002$, $p = 0.968$, $\eta_p^2 < 0.001$], and group effects [$F(2,50) = 0.42$, $p = 0.657$, $\eta_p^2 = 0.017$] on the amplitudes of cue-locked P3 were not significant.

Post hoc analyses on trial × site × group effect were conducted separately at each level of validity. The trial × site × group effect was only significant in the 100% validity condition [$F(3.351,85.442) = 5.46$, $p = 0.001$, $\eta_p^2 = 0.177$] but not in the 75% [$F(2.752,70.179) = 1.11$, $p = 0.348$, $\eta_p^2 = 0.042$] and 50%

validity conditions [$F(3.249,82.856) = 0.13$, $p = 0.952$, $\eta_p^2 = 0.005$]. Hence, the trial × group effect for the 100% validity condition was further examined at each of the electrode sites. The trial × group effect was significant at Fz [$F(2,51) = 6.82$, $p = 0.002$, $\eta_p^2 = 0.211$] and Pz [$F(2,51) = 8.03$, $p = 0.001$, $\eta_p^2 = 0.239$] but not at Cz [$F(2,51) = 1.32$, $p = 0.28$, $\eta_p^2 = 0.049$]. At Fz, the open-skilled group showed marginally less positive-going cue-locked P3 in the switch than repeat trials ($p = 0.056$) in the 100% validity condition, whereas the closed-skilled group showed an opposite trend whereby the cue-locked P3 was significantly more positive-going in the switch than repeat trials ($p = 0.003$). The control group did not show significant between-trial-type differences in the amplitudes of cue-locked P3 at Fz ($p = 0.822$) (**Figure 3C**). At Pz, the open-skilled group showed a less positive-going cue-locked P3 in the switch than repeat trials ($p = 0.011$). The closed-skilled group, however, did not show significant between-trial-type differences in the amplitudes of cue-locked P3 at Pz ($p = 0.523$). The control group showed significantly more positive-going cue-locked P3 at Pz in the switch than repeat trials ($p = 0.004$).

CNV (1200–1500 ms)

In CNV, the covariate of years of professional motor skill practices was found significant [$F(1,50) = 4.21$, $p = 0.045$,

FIGURE 3 | Analyses of the cue-locked P3 among three groups of participants. (A) Topographical distributions within the 350–550 ms post-cue time-window regardless of trial types and validities; the Pz site is indicated with the green circle. (B) Mean amplitudes of the waveforms of the repeat (in green) and switch (in red) trial-types regardless of validities; the 350–550 ms time-window at Pz is indicated with the rectangular boxes. (C) Comparisons of the mean amplitudes between the repeat and switch trial-types at the Pz site for the 100% (left panel) and 50% (right panel) validity conditions. Error bars indicate SEM; $*p < 0.050$; $**p < 0.010$.

$\eta_p^2 = 0.078]$. The validity × trial × site × group effect $[F_{(6.118,152.946)} = 0.55, p = 0.772, \eta_p^2 = 0.022]$ on the amplitudes of CNV was not significant (**Figure 4**). The validity effect $[F_{(2,100)} = 7.20, p = 0.001, \eta_p^2 = 0.126]$ was significant. However, trial $[F_{(1,50)} = 0.06, p = 0.811, \eta_p^2 = 0.001]$, site $[F_{(2,100)} = 1.54, p = 0.221, \eta_p^2 = 0.030]$, and group effects $[F_{(2,50)} = 1.06, p = 0.354, \eta_p^2 = 0.041]$ were not significant. *Post hoc* analysis showed that CNV in the 100% validity condition was more negative-going than those in 75% ($p < 0.001$) and 50% ($p < 0.001$) validity conditions.

N2 (200–300 ms)

The covariate of years of professional motor skill practices was found not statistically significant $[F_{(1,50)} = 0.29, p = 0.596, \eta_p^2 = 0.006]$. The validity $[F_{(1.656,82.825)} = 1.85, p = 0.170, \eta_p^2 = 0.036]$, trial $[F_{(1,50)} = 1.63, p = 0.208, \eta_p^2 = 0.031]$, group $[F_{(2,50)} = 0.42, p = 0.658, \eta_p^2 = 0.017]$, and the validity × trial × site × group effects $[F_{(5.866,146.655)} = 1.23, p = 0.295, \eta_p^2 = 0.047]$ on the amplitudes of N2 were also not significant. However, the site $[F_{(1.275,63.730)} = 5.86, p = 0.012, \eta_p^2 = 0.105]$, and validity × trial × site × covariate (years of professional motor skill practices) effect were significant $[F_{(2.933,146.655)} = 2.80, p = 0.043, \eta_p^2 = 0.053]$. *Post hoc* analysis showed that N2 in 100% validity conditions was more negative-going in the switch than repeat trials at Fz ($p = 0.034$), Cz ($p < 0.01$), and Pz ($p < 0.01$), but this effect was not found in 75 and 50% validity conditions. In both 75 and 50% validity conditions, N2 was more negative-going at Fz ($ps < 0.05$) than Cz and Pz. **Figure 5A** presents the topographic maps of the N2 component.

Stimulus-Locked P3 (300–600 ms)

The topographic maps (5A) and waveforms (5B) of the stimulus-locked P3 for the open-skilled, closed-skilled, and control groups are presented in **Figure 5**. The covariate of years of professional motor skill practices was significant $[F_{(1,50)} = 4.46, p = 0.040, \eta_p^2 = 0.082]$. The validity × trial × site × group effect $[F_{(5.671,141.771)} = 2.25, p = 0.045, \eta_p^2 = 0.082]$ on the amplitudes of stimulus-locked P3 was found significant. The site $[F_{(1.474,73.719)} = 4.14, p = 0.031, \eta_p^2 = 0.076]$ and group effects $[F_{(2,50)} = 5.27, p = 0.008, \eta_p^2 = 0.174]$ were significant. The trial effect was marginally significant $[F_{(1,50)} = 3.93, p = 0.053, \eta_p^2 = 0.073]$. However, the validity $[F_{(2,100)} = 0.33, p = 0.719, \eta_p^2 = 0.007]$ effect was not significant.

By adjusting years of professional motor skill practices, *post hoc* analysis showed significant trial × site × group effect only in the 50% validity condition $[F_{(3.267,81.670)} = 3.48, p = 0.017, \eta_p^2 = 0.122]$ (**Figure 5C**), but not in the 100% $[F_{(2.824,70.610)} = 1.81, p = 0.322, \eta_p^2 = 0.045]$ and 75% validity conditions $[F_{(2.897,72.437)} = 1.37, p = 0.260, \eta_p^2 = 0.052]$. Interestingly, the results were different from those found in cue-locked P3, in which the same three-way interaction effect was significant in the 100% validity condition. The trial × group effect in the 50% validity condition was further tested separately for Fz, Cz, and Pz. The trial × group effect was found significant

at the Cz and Pz $[F_{(2,51)} = 3.52, p = 0.037, \eta_p^2 = 0.121; F_{(2,51)} = 4.14, p = 0.021, \eta_p^2 = 0.140,$ respectively], but not at the Fz $[F_{(2,51)} = 2.02, p = 0.143, \eta_p^2 = 0.073]$. At Pz, both the open-skilled ($p = 0.008$) and closed-skilled ($p = 0.002$) groups showed significantly less positive-going stimulus-locked P3 in the switch than repeat trials (**Figure 5C**). The control group, however, showed no significant between-trial-type difference in the stimulus-locked P3 amplitudes at Pz ($p = 0.630$). At Cz, the open-skilled ($p < 0.001$) and closed-skilled ($p = 0.038$) groups had significantly less positive-going stimulus-locked P3 in the switch than repeat trials; whereas the control group did not show any significant between-trial-type differences in the amplitudes of stimulus-locked P3 ($p = 0.838$).

Hierarchical Stepwise Regression

In the 100% validity condition, the regression model was significant, $R^2 = 0.226, F_{(4,49)} = 3.579, p = 0.012$, with the only significant regressor in the model being the open-skilled group as a group identity ($\beta = -0.573, p = 0.001$) (**Table 2**). Other regressors (e.g., $cP3_S$-$cP3_R$ and $sP3_S$-$sP3_R$, closed-skilled group) were not significant ($|\beta| < 0.285, ps > 0.057$). The changes in the variance explained by the open-skilled group regressor were also significant $[\Delta R^2 = 0.182, F_{(4,45)} = 3.455, p = 0.015]$. The effect of $cP3_S$-$cP3_R$ × open-skilled was significant ($\beta = 0.475, p = 0.007$), whereas $cP3_S$-$cP3_R$ ($\beta = -0.431, p = 0.072$) and $cP3_S$-$cP3_R$ × closed-skilled ($\beta = 0.247, p = 0.205$) did not show significant impacts (**Figure 6A**). These results suggested that the $cP3_S$-$cP3_R$, which was associated with proactive control for task switching, showed significant correlation with the switch cost of response times among the open-skilled participants but not among the closed-skilled and control participants.

In the 50% validity condition, the regression model was significant, $R^2 = 0.211, F_{(4,49)} = 2.782, p = 0.037$, for both the open- and closed-skilled groups as group identities were identified as significant regressors ($\beta = -0.474, p = 0.003; \beta = -0.397, p = 0.012$, respectively) (**Figure 6B**). The other two regressors, $cP3_S$-$cP3_R$ and $sP3_S$-$sP3_R$, were not significant ($\beta = -0.008, p = 0.949; \beta = -0.062, p = 0.643$, respectively). The changes in the variance are explained by the significance of the open- and closed-skilled groups' regressors $[\Delta R^2 = 0.157, F_{(4,45)} = 3.057, p = 0.043]$. The effects of $sP3_S$-$sP3_R$ ($\beta = -0.550, p = 0.020$), open-skilled ($\beta = -0.363, p = 0.031$), closed-skilled ($\beta = -0.347, p = 0.023$), $sP3_S$-$sP3_R$ × open-skilled ($\beta = 0.464, p = 0.022$), and $sP3_S$-$sP3_R$ × closed-skilled ($\beta = 0.519, p = 0.012$) were the significant predictors, suggesting that the correlations between $sP3_S$-$sP3_R$ and the switch cost of response time in the open- and closed-skilled groups were significantly different from those of the control group. A follow-up analysis suggested that such correlation was significantly negative in the control participants ($r = -0.534, p = 0.022$), but not significant in the open-skilled ($r = 0.262, p = 0.294$) and closed-skilled ($r = 0.274, p = 0.270$) participants (**Figure 6B**).

In the 75% validity condition, the regression model was also significant, $R^2 = 0.255, F_{(4,49)} = 4.203, p = 0.005$, with $cP3_S$-$cP3_R$ ($\beta = 0.320, p = 0.023$) and $sP3_S$-$sP3_R$ ($\beta = -0.378, p = 0.004$)
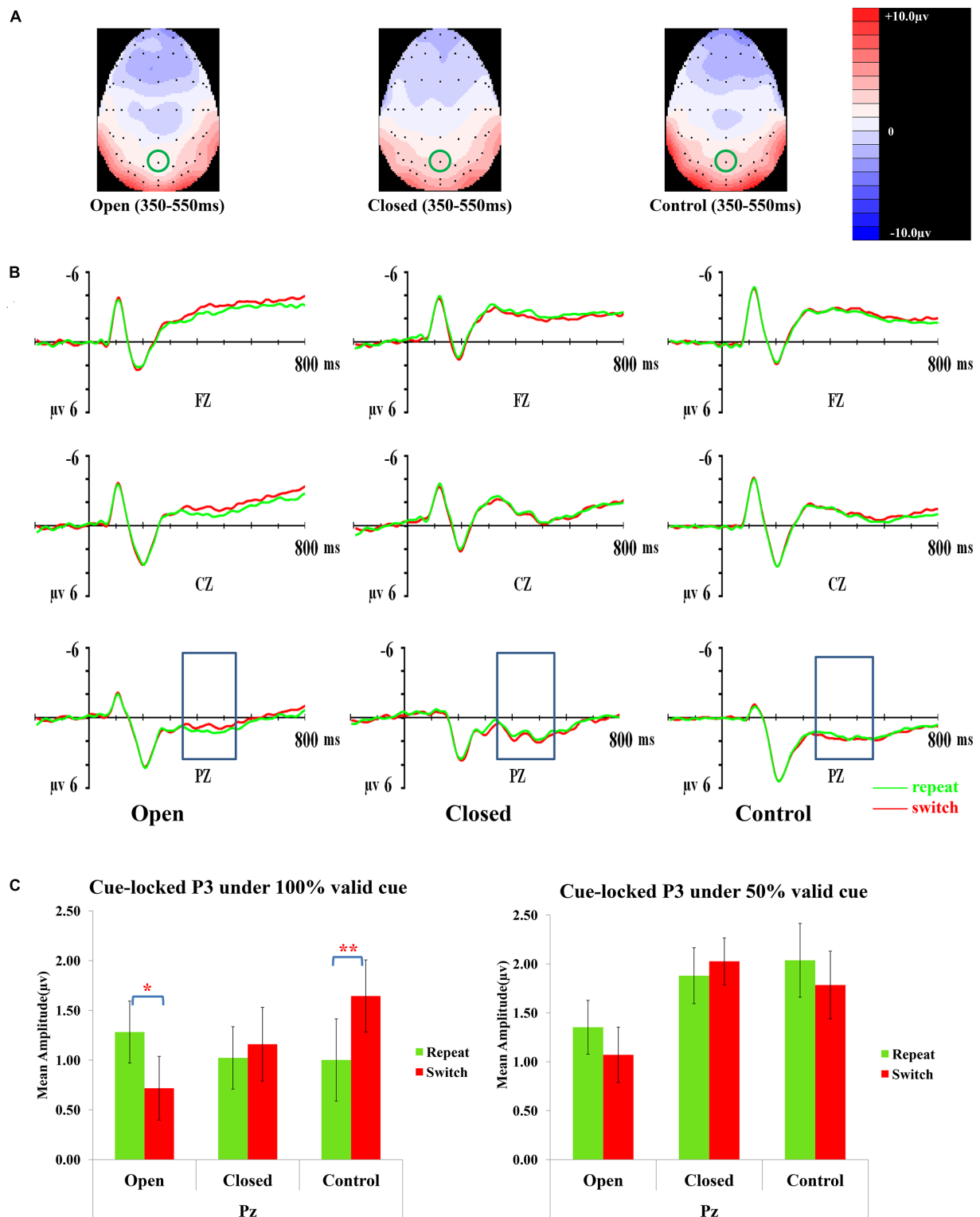
FIGURE 4 | Analyses of CNV among three groups of participants. **(A)** Topographical distributions within the 1000–1500 ms post-cue time-window regardless of trial types and validities; the Cz site is indicated with the red circle. **(B)** Mean amplitudes of the waveforms of the repeat (in green) and switch (in red) trial-types extracted at the Cz site regardless of validities; the 1000–1500 ms time-window is indicated with the rectangular boxes.

being significant regressors in the model. No other significant regressors were found [$R^2 = 0.277$, $F(8,45) = 2.151$, $p = 0.050$] (**Figure 6C**). Both cP3s-cP3r ($r = 0.317$, $p = 0.020$) and sSP3s-sSP3r ($r = -0.333$, $p = 0.014$) showed significant correlations with the participants' switch cost of response times, regardless of the subgroups.

DISCUSSION

The current study investigated how motor skill experiences modulated proactive and reactive controls of executive function in healthy adults. New findings are that the open-skilled participants, when compared with the other two groups, showed significantly less positive-going parietal cue-locked P3 in switch than repeat trials, which coupled with better performances on task-switching in the predictive condition. These suggest that proactive control was unique to the open-skilled participants. It appears that they might have been able to deploy fewer attentional resources in proactively updating the new action rule than the closed-skilled participants. These findings corroborate with the results of the regression analysis, which show that better proactive control for task switching was associated with the between-trial difference in the cue-locked P3 amplitudes for the open- but not closed-skilled participants. On the contrary, in the non-predictive condition both the open- and closed-skilled participants showed significantly less positive-going parietal stimulus-locked P3 in the switch than repeat trials, which could not be found in the control participants.

These results indicate that, prior experiences in motor skill training, regardless of the types of motor skills developed, would result in fewer deployments of attentional resources for reactively updating the new action rule under non-anticipatory circumstances. Our findings further confirm the dissociation of proactive and reactive controls in relation to their modulations by different motor-skill experiences. In particular, both proactive and reactive controls of executive functions could be enhanced by intensively exposing individuals to anticipatory and non-anticipatory enriched environments.

The results of the 75% validity condition will not be discussed because no significant findings were revealed in the comparisons of the ERP data.

Proactive Control and Open-Skill Experience

The experimental task used in the present study required the participants to switch between two sets of action rules. Our ERP results show less positive cue-locked P3 in switch than repeat trials observed among the open-skilled participants, which was not the case in the closed-skilled and control groups. These findings indicate that open-skilled participants deployed fewer attentional resources when proactively updating the new action rule than their closed-skilled counterparts. In addition, behavioral results on the same groups of participants reported by in Yu et al. (2017) showed that open-skilled participants exhibited smaller switch cost of response times in the predictive condition (100% validity) than closed-skilled participants. The ERP findings of the cue-locked P3 and the published behavioral
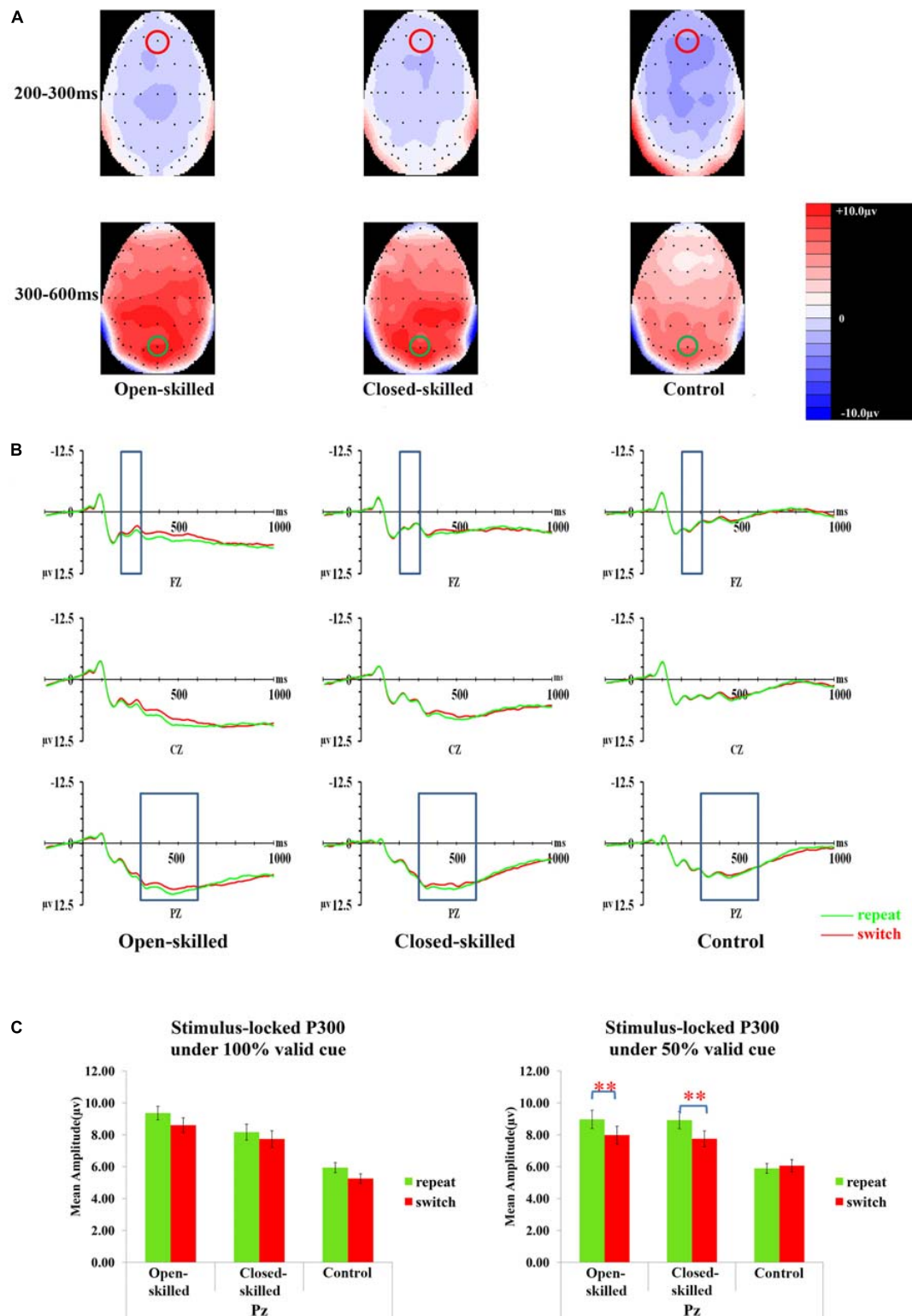
FIGURE 5 | Analyses of the N2 and stimulus-locked P3 among three groups of participants. (A) Topographical distributions of N2 (200–300 ms) and stimulus-locked P3 (300–600 ms) post-target time-window regardless of trial types and validities; the Fz site is indicated with the red circle, and the Pz site is indicated with the green circle. (B) Mean amplitudes of the waveforms of the repeat (in green) and switch (in red) trial-types regardless of validities; the 200–300 ms time-window at Fz and 300–600 ms time-window at Pz are indicated with the rectangular boxes. (C) Comparisons of the mean amplitudes between the repeat and switch trial-types at the Pz site for the 100% (left panel) and 50% (right panel) validity conditions. **$p < 0.010$.

TABLE 2 | Results of hierarchical stepwise regression of amplitudes of the cue- and stimulus-locked P3 for predicting the switch cost of response times in the 100, 50, and 75% cue validity conditions.

	100% cue validity	50% cue validity	75% cue validity
	Step 1	Step 1	Step 1
	Standardized regression coefficients (β)	Standardized regression coefficients (β)	Standardized regression coefficients (β)
$cP3_S$-$cP3_R$	−0.075	−0.008	0.320*
$sP3_S$-$sP3_R$	0.029	−0.062	−0.378**
Open-skilled group	−0.573**	−0.474*	−0.139
Closed-skilled group	−0.284	−0.397**	−0.051
R^2	0.226	0.185	0.255
Adjusted R^2	0.163	0.119	0.195
F	3.579*	2.782*	4.203**
	Step2	Step2	Step2
	Standardized regression coefficients (β)	Standardized regression coefficients (β)	Standardized regression coefficients (β)
$cP3_S$-$cP3_R$	−0.431	0.010	0.445
$sP3_S$-$sP3_R$	−0.262	−0.550*	−0.138
Open-skilled group	−0.393*	−0.363*	−0.133
Closed-skilled group	−0.265	−0.347*	−0.057
$cP3_S$-$cP3_R$ × Open-skilled	0.475**	0.066	−0.070
$cP3_S$-$cP3_R$ × Closed-skilled	0.247	0.131	−0.099
$sP3_S$-$sP3_R$ × open-skilled	0.293	0.464*	−0.096
$sP3_S$-$sP3_R$ × closed-skilled	0.225	0.519*	−0.255
R^2	0.408	0.342	0.277
Adjusted R^2	0.303	0.225	0.148
F	3.876**	2.925**	2.151
ΔR^2	0.182	0.157	0.021
ΔF	3.455*	2.685*	0.329

*$p < 0.050$; **$p < 0.010$; RT denotes response time; $cP3_S$-$cP3_R$ denotes difference of mean amplitudes between switch and repeat trial-types of cue-locked P3; $sP3_S$-$sP3_R$ denotes difference of mean amplitudes between switch and repeat trial-types of stimulus-locked P3; 100% denotes 100% valid cue; 75% denotes 75% valid cue; 50% denotes 50% valid cue.

data suggest higher efficiency in updating the new action rule in proactive control than the closed-skilled participants. Our results are consistent with those reported in previous studies, which found participants engaged in open-skilled physical activities have higher efficiency in the updating process related to motor preparation. Open-skilled participants were proposed to have better motor-reprograming processes in terms of smaller timing errors than closed-skilled participants (Nakamoto and Mori, 2012). Jin et al. (2011) also found that professional badminton players had more accurate judgments of the placement of badminton strokes and larger P3a amplitude in the proactive anticipation than non-professional players, suggesting good anticipation ability. Mcrobert et al. (2011) also reported the open-skilled participants had the higher online updating ability, with which they were more adaptive to dynamic environments in the anticipation. Bertollo et al. (2016) further explained that the higher efficiency in switching between the automatic and controlled processes among sport experts than amateurs is a reason for the former to proficiently adapt to the changing environment in open sports. When compared with amateur athletes, the high efficiency found among the open-skilled

athletes was revealed to result in the employment of less cognitive resources in controlled processes (Babiloni et al., 2010).

In the present study, the open-skilled participants are badminton athletes whose experience is in updating opponent's changed kinematics information and overcoming interferences from previous deceptive movement patterns (Müller and Abernethy, 2012). They succeeded in winning the game mostly depending on how well they update environmental changes and anticipate their opponent's actions (Bianco et al., 2017b). With such a background, the open-skilled participants tend to deploy fewer attentional resources when updating the new action rules in proactive control for the switch trials, which was not the case in the closed-skilled and control participants. The significant positive correlations between the amplitudes of cue-locked P3 and the switch cost of response time in our results were revealed only in the participants of the open-skilled group, which further substantiate the uniqueness of proactive control to the open-skilled participants.

No differences were revealed in the between-trial and between-group comparisons for the CNV. The non-significant between-group difference, supported by those reported in other

FIGURE 6 | Scatter plots showing the relationships between the differences of the mean amplitudes of the cue-locked P3 between switch and repeat trial-types (cP3$_S$-cP3$_R$) (right panel) or that of the stimulus-locked P3 (sP3$_S$-sP3$_R$) (right panel) and the switch cost of response time for the 100, 50, and 75% validity conditions (from a to c) among the three groups of participants. **(A)** In the 100% validity condition, one significant correlation coefficient ($r = 0.605$, $p = 0.008$) is revealed in the open-skilled group for the cue-locked P3 (cP3s-cP3r); the blue regression line presents the significant positive correlation between the two variables. **(B)** In the 50% validity condition, one significant correlation coefficient ($r = -0.534$, $p = 0.022$) is revealed in the control group for the stimulus-locked P3 (sP3s-sP3r); the green regression line presents the significant negative correlation between the two variables. Only control group shows significant correlation. **(C)** In the 75% validity condition, no significant correlations are revealed. *$p < 0.050$; **$p < 0.010$.

study (Bianco et al., 2017b), suggests that the level of the top–down attentional control preparing for task switching was comparable across the three groups of participants. No

difference between switch and repeat trials indicates that motor preparation levels were not affected by switch and repeat conditions, which was consistent with the findings in

Gajewski and Falkenstein (2015). These results further support the notion that the CNV appears to be not an important neural marker related to the executive functions in the proactive control process. The years of professional motor skill practices is a significant covariate for CNV component, which suggests that participants with professional motor skill practices, regardless of any type of motor skills, had more negative-going CNV than those without professional motor skill practices. These findings were in line with those reported by Bianco et al. (2017a,b). The tonic activity in motor preparation was related to speed control in pre-supplementary motor area (Bianco et al., 2017a). In the present study, the speed requirement in both open (badminton) and closed (most were runners) skills was high, which may contribute to no group differentiation in CNV amplitude.

Reactive Control and Motor Skill Experiences

The open- and closed-skilled participants showed less positive-going parietal stimulus-locked P3 in the switch than repeat trials in the non-predictive condition. Less positivity in the switch than repeat trials could not be found in the control group, who reported having no habit of practicing any types of physical activities. The results indicate that fewer attentional resources would have been deployed by the participants in updating the action rule in reactive control for switching in both open- and closed-skilled groups than in the control group. No significant differences were shown between these two motor-skilled groups, however. The comparable performance in stimulus-locked P3 for the open- and closed-skilled groups was not in line with our hypothesis, but was partly consistent with the findings in Wang et al.'s (2017). The findings of Wang and colleagues study revealed that open- and closed-skilled participants had comparable frontal N2 component and theta power for reactive control in a flanker task. The open-skilled participants, however, showed greater theta phase coherence (0–500 ms, 4 Hz; 300–400 ms, 5 Hz) for incongruent trials compared to congruent trials, but this effect could not be found in the closed-skilled group. These findings suggested that the dissociation between open- and closed-skilled participants appears to be the stability level of neural process rather than the level of allocated cognitive resources. Compared with those engaged in closed motor skills, the open motor skills required the participants to give a response within a limited time (Jin et al., 2011). Thus, the superior performance in reactive control of open-skilled participants could be showed in the paradigm with short interval between the response and the next target stimulus, like 500 ms in Tsai and Wang's (2015), but not long interval between the response and the next target stimulus, like 2500 ms in the present study.

The behavioral results reported in Yu et al. (2017) indicated that open- and closed-skilled participants had significantly smaller switch cost of response times than controls. The ERP findings reported in this study reveal that both open-and closed-skilled groups had higher efficiency in updating the new action rule in reactive control, reflected by less positive-going stimulus-locked P3. These findings were supported by those reported in previous studies (Kamijo and Takeda, 2010; Zhang et al., 2015).

Kamijo and Takeda (2010) reported that the participants with regular physical training showed less switch cost and less positive stimulus-locked P3 in the switch than repeat trials than the sedentary controls, suggesting better reactive control in task switching. One plausible reason to account for the enhanced reactive control among the participants could have been the inevitable gains in cardiorespiratory fitness due to the intensive training received by both open- and close-skilled groups (Scisco et al., 2008; Tsai and Wang, 2015). Zhang et al. (2015) indicated that experienced fencers deployed less cognitive effort in the reactive inhibition process (as reflected from the less positive-going P3 in Nogo condition) than their novice counterparts, which is consistent with the results revealed in this study. Nevertheless, Yamashiro et al. (2015) revealed more positive somatosensory Nogo-P3 in the open-skilled participants for reactive control in the Go/Nogo task than the closed-skilled participants. The inconsistent findings reported in Yamashiro et al. (2015) study were likely due to a lack of controlling the years of professional motor skill practices among the participants, which could confound the results in the reactive control condition. Another reason may be that the superior performance in baseball group resulting from the baseball specific training, which could not be generalized to the closed-skilled participants.

Significant correlations were revealed between the amplitudes of stimulus-locked P3 and the switch cost of response time only in the control group. The results are unexpected, as significant correlations were anticipated among the open- and closed-skilled participants. A plausible explanation for the non-significant findings could be due to the heterogeneity of the strategies employed by the open- or closed-skilled participants. To further test this proposition, a median-split method (Tamura et al., 2010) was applied to subdivide the open- and closed-skilled groups into higher and lower ability subgroups based on the participants' performances on between-trial difference in the stimulus-locked P3 amplitudes ($sP3_S$-$sP3_R$). Significant correlations were found in the higher but not the lower open-skilled ability subgroups; whilst significant correlations but opposite in direction were found in both the higher and lower closed-skilled ability subgroups. The small sample size for each of the subgroups ($n = 9$) only allowed us to suspect within-group heterogeneity as a possibly confounding factor to the non-significant relationships between the ERP and behavioral results. Future studies should explore possible deployment of different strategies by the participants in the same open- or closed-skilled group and the differences in the neural processes associated with proactive or reactive controls.

Our negative-going frontal N2 findings in the switch than repeat trials (in the 100 validity) at the Fz, Cz, and Pz electrodes are consistent with those reported by Hsieh and Wu (2011). The results suggested the possible involvement of response-set switching and suppression of conflict response processes unique to the behavioral task used in this study (Nakamoto and Mori, 2008; Wang et al., 2017), as no significant group difference was revealed in the N2 among all the participant groups.

Limitations

First, the proactive and reactive control processes were prescribed by the switching task employed in this study. The results may

not be directly generalized to other executive functions, such as inhibition or self-regulation. Second, the sample sizes of the open- and closed-skilled participants of this study were relatively small, which could have lowered the power of the analyses. Readers should be cautious when interpreting the results. Third, it is unclear whether badminton and track and field can best represent open- and closed-skilled physical activities, respectively. Any generalization of the findings should be restricted to participants of the same type of physical activities and level of competence. Future studies may consider recruiting participants of other types of physical activities and levels of competence. Fourth, the participants in the control group were those who had not been engaged in professional or amateur sports. The levels of physical activity engaged by these participants were not controlled. The existing differences between the open/closed-skilled and control groups could have been confounded by the differences in other parameters such as the levels of physical fitness rather than the types of motor skills, in case that the majority of the participants in the control group had been leading a sedentary lifestyle. Future study should recruit individuals who have comparable levels of physical fitness and/or physical training, but not at the professional level, as the control group. Fifth, the proposition of potential heterogeneity in strategies taken by the open- and closed-skilled groups was based on small sample sizes and without triangulation. Future studies should employ a more stringent research design and a larger sample size to address this issue.

CONCLUSION

This study explored how experiences in open and closed motor skills modulate individuals' proactive and reactive control processes. Compared with closed-skilled experiences, intensive open-skilled experiences were related to better proactive control for task switching characterized by lower switch cost and significant difference between switch and repeat trials of cue-locked P3 amplitudes. The enhanced proactive control is likely the result of high demand of anticipating environmental changes in open-skilled physical activities. Intensive open- and closed-skilled experiences were related to better reactive control for task switching than the experiences of control participants, which

was most likely resulted from higher cardiorespiratory fitness. Proactive and reactive controls as part of the process of motor-skill development seem to be transferable to domain-general executive functions.

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 The Research Grant Committee of the Hong Kong Polytechnic University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Abreu, A. M., Macaluso, E., Azevedo, R. T., Cesari, P., Urgesi, C., and Aglioti, S. M. (2012). Action anticipation beyond the action observation network: a functional magnetic resonance imaging study in expert basketball players. *Eur. J. Neurosci.* 35, 1646–1654. doi: 10.1111/j.1460-9568.2012.08104.x

Babiloni, C., Marzano, N., Infarinato, F., Iacoboni, M., Rizza, G., Aschieri, P., et al. (2010). "Neural efficiency" of experts' brain during judgment of actions: a high-resolution EEG study in elite and amateur karate athletes. *Behav. Brain Res.* 207, 466–475. doi: 10.1016/j.bbr.2009.10.034

Bertollo, M., Di Fronso, S., Filho, E., Conforto, S., Schmid, M., Bortoli, L., et al. (2016). Proficient brain for optimal performance: the MAP model perspective. *PeerJ* 4:e2082. doi: 10.7717/peerj.2082

Bianco, V., Berchicci, M., Perri, R. L., Quinzi, F., and Di Russo, F. (2017a). Exercise-related cognitive effects on sensory-motor control in athletes and drummers

compared to non-athletes and other musicians. *Neuroscience* 360, 39–47. doi: 10.1016/j.neuroscience.2017.07.059

Bianco, V., Di Russo, F., Perri, R. L., and Berchicci, M. (2017b). Different proactive and reactive action control in fencers' and boxers' brain. *Neuroscience* 343, 260–268. doi: 10.1016/j.neuroscience.2016.12.006

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends Cogn. Sci.* 16, 106–113. doi: 10.1016/j.tics.2011.12.010

Chan, J. S., Wong, A. C., Liu, Y., Yu, J., and Yan, J. H. (2011). Fencing expertise and physical fitness enhance action inhibition. *Psychol. Sport Exerc.* 12, 509–514. doi: 10.1016/j.psychsport.2011.04.006

Chevalier, N., Martis, S. B., Curran, T., and Munakata, Y. (2015). Metacognitive processes in executive control development: the case of reactive and proactive control. *J. Cogn. Neurosci.* 27, 1125–1136. doi: 10.1162/jocn_a_00782

Funderud, I., Lovstad, M., Lindgren, M., Endestad, T., Due-Tonnessen, P., Meling, T. R., et al. (2013). Preparatory attention after lesions to the lateral or orbital

prefrontal cortex–an event-related potentials study. *Brain Res.* 1527, 174–188. doi: 10.1016/j.brainres.2013.06.017

Gajewski, P. D., and Falkenstein, M. (2011). Diversity of the P3 in the task-switching paradigm. *Brain Res.* 1411, 87–97. doi: 10.1016/j.brainres.2011.07.010

Gajewski, P. D., and Falkenstein, M. (2015). Lifelong physical activity and executive functions in older age assessed by memory based task switching. *Neuropsychologia* 73, 195–207. doi: 10.1016/j.neuropsychologia.2015.04.031

Grane, V. A., Brunner, J. F., Endestad, T., Aasen, I. E., Kropotov, J., Knight, R. T., et al. (2016). ERP correlates of proactive and reactive cognitive control in treatment-naive adult ADHD. *PLoS One* 11:e0159833. doi: 10.1371/journal.pone.0159833

Hillman, C. H., Kramer, A. F., Belopolsky, A. V., and Smith, D. P. (2006). A cross-sectional examination of age and physical activity on performance and event-related brain potentials in a task switching paradigm. *Int. J. Psychophysiol.* 59, 30–39. doi: 10.1016/j.ijpsycho.2005.04.009

Hsieh, S., and Wu, M. (2011). Electrophysiological correlates of preparation and implementation for different types of task shifts. *Brain Res.* 1423, 41–52. doi: 10.1016/j.brainres.2011.09.018

Jamadar, S., Hughes, M., Fulham, W. R., Michie, P. T., and Karayanidis, F. (2010). The spatial and temporal dynamics of anticipatory preparation and response inhibition in task-switching. *Neuroimage* 51, 432–449. doi: 10.1016/j.neuroimage.2010.01.090

Jin, H., Xu, G., Zhang, J. X., Gao, H., Ye, Z., Wang, P., et al. (2011). Event-related potential effects of superior action anticipation in professional badminton players. *Neurosci. Lett.* 492, 139–144. doi: 10.1016/j.neulet.2011.01.074

Kamijo, K., and Takeda, Y. (2010). Regular physical activity improves executive function during task switching in young adults. *Int. J. Psychophysiol.* 75, 304–311. doi: 10.1016/j.ijpsycho.2010.01.002

Kida, N., Oda, S., and Matsumura, M. (2005). Intensive baseball practice improves the Go/Nogo reaction time, but not the simple reaction time. *Cogn. Brain Res.* 22, 257–264. doi: 10.1016/j.cogbrainres.2004.09.003

Li, L., Wang, M., Zhao, Q. J., and Fogelson, N. (2012). Neural mechanisms underlying the cost of task switching: an ERP study. *PLoS One* 7:e42233. doi: 10.1371/journal.pone.0042233

Linssen, A. M., Sambeth, A., Riedel, W. J., and Vuurman, E. F. (2013). Higher, faster, stronger: the effect of dynamic stimuli on response preparation and CNV amplitude. *Behav. Brain Res.* 237, 308–312. doi: 10.1016/j.bbr.2012.09.050

Mcrobert, A. P., Ward, P., Eccles, D. W., and Williams, A. M. (2011). The effect of manipulating context-specific information on perceptual-cognitive processes during a simulated anticipation task. *Br. J. Psychol.* 102, 519–534. doi: 10.1111/j.2044-8295.2010.02013.x

Müller, S., and Abernethy, B. (2012). Expert anticipatory skill in striking sports: a review and a model. *Res. Q. Exerc. Sport* 83, 175–187. doi: 10.5641/027013612800745059

Nakamoto, H., and Mori, S. (2008). Effects of stimulus-response compatibility in mediating expert performance in baseball players. *Brain Res.* 1189, 179–188. doi: 10.1016/j.brainres.2007.10.096

Nakamoto, H., and Mori, S. (2012). Experts in fast-ball sports reduce anticipation timing cost by developing inhibitory control. *Brain Cogn.* 80, 23–32. doi: 10.1016/j.bandc.2012.04.004

Nicholson, R., Karayanidis, F., Bumak, E., Poboka, D., and Michie, P. T. (2006). ERPs dissociate the effects of switching task sets and task cues. *Brain Res.* 1095, 107–123. doi: 10.1016/j.brainres.2006.04.016

Rushworth, M. F., Passingham, R. E., and Nobre, A. C. (2002). Components of switching intentional set. *J. Cogn. Neurosci.* 14, 1139–1150. doi: 10.1162/089892902760807159

Scheibe, C., Schubert, R., Sommer, W., and Heekeren, H. R. (2009). Electrophysiological evidence for the effect of prior probability on response preparation. *Psychophysiology* 46, 758–770. doi: 10.1111/j.1469-8986.2009.00825.x

Scisco, J. L., Leynes, P. A., and Kang, J. (2008). Cardiovascular fitness and executive control during task-switching: an ERP study. *Int. J. Psychophysiol.* 69, 52–60. doi: 10.1016/j.ijpsycho.2008.02.009

Swainson, R., Jackson, S. R., and Jackson, G. M. (2006). Using advance information in dynamic cognitive control: an ERP study of task-switching. *Brain Res.* 1105, 61–72. doi: 10.1016/j.brainres.2006.02.027

Taddei, F., Bultrini, A., Spinelli, D., and Di Russo, F. (2012). Neural correlates of attentional and executive processing in middle-age fencers. *Med. Sci. Sports Exerc.* 44, 1057–1066. doi: 10.1249/MSS.0b013e31824529c2

Tamura, R., Kitamura, H., Endo, T., Hasegawa, N., and Someya, T. (2010). Reduced thalamic volume observed across different subgroups of autism spectrum disorders. *Psychiatry Res.* 184, 186–188. doi: 10.1016/j.pscychresns.2010.07.001

Tarantino, V., Mazzonetto, I., and Vallesi, A. (2016). Electrophysiological correlates of the cognitive control processes underpinning mixing and switching costs. *Brain Res.* 1646, 160–173. doi: 10.1016/j.brainres.2016.05.048

Tsai, C. L., and Wang, W. L. (2015). Exercise-mode-related changes in task-switching performance in the elderly. *Front. Behav. Neurosci.* 9:56. doi: 10.3389/fnbeh.2015.00056

Verburgh, L., Scherder, E. J., Van Lange, P. A., and Oosterlaan, J. (2014). Executive functioning in highly talented soccer players. *PLoS One* 9:e91254. doi: 10.1371/journal.pone.0091254

Voss, M. W., Kramer, A. F., Basak, C., Prakash, R. S., and Roberts, B. (2010). Are expert athletes 'expert'in the cognitive laboratory? A meta-analytic review of cognition and sport expertise. *Appl. Cogn. Psychol.* 24, 812–826. doi: 10.1002/acp.1588

Wang, C. H., Chang, C. C., Liang, Y. M., Shih, C. M., Chiu, W. S., Tseng, P., et al. (2013a). Open vs. closed skill sports and the modulation of inhibitory control. *PLoS One* 8:e55773. doi: 10.1371/journal.pone.0055773

Wang, C. H., Chang, C. C., Liang, Y. M., Shih, C. M., Muggleton, N. G., and Juan, C. H. (2013b). Temporal preparation in athletes: a comparison of tennis players and swimmers with sedentary controls. *J. Motor Behav.* 45, 55–63. doi: 10.1080/00222895.2012.740522

Wang, C. H., and Tu, K. C. (2017). Neural correlates of expert behavior during a domain-specific attentional cueing task in badminton players. *J. Sport Exerc. Psychol.* 39, 209–221. doi: 10.1123/jsep.2016-0335

Wang, C. H., Yang, C. T., Moreau, D., and Muggleton, N. G. (2017). Motor expertise modulates neural oscillations and temporal dynamics of cognitive control. *Neuroimage* 158, 260–270. doi: 10.1016/j.neuroimage.2017.07.009

West, R., Bailey, K., Tiernan, B. N., Boonsuk, W., and Gilbert, S. (2012). The temporal dynamics of medial and lateral frontal neural activity related to proactive cognitive control. *Neuropsychologia* 50, 3450–3460. doi: 10.1016/j.neuropsychologia.2012.10.011

Yamashiro, K., Sato, D., Onishi, H., Sugawara, K., Nakazawa, S., Shimojo, H., et al. (2015). Skill-specific changes in somatosensory nogo potentials in baseball players. *PLoS One* 10:e0142581. doi: 10.1371/journal.pone.0142581

Yu, Q., Chan, C. C. H., Chau, B., and Fu, A. S. N. (2017). Motor skill experience modulates executive control for task switching. *Acta Psychol.* 180, 88–97. doi: 10.1016/j.actpsy.2017.08.013

Yu, Q., Fu, A. S. N., Kho, A., Li, J., Sun, X.-H., and Chan, C. C. H. (2016). Imagery perspective among young athletes: differentiation between external and internal visual imagery. *J. Sport Health Sci.* 5, 211–218. doi: 10.1016/j.jshs.2014.12.008

Zhang, D., Ding, H., Wang, X., Qi, C., and Luo, Y. (2015). Enhanced response inhibition in experienced fencers. *Sci. Rep.* 5:16282. doi: 10.1038/srep16282

The Effect of Meditation on Comprehension of Statements About One-Self and Others: A Pilot ERP and Behavioral Study

Alexander Savostyanov [1,2,3]*, Sergey Tamozhnikov [2], Andrey Bocharov [2,3], Alexander Saprygin [2], Yuriy Matushkin [1], Sergey Lashin [1,3], Galina Kolpakova [3], Klimenty Sudobin [3] and Gennady Knyazev [2]

[1] Laboratory of Psychological Genetics, Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences (SB RAS), Novosibirsk, Russia, [2] Laboratory Differential Psychophysiology, State-Research Institute of Physiology and Basic Medicine, Novosibirsk, Russia, [3] Laboratory of Biological Markers of Human Social Behavior, Humanitarian Institution at the Novosibirsk State University, Novosibirsk, Russia

The main goal of this study was to examine the effect of long-term meditation practice on behavioral indicators and ERP peak characteristics during an error-recognition task, where participants were presented with emotionally negative (evoking anxiety or aggression) written sentences describing self-related or non-self-related emotional state and personality traits. In total, 200 sentences written in Russian with varying emotional coloring were presented during the task, with half of the sentences containing a grammatical error that the participants were asked to identify. The EEG was recorded in age-matched control individuals ($n = 17$) and two groups of Samatha meditators with relatively short- (3–5 years' experience, $n = 18$) and long-term (10–30 years' experience, $n = 18$) practice experience. Task performance time (TPT) and accuracy of error detection (AED) were chosen as behavioral values. Amplitude, time latency and cortical distribution of P300 and P600 peaks of ERP were used as a value of speech-related brain activity. All statistical effects of meditation were estimated, controlling for age and sex. No behavioral differences between two groups of meditators were found. General TPT was shorter for both groups of meditators compared to the control group. Non-meditators reacted significantly slower to sentences about aggression than to sentences about anxiety or non-emotional sentences, whereas no significance was found between meditator groups. Non-meditators had better AED for the sentences about one-self than for the sentences about other people, whereas the meditators did not show any significant difference. The amplitude of P300 peak in frontal and left temporal scalp regions was higher for long-term meditators in comparison with both intermediate and control groups. The latency of P300 and P600 in left frontal and temporal regions positively correlated with TPT, whereas the amplitude of P300 in these regions had a negative correlation with TPT. We demonstrate that long-term meditation practice increases the ability of an individual to process negative emotional stimuli. The differences in behavioral reactions after onset of negative information that was self-related and non-self-related,

which is typical for non-meditators, disappeared due to the influence of meditation. ERP results could be interpreted as a value of increase in voluntary control over emotional state during meditational practice.

INTRODUCTION

Assessment of self and others (non-self) is one of the most important elements of regulation of social behavior. One of these methods, attribution theory, was suggested by social psychologist Heider (1958) to evaluate how people perceive their own behavior, as well as the behaviors of others. According to this theory, an act of attribution is a prediction of ongoing behaviors, such as emotions and motives for self or any other person. The main question of research in this area of self-feeling (our "I") is whether it has a special status in human consciousness or if it is functionally equivalent to semantic processing of any other classes of stimuli. Is our consciousness capable of paying more attention to information in which we are somehow involved? Symons and Johnson (1997) showed that information related to self in any way is remembered better than impersonal. This phenomenon became known as "self-reference effect". Neuroscientific studies on self-concept have mainly relied on experimental paradigms in which participants evaluated sentences that described some traits about the self. These studies revealed consistent activation in a medial prefrontal cortex (mPFC) during self-evaluation (see meta-analyses: Denny et al., 2012; Murray et al., 2012). In the study of Fossati et al. (2003), influence of emotional valence of stimuli was studied in addition to localization of the information processing relating to self. Activation in the right dorsomedial prefrontal cortex was found for self-referencing stimuli, which was independent from the valence of the words. Differences in processing of positive and negative adjectives were noticed in this study in areas outside of the medial prefrontal cortex. Other studies (Moran et al., 2006; Qin and Northoff, 2011) showed that besides mPFC, such areas as the cingulate cortex, precuneus, and temporal-parietal cortex are also involved in the processing of information about self.

One of the experimental approaches to studying evaluation of self and non-self is a comparison of groups of people that vary significantly in behavioral values related to social perception. Such groups can include people of different age (Pfeifer et al., 2009), gender (Cross and Madson, 1997), or ethno-cultural traits (Markus and Kitayama, 1991; Cross et al., 2011) that affect self-evaluation. In addition, healthy subjects can be compared to psychiatric patients with impaired self-perception processes (Sass and Parnas, 2003; Nelson et al., 2012; Mishara et al., 2014). In this study, we aimed to examine people practicing Buddhist Samatha meditation through neuroscientific approaches.

Meditation is a religious practice of great interest to neuroscience, as it is a model that allows the studying of changes occurring in regards to brain function, which occur due to voluntary and conscious efforts of the subject. Meditation's effects on attention have been previously studied using attention network task (ANT) and other experimental paradigms (Chiesa et al., 2011). Some studies report improvement in conflict monitoring (Wenk-Sormaz, 2005; Chan and Woollacott, 2007; Slagter et al., 2007; Tang et al., 2007; Moore and Malinowski, 2009). For example, it is shown that several days (20 min each day) of integrative mental-body training led to improved conflict monitoring (Tang et al., 2007). Additionally, longitudinal studies of 3-month long awareness meditation have shown decrease in attention blinking due to exercise (Slagter et al., 2007; van Leeuwen et al., 2009). In cross-section studies, experienced meditators demonstrated the best results in monitoring conflicts (van den Hurk et al., 2010).

According to the report of Aftanas and Golosheykin (2005), Sahaja Yoga meditators manifested smaller emotional arousal while watching aversive video clips, which was reflected in changes of spectral power in their EEG records. In addition, meditators manifested larger power values in theta-1 (4–6 Hz), theta-2 (6–8 Hz) and alpha-1 (8–10 Hz) frequency bands compared to control subjects in eye-closed conditions without external stimulation. These results were interpreted as proof of meditators having better capabilities to moderate intensity of emotional arousal.

Changes in amplitude of N1, P2, and P3 peaks in meditators compared to non-meditators were revealed in many ERP studies of different kinds of meditation (Cahn and Polich, 2009; Atchley et al., 2016; Biedermann et al., 2016). P300 amplitude increased in meditators immediately after meditation practice compared to the pre-meditation condition (Telles et al., 2019). The study of Vipassana meditation practice has found that the dynamics of P300 in conjunction with the theta and alpha bands spectral power indices reflected differences between people with different prolongation of meditation experience (Kakumanu et al., 2019). The amplitude of frontal P300 reflects a degree of concentration of voluntary attention. Therefore, most researchers interpret the influence of meditation on the amplitude of P300 as an index of the increase in the ability of meditators to voluntary concentrate attention on themselves and/or external events.

Several functional and structural MRI studies of awareness meditation have focused on neuroplasticity in areas responsible for attention control. Awareness practice effects on attention are mostly related to the anterior cingulate (van Veen and Carter, 2002; Cahn and Polich, 2006; Hölzel et al., 2007; Tang et al., 2010, 2012a,b, 2013; MacCoon et al., 2014; Tang and Posner, 2014). The anterior cingulate is responsible for executive attention and control through detecting conflicts caused by incompatible information processing flows (van Veen and Carter, 2002; Posner et al., 2007; Tang and Tang, 2014). The anterior cingulate cortex and the frontal cortex form a part of the network, which, due

to effective distant connections with other areas of the brain, facilitates cognitive processing (Sridharan et al., 2008; Tang et al., 2012b). Compared to the control group, experienced practitioners have shown an increase in activation of areas of the anterior cingulate during meditation (Hölzel et al., 2007) or with a conscious expectation of pain stimulus (Gard et al., 2012). In a controlled randomized longitudinal study, high activation of the ventral and/or rostral anterior cingulate cortex have also been identified during the rest period following 5 days of integrative mental-body training (MacCoon et al., 2014). Activation of the anterior cingulate cortex may increase during the initial stages of the practice and decrease with prolonged training (Brefczynski-Lewis et al., 2007).

Other brain areas related to attention, in which functional changes due to awareness meditation were witnessed, include the dorsolateral prefrontal cortex (dlPFC), where an increased response during executive processing was observed (Allen et al., 2012), and the posterior parietal cortex (PPC), which had shown higher activation after the MBSR course in subjects with social anxiety (Goldin and Gross, 2010).

Thus, the areas of the brain where activity changes under the influence of meditation and the areas involved in the self-assessment and evaluation of other people overlap to a significant degree. According to the self-report of participants, Samatha is a religious practice, with the end goal being the "dissolution" of Myself in Universe. Self-awareness is considered by adherents of such meditation an illusion that must be discarded. We have proposed that the long practice of Samatha can change the processes of evaluation of one-self and/or of other people, which can be witnessed using brain electrical activity analysis. We chose event-related potentials (ERP), occurring during the recognition of written sentences, which include emotionally negative descriptions of one-self's conditions or non-self traits and characteristics, as the method of our study. According to our hypothesis, in such a paradigm non-meditators should show differences in responses to messages about self and non-self, whereas the consequence of meditation should be the disappearance of such differences.

MATERIALS AND METHODS

Participants

The study was conducted at the Baikal Meditational Center[1]. The experimental sample included 53 healthy right-handed participants from 25 to 66 years (32 men; mean age = 41.0, SD = 8.3). The participants were divided into three groups: (1) non-meditators-17 persons (10 men, mean age = 40.5, SD = 8.5) who refused to take part in a meditational practice; (2) intermediate group–18 persons (11 men, mean age = 40.3, SD = 8.0) with 3–5 years of experience of meditation; and (3) long-term meditators–18 persons (11 men, mean age = 42.7, SD = 9.3) who had long-term (more than 10 years) experience of a meditation practice. All participants from the long-term group have the status of great masters of meditation, recognized in the Buddhists community. After completion of our study,

[1]http://www.geshe.ru/

they took part in "a retreat," i.e., semi-annual continuous meditation. The groups of participants were balanced in age and sex. All participants had no history of neurological, psychiatric, or major somatic disorders. According to the self-report, they denied use of narcotic drugs or other psychoactive substances. All participants were native Russian-speakers or natural bilinguals (Russian and one of the Siberian languages) and had normal or corrected-to-normal visual acuity. All participants (including non-meditators) are related to the lamaistic Buddhism community of Russia.

After the experiment all participants filled a Russian version of Goldberg's IPIP Big-Five Factor Markers (validated by Knyazev et al., 2010) for estimation of their psychological traits. In addition, a set of psychological questionnaires was used for estimation of participants' differences in emotional intelligence scores, trait anxiety level, vulnerability to depression and anxiety disorder, etc.

All subject protection guidelines and regulations were followed in accordance with the Declaration of Helsinki. The study goals were explained to all participants and they signed the informed consent. The study and the consent form were approved by the Institute of Physiology and Basic Medicine ethics committee and by the spiritual leader of the lamaistic Buddhism community of Russia.

Experimental Procedure

Two-hundred sentences in Russian language were selected for the experiment (see **Table 1**). Half of the sentences (100) contained a grammatical error. In preliminary testing on another group, the grammatically wrong sentences were selected so that they were easily recognized (more than 80% accuracy) by all participants. The text stimuli were presented in white and black (Arial, 36 pt) *via* a 24.4 cm × 18.3 cm monitor located 60 cm away in front of a participant. A warning signal (cross) appeared at the center of the screen for 0.5 s before the task onset. The participants were instructed to answer whether a presented sample contains an error by pressing one of two buttons with their dominant hand. Participants had three practice trials before task execution. Time for decision making was not limited for participants, but they were instructed to perform the task as quickly as possible and with maximal accuracy.

The experiment also contained a hidden condition: the sentences were grouped by emotional coloring, of which participants were not informed. There were five groups of sentences: (1) sentences describing aggression of a participant; (2) sentences describing aggression of other people; (3) sentences describing anxiety of a participant; (4) sentences describing anxiety of other people; and (5) neutral sentences about inanimate objects. The sentences from different categories were presented in random order. The sentences about anxiety were selected from the Russian version of Spielberger's State-Trait Anxiety Inventory (Spielberger et al., 1970; translated to Russian and validated by Hanin, 1976). The sentences about aggression were taken from the Buss-Perry aggression questionnaire (Buss and Perry, 1992; translated to Russian and validated by Yenikolopov and Tsibulsky, 2007). All sentences

TABLE 1 | Examples of sentences from the different categories.

Category of sentences	Right sentences	Wrong sentences
My aggression	Я дерусь чаще, чем окружающие.	Я драться чаще, чем окружающие.
Aggression of others	Он дерется чаще, чем окружающие.	Он драться чаще, чем окружающие.
My anxiety	Меня волнуют возможные неудачи.	Меня волновать возможные неудачи.
Anxiety of others	Её волнуют возможные неудачи.	Её волновать возможные неудачи.
Inanimate objects	Зеленое дерево стоит у дороги.	Зеленое дерево стоять у дороги.

about one-self were taken from validated Russian versions of psychological tests. Translations of these tests into Russian were performed by professional translators. All questionnaires were repeatedly validated in different samples in Russia and in other countries where Russian is widely spoken. The sentences about non-self anxiety or aggression were created by the replacement of a pronoun "I" or "Me" on accidentally chosen pronouns "He," "She," "They," "Him," "Her" or "Them." The sentences about objects, anxiety, and aggression were balanced in number of words and grammatical structure. The sentences about self and non-self differed only by pronouns and connected verbs. The order of sentences was randomized across participants.

EEG Recording

EEGs were recorded using 130 channels (128 EEG, VEOG, ECG) *via* Ag/AgCl electrodes. The EEG electrodes were placed on 128 head sites according to the extended International 5–10% system and referred to Cz with ground at FzA. The Quik-Cap128 NSL was used for electrode fixation. The electrode resistance was maintained below 5 kΩ. The signals were amplified using NVX 136 amplifier (MCS, Russia), with 0.1–100 Hz analog bandpass and continuously digitized at 1,000 Hz.

Behavioral Data Processing

The task performance time (TPT, in milliseconds) and accuracy of error detection (AED, a percent of correctly recognized error in wrong sentences and a percent of sentences without error correctly marked as "right") were used as the two behavioral values.

Initially, these values were used for repeated measures ANOVA with the Greenhouse-Geisser correction to test the main effects of such factors as "group" (non-meditators vs. meditators), "correctness" (right or wrong sentences), "sentence category" (five levels for different categories of sentences), "me_other" (sentences about participant or about other people, results for neutral sentences were excluded from such analysis), "aggression_anxiety" (sentences about aggression vs. about anxiety, results for neutral sentences were excluded from such analysis), age, sex, and interactions between these factors. However, any significant differences in the behavioral values between the participants from intermediate and long-term experience groups were not revealed. Both groups of meditators showed same differences from the non-meditators. For this reason, behavioral values were compared between the united group of meditators and the control group of non-meditators. Besides, no inter-group differences in TPT and AED were found

for neutral sentences. Therefore, behavioral values for neutral sentences were excluded from statistical analysis. Finally, for statistical comparison of TPT and AED values, the repeated measures ANOVA with the Greenhouse-Geisser correction with factors "group" (non-meditators vs. meditators of both groups), "correctness," "me_other," "aggression_anxiety," age, sex, and interactions between these factors was used. All statistical effects of meditation were estimated controlling for age and sex.

EEG Pre-processing and ERP Analysis

Recordings were processed in the EEGLAB toolbox (Delorme and Makeig, 2004). The trials containing artifacts were rejected from analysis. For each subject, 180–195 trials with sentence onset were used. The time intervals from 1.5 s before to +3.0 s after the fixation cross onset were analyzed. Time intervals from −1.5 to −0.75 s before fixation cross onset were used for baseline-correction.

EEGs were preliminary band-pass filtered in 1–40 Hz using elliptic filters. Following the suggestion by Delorme and Makeig (2004), re-reference (averaged reference) and baseline adjustment procedures were performed during data pre-processing. Independent component analysis (ICA) was used for correction of eye-movement and eye-blinking artifacts. First, the component's weights were computed individually for each record. Then, components corresponding to eye's artifacts were disclosed by visual inspection of component sets together with VEOG and ECG records. Artifactual components were removed in the pre-processing of EEGs.

To assess changes in signal amplitude, associated sentence onset, event-related potentials (ERPs) were calculated in the ERPLAB toolbox[2]. After removing artifacts, we computed ERP-values using the Erplab toolbox separately for every EEG-channel, subject and experimental condition, and cutoff filter at 12 Hz was applied to them. After that, maximal peak amplitudes, mean peak amplitude, and peak latencies in 150–500 ms (that corresponded to location of P300 peak during the visual peak analysis) and 600–1,000 ms (i.e., peak P600) time ranges were averaged across nine scalp regions of interest (ROI: left frontal, medial frontal, right frontal, left temporal, central, right temporal, and left, medial and right parietal-occipital scalp regions) for each subject. These values were used for repeated measures ANOVA with the Greenhouse-Geisser correction to test the main effects of such factors as "region of interest," "group," "correctness,"

---

[2]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3995046/

"sentence category," age, sex, and interactions between these factors.

RESULTS

Behavioral Results

No statistically significant inter-group differences in Big-Five factor markers were found in all scales of Goldberg's inventory. The main effects of factors "age" and "sex" and their interactions were statistically insignificant both for TPT (for age $p = 0.95$, for sex $p = 0.51$) and AED (for age $p = 0.33$, for sex $p = 0.12$).

As already mentioned, significant differences in the behavioral values between the participants from intermediate and long-term experience groups were not revealed. The main effect of "correctness" factor controlling for age and sex was highly significant as for TPT, $F_{(1,51)} = 23.78$; $p < 0.0001$, as for AED, $F_{(1,51)} = 15.65$; $p < 0.0001$. The correct sentences were recognized slower (correct sentences: mean TPT 4.0 ± 0.2 s; wrong sentences: 3.4 ± 0.2 s), but with better accuracy (correct sentences: mean accuracy $96.5 \pm 1.3\%$; wrong sentences: $90.6 \pm 0.9\%$), than the sentences with error. However, the interactions between factors "correctness," "group," "age," and "sex" were insignificant for both values ($p > 0.5$).

For TPT, the main effect of the "group" factor was statistically significant, $F_{(1,51)} = 2.95$; $p = 0.052$. Meditators solved the task faster (mean time 3.3 ± 0.2 s) than non-meditators (4.0 ± 0.2 s). In addition, for value of TPT, the main effect of the "me/other" factor was highly significant, $F_{(1,51)} = 22.69$; $p < 0.0001$. Sentences about one-self were recognized faster (mean time 3.5 ± 0.2 s), than about others (3.9 ± 0.2 s). Interaction between factors "Me/Other" and "group" for value of reaction time was statistically insignificant ($p > 0.3$). Additionally, a highly significant main effect of the "Anxiety/Aggression" factor, $F_{(1,51)} = 51.24$; $p < 0.0001$, and an interaction between "Anxiety/Aggression" and "group" factors, $F_{(1,51)} = 5.86$; $p = 0.019$ were revealed for TPT. For all groups of participants, recognition time for sentences about anxiety (mean 3.4 ± 0.2 s) was lower than for sentences about aggression (4.0 ± 0.2 s). However, meditators had smaller differences in recognition time for sentences about anxiety and aggression (anxiety: 3.4 ± 0.2 s; aggression: 3.6 ± 0.2 s) than non-meditators (3.7 ± 0.3 s и 4.4 ± 0.3 s, respectively; **Figure 1**).

The main effect of the "group" factor for AED value was not significant ($p > 0.4$). For AED, a statistically significant main effect of the "me/other" factor, $F_{(1,51)} = 7.46$; $p = 0.009$, and an interaction between "Me/Other" and "group" factors, $F_{(1,51)} = 9.11$; $p = 0.004$ were revealed. Averages between all group AED values were higher for sentences about one-self ($94.3 \pm 1.0\%$), than about others ($92.8 \pm 0.8\%$). However, these differences manifested only in the non-meditation group (one-self: $94.5 \pm 1.7\%$; other: $91.3 \pm 1.3\%$), while meditators had no such differences (one-self: $94.1 \pm 1.1\%$; others: $94.3 \pm 0.8\%$; **Figure 2**). The main effect of the "Anxiety/Aggression" factor ($p > 0.9$), and the interaction between "Anxiety/Aggression" and "group" ($p > 0.8$) were not significant. Interactions between "Me/Other," "Anxiety/Aggression," "age," and "sex" were also insignificant.

FIGURE 1 | Differences in time of task performance for the sentences about anxiety (light gray) and aggression (darkly gray) for both groups of meditators (left side) and non-meditators (right side).
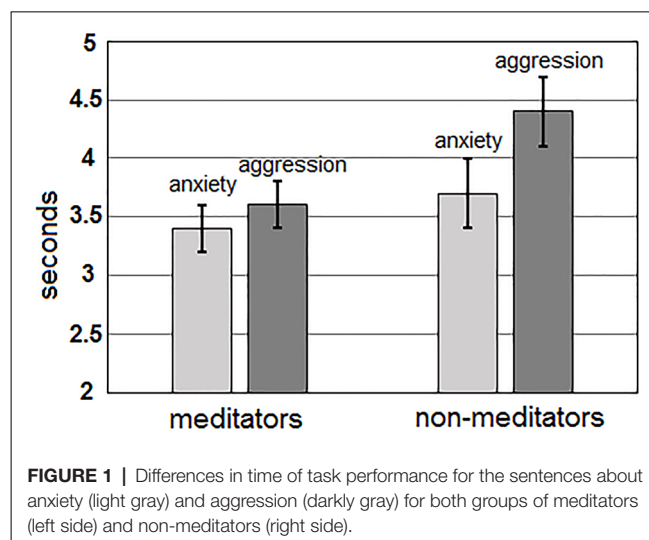
ERP Results

Time-amplitude plots of ERP in different cortical areas among different groups of participants are presented in **Figure 3**; cortical topography is presented in **Figure 4**. P300 and P600 peaks can be identified by visual inspection of these plots. P300 peak is clearly noticeable in the left temporal and all frontal cortical regions with amplitude maximum about 300 ms after sentence onset. Negative peak is detected in occipital-parietal cortical regions with amplitude maximum about 350 ms after sentences onset. The P300 has the highest amplitude in the group of long-term meditators in comparison with other groups. P600 peak is noticeable only in left temporal regions, i.e., in Broca's and Wernicke's areas. Amplitude of P600 among different groups is maximal in short-term meditators, but this difference was not significant. The topographic distribution of P300 and P600 peaks over cortical regions was not principally different among groups of participants. Though, general time-amplitude parameters and

FIGURE 2 | Differences in accuracy of error detection (AED) for the sentences about one-self (light gray) and others (darkly gray) for both groups of meditators (left side) and non-meditators (right side).

FIGURE 3 | Time-amplitude plots of ERP in nine cortical ROI for three groups of participants. Blue line—ERP for control group; pink line—ERP for short-term meditators; orange line—ERP for long-term meditators. First vertical line notices a moment of fixation cross onset; second vertical line notices a moment of sentence onset.

cortical distribution of P300 and P600 peaks were in consistence with standard patterns for tasks on language recognition.

In all subject groups, amplitudes of P300 и P600 peaks were significantly higher for sentences about aggression than for sentences about anxiety and inanimate objects. Amplitudes of these peaks were higher for incorrect sentences, however, no significant interactions between "group" and "category" or "correctness" were revealed. Also, there were no differences between ERP values for sentences about self and non-self. That's why we averaged ERP values for all sentence categories.

For the majority of measured ERP values (i.e., time latency of P300 and P600, averaged and maximal amplitude of P600, and averaged amplitude of P300) any inter-group differences were not revealed. Significant inter-group differences were revealed for the maximal amplitude of the P300 peak. The main effect of age was significant, $F_{(1,46)} = 6.16$; $p = 0.017$ while the main effect of sex was insignificant ($p > 0.7$). The statistical significance of effects of meditation to P300 amplitude controlling for age and sex was higher, than without such control. The main effect of the "group" factor for P300 values was insignificant ($p > 0.2$). However, the interaction of effects of factors "Region" and "Group" compared in three groups of participants was statistically significant, $F_{(16,368)} = 1.83$;

$p = 0.026$. Using one-way ANOVA, we have compared all pair combinations of three groups (i.e., control vs. intermediate; control vs. long-term meditators; intermediate vs. long-term). The comparison was performed separately for each cortical region, controlling for age and sex. One-way ANOVA revealed significant differences between the long-term meditators and other participants in left frontal, medial frontal, right frontal, and left temporal cortical regions. In all of these regions, long-term meditators showed higher amplitude of peak P300 in comparison with non-meditators and people with short-term experience of meditation (see **Table 2**). These effects were significant both for comparison between control participants and long-term meditators, and between short-term and long-term meditators. The significant differences between a control group and short-term meditators were revealed in right frontal and right temporal cortical regions. The amplitude of P300 in these regions was higher in control participants than in short-term meditators.

Correlations Between Behavioral and ERP Values

Two-tailed Pearson's correlation coefficients were calculated between the values of TPT and AED, averaged across all

FIGURE 4 | Cortical distribution of P300 (left side) and P600 (right side) amplitude averaged across all participants.
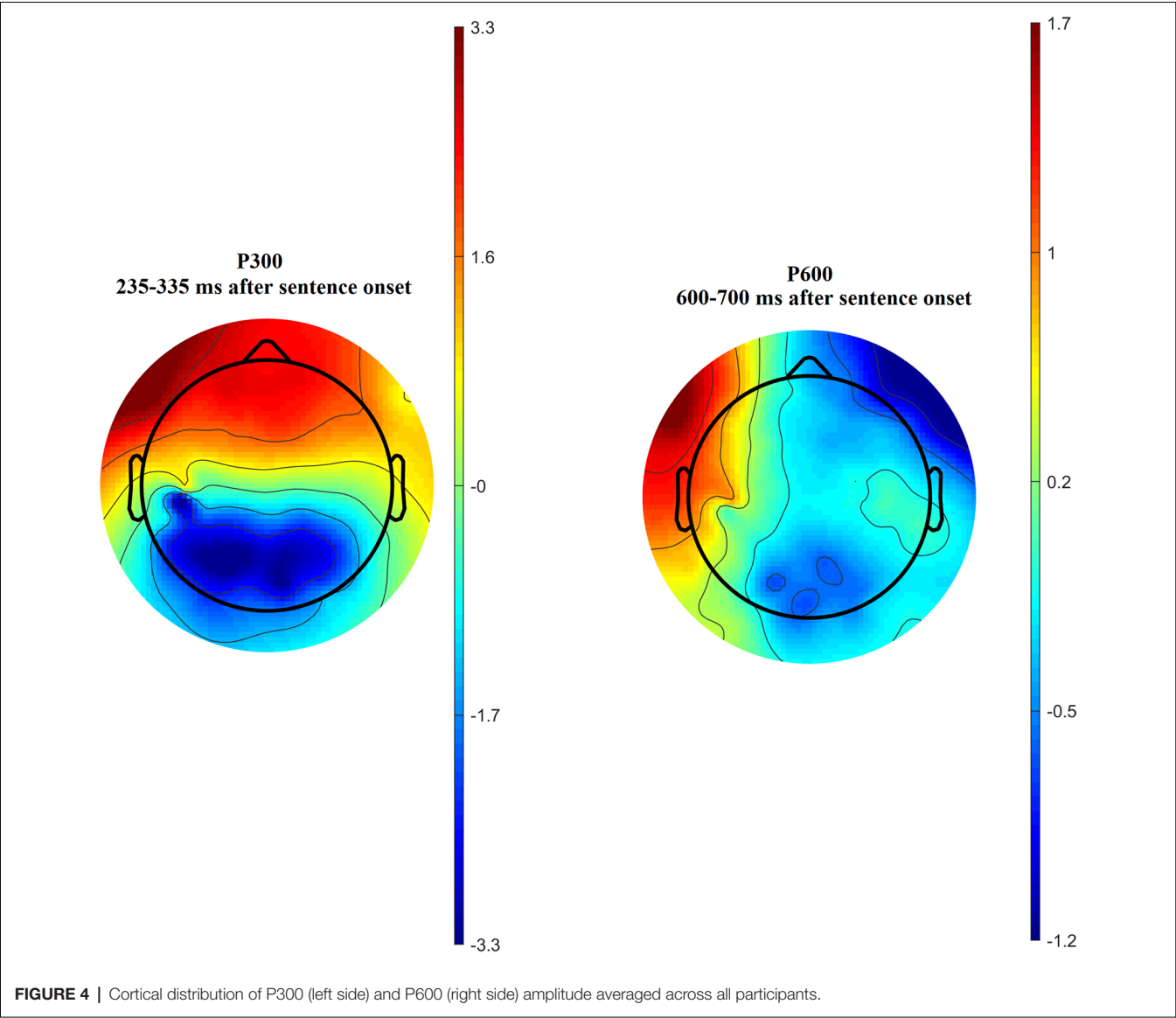
TABLE 2 | The differences in the P300 maximal amplitude between various experimental groups among cortical regions.

Region of interest	Non-meditators (group 1)	Intermediate meditators (group 2)	Long-term meditators (group 3)	P-level between 1 and 2 groups	P-level between between 1 and 3 groups	P-level between between 2 and 3 groups
Left frontal	**2.5 ± 0.3**	**2.2 ± 0.3**	**3.1 ± 0.3**	0.56	**0.043**	**0.039**
Medial frontal	**2.5 ± 0.3**	**2.1 ± 0.2**	**2.9 ± 0.3**	0.27	**0.049**	**0.045**
Right frontal	**2.4 ± 0.3**	**1.8 ± 0.2**	**2.9 ± 0.3**	**0.052**	**0.050**	**0.050**
Left temporal	**1.9 ± 0.3**	**1.6 ± 0.3**	**2.8 ± 0.4**	0.48	**0.013**	**0.001**
Central	1.0 ± 0.2	1.0 ± 0.1	0.9 ± 0.1	0.92	0.51	0.46
Right temporal	1.8 ± 0.2	1.2 ± 0.2	1.7 ± 0.2	**0.047**	0.41	0.24
Left parietal-occipital	1.9 ± 0.3	1.7 ± 0.2	1.9 ± 0.3	0.47	0.95	0.47
Medial parietal-occipital	2.1 ± 0.3	1.9 ± 0.2	2.1 ± 0.3	0.57	0.81	0.74
Right parietal-occipital	2.2 ± 0.3	2.1 ± 0.2	2.1 ± 0.3	0.80	0.90	0.91

Significantly different values are marked by a bold font.

categories of sentences, and values of maximal amplitude and time latency of P300 and P600 peaks of ERP separately in all cortical ROI. Maximal amplitude of P300 had significantly negative correlation with the TPT in left frontal ($r = -0.34$;

$p = 0.014$) and left temporal ($r = -0.31$; $p = 0.028$) cortical regions. The latency of P300 significantly positively correlated with TPT values in left frontal ($r = 0.34$; $p = 0.014$) and medial frontal ($r = 0.31$; $p = 0.029$) cortical regions. The latency of

P600 had a positive highly significant correlation with TPT in the left frontal ($r = 0.53$; $p < 0.0001$) and left temporal ($r = 0.56$; $p < 0.0001$) regions, and negative correlation with accuracy of error recognition in the left temporal area ($r = -0.30$; $p = 0.035$). Separated statistical analysis did not reveal any differences in correlations between behavioral and ERP values among different groups of participants and different categories of sentences.

Results Summary

Short-term and long-term experience groups of meditators were not statistically different in their behavioral values, however, they showed significant differences to the control group. Non-meditators showed decrease in TPT for sentences about aggression when compared to sentences about anxiety, while both meditators groups had these differences decreased due to faster recognition of sentences containing aggression. Additionally, non-meditators performed better in sentences about self than about non-self, while meditators had no such differences due to better performance in sentences about non-self. In ERP values long-term meditators showed higher amplitudes of P300 peak in frontal and left temporal cortical regions compared to other groups, while differences between non-meditators and short-term meditators were found in the right frontal and temporal regions. For all groups and all sentences, categories peak latencies of both P300 and P600 correlated positively with TPT. P300 peak amplitude in left frontal and temporal regions correlated negatively with TPT.

DISCUSSION

The aim of our study was to investigate how meditation changes perception of information connected with negative assessment of one-self or other people. The task reflecting implicit perception of emotional coloring of written speech was used as an experimental method. In our experiment, the voluntary attention of a participant was concentrated on detection of a grammatical error in the presented sentence, with the tasks being simple for the participants (mean AED was more than 90% for all participants). However, we compared speeds and accuracy of the answers for different categories of sentences which differed in their emotional coloring (neutral, anxious and aggressive), and whether the sentence was in relation to themselves or to other people. In this study, we have analyzed differences in behavior that participants did not control voluntarily and did not concentrate their attention on. It is possible to assume that the recorded differences in these measurements between meditators and non-meditators reflect some long-term changes in their behavior, which can be shown in other test tasks and ordinary life.

We have found two significant effects of meditation on the behavioral values and one effect of meditation on the ERP values. The behavioral differences were revealed in both groups of meditators (i.e., both for people with relatively shorter and longer experience of such practice), whereas the ERP differences distinguished the group with long-term experience both from non-meditators and meditators with

shorter experience. A slight difference in the P300 amplitude in the right frontal and right temporal areas was also found between the control participants and meditators with shorter experience.

The significant differences in time and accuracy in performing the set tasks for anxiety and aggression colored sentences, as well as for sentences about self and non-self were observed in non-meditators compared to meditators (the values for neutral sentences did not significantly differ from values for sentences about self-related anxiety). Non-meditators have recognized the sentences about anxiety quicker (difference approximately in 0.7 s) than sentences about aggression. The quality of recognition of sentences about anxiety and aggression in non-meditators did not differ. Also, non-meditators recognized the sentences about themselves quicker (difference in 0.4 s) and with better accuracy (difference about 3%), than the sentences about others. It is essential that the sentences about themselves and about others differed only with a pronoun and connected verb and contained identical grammatical errors. The order of presentation of these sentences have been randomized between participants. Thus, the differences in behavioral values for these sentences cannot be connected neither with their grammatical properties, nor with a presentation order, and reflect personal features of perception of negative information about self and others.

It is also important to note that all statistically significant effects of meditation were calculated controlling for participants' age and sex. Thus, behavioral and ERP differences between groups cannot be explained as a random result of influence of age and sex differences between groups of participants. We revealed a general correlation between TPT and AED, which also correlated with P300 amplitudes for all groups, as noted in the results on correlations between different indicators. However, long-experienced meditators showed a general decrease in TPT as compared to other groups, but the main effect of the "group" factor for AED value was not significant. Therefore, the effect of meditation on AED cannot be explained by the decrease of reaction time. In addition, the most important difference between meditators and non-meditators was not the overall accuracy or task response time, but the difference in AED when searching for an error in self-related and non-self-related sentences. This effect did not depend on the tasks response time. Also, the statistical significance of this effect increased when controlling for age and sex. Overall, the most significant behavioral effect of meditation did not depend neither on the general response time, nor on general quality of task performance, nor sex, nor age of the participants.

It has been shown that the recognition of grammatical errors in sentences depends on their emotional coloring. Sentences with emotionally negative signals usually induce a slower error recognition response, with worse accuracy scores when compared to emotionally positive or neutral sentences (Ayusheeva et al., 2018). In our results, it can be observed that in non-meditators the greatest delay and decline in quality of task performance happens in sentences containing aggression towards others. It could be interpreted as an indicator that the description of aggressive behavior of other people causes the

most negative emotions, which is shown in behavioral data of a linguistic task performance.

The effect of meditation consists in the general acceleration of performance of a linguistic task for all categories of emotional (but not neutral) sentences and in reduction of differences in the TPT between anxiety and aggression related sentences. This correlates with the assumption that meditation improves voluntary control over own emotions (Aftanas and Golosheykin, 2005; Marchand, 2012; Kasala et al., 2014). Our result could be interpreted as an indicator that meditators reduce the extent of negative perception of sentences about aggression that improves their control over execution of a grammatical task.

Other behavioral effects of meditation are connected with the alignment in AED for sentences about self and non-self. This effect happens due to improvement (in comparison with non-meditators) in accuracy in sentences about non-self with maintaining the accuracy for self. Differences in the TPT for the sentences about one-self and others in meditators remain the same, as well as in non-meditators. These results could be hypothetically interpreted based on data about cross-cultural differences in the neuronal processes underlying an assessment of self and non-self. It is known that in people who originate from individualistic (so-called "western") cultures, a significantly stronger activation of the medial prefrontal cortex was associated when thinking about self rather than non-self, including when thinking about relatives (Kelley et al., 2002). In contrast to this, in people from collectivistic ("oriental") cultures the thinking about self causes the same levels of activation of the prefrontal cortex as when promoted to think about relatives (Zhu et al., 2007; Wang et al., 2012). According to the self-report of our participants, the purpose of the Buddhist meditational practices is subjectively realized as "a dissolution of consciousness in the Universe." As meditators aspire to reach this goal, the differences between themselves and others are perceived as an illusion that needs to be eliminated. It is possible to speculate that the lack of differences in accuracy of recognition of sentences about self and others in meditators is an indicator reflecting the effect of meditation on their consciousness. On the other hand, all of our participants (including non-meditators) consider the Buddhist community of Russia and Kazakhstan to belong to oriental, rather than to western culture. Moreover, the differences in TPT for sentences about self and non-self were still present in the meditators group. Another possible explanation for the effect of meditation is the general decrease in negative emotions that improves attention to those sentences which cause the most negative reaction in non-meditators. Such an assumption is in agreement with the results received when comparing reactions to anxiety and aggression related sentences.

Amplitude of frontal P300 peak is one of the most frequently noted neurophysiological markers of meditation (Cahn and Polich, 2009; Atchley et al., 2016; Telles et al., 2019). In the study of Vipassana meditation, it was revealed that dynamics of P300 together with theta and alpha spectral power differ in people with various durations of meditation practice (Kakumanu et al., 2019). Our ERP results showed the differences at the level of brain activity between meditators with long-term experience both from non-meditators and intermediate group of meditators. In long-term meditators the amplitude of P300 peak in all frontal and left temporal cortical regions was increased in comparison with other groups. At the same time, in all groups of participants higher amplitude of P300 in left frontal and temporal regions correlated with faster task solution for all categories of sentences. Frontal amplitude of the P300 peak is a well-known correlate of voluntary-controlled attention (Heinze et al., 1999). Our findings could be interpreted as the indicator reflecting the general improvement of ability to concentrate attention as a result of long-term meditation. It is known, that the right fronto-temporal areas are activated under processing of unconscious negative emotion (Satto and Aoki, 2006). Therefore, a decrease in the P300 amplitude in meditators with shorter experience can be hypothetically interpreted as an indicator of a general decrease of sensitivity to negative emotions.

The differences revealed in behavior may have several interpretations. Probably, it is a direct effect of meditation practice on self-assessment behavior. Another possible reason of revealed behavioral effects is differences in the personality traits of the participants. In our study, intergroup differences in the Big-Five factor markers were not identified. However, some of psychological properties, not covered by the Goldberg's inventory, could exist between non-meditators and meditators before the latter began to practice meditation. These properties can have influence on recognition of self-related sentences.

It is necessary to notice that the differences in brain activity between the long-term meditators and other groups have no direct relation to the differences in the behavioral values. First of all, the behavioral differences have been found in the intermediate meditators as well, whereas the ERP differences were detected only in long-term meditators. Besides, ERP differences among the experimental groups did not concern the differences in reactions of the sentences about anxiety and aggression or about one-self and others, and were revealed for the reactions to all sentences, including the neutral ones. Respectively, in the present study we did not find the indicators of brain activity which are connected with the changes in perception of one-self and others, observed in behavioral assessment. On the contrary, our ERP findings reflect the long-term effect of meditation, which is shown at the perception of any speech-related information, including unemotional information about inanimate objects. A more detailed comparison of various behavioral and neurophysiological effects of meditation is a topic for our future research.

LIMITATIONS

1. Although the groups of participants were balanced by sex and age to the best of our ability, inside of all groups there was significant age variability. The age range was between 20 and 66 years old: it was connected to age heterogeneity in groups of meditators. Besides, the researched samples were not few in number of participants. Our hypotheses about the effects of

meditation on behavior and brain activity should be tested in larger groups.

2. Influence of inter-individual differences in the participants' personality traits on the effect of meditation was not observed in this study and remains a subject for future investigations.

3. There were no direct relationships between intergroup differences in the recognition of self-assessment sentences with differences in amplitudes and latencies of ERP. ERP indicators correlated with the general time of a linguistic task performance for all categories of sentences and did not relate to the perception of sentences about one-self or others. Therefore, the relationship between behavioral and electrophysiological effects of meditation remains a topic of future research.

4. Non-meditators and meditators were tested during their engagement in the religious rites of the Lamaistic community. In a period of study execution, all participants lived in a special camp in temporary isolation from their usual living conditions. Possible influence of Buddhist rituals on the psychological state of non-meditators should be taken into account. This topic requires comparing all research participants with people not involved in the ritual practice of Buddhism, which is also a topic of further research.

CONCLUSION

Meditative practice changes a perception of emotional coloring of written speech. Meditators have an increase in behavioral control of recognition of emotionally negative sentences about aggression, which is reflected in an increase in speed of performance of grammatical tasks. Meditation changes a perception of information about one-self and others. The differences in a recognition of sentences about one-self and others, which are characteristic of non-meditators, have not been observed in meditators. Effect of long-term meditation on brain activity is an increase in amplitude of P300 peak of ERP in frontal and left temporal cortical areas, which correlates with the reduction of the TPT. Revealed in this study is that ERP-effects of meditation are not specific to the emotional category of a sentence, and are not directly connected with processes of assessment of one-self or others.

REFERENCES

Aftanas, L., and Golosheykin, S. (2005). Impact of regular meditation practice on EEG activity at rest and during evoked negative emotions. *Int. J. Neurosci.* 115, 893–909. doi: 10.1080/00207450590897969

Allen, M., Dietz, M., Blair, K. S., van Beek, M., Rees, G., Vestergaard-Poulsen, P., et al. (2012). Cognitive-affective neural plasticity following active-controlled mindfulness intervention. *J. Neurosci.* 32, 15601–15610. doi: 10.1523/JNEUROSCI.2957-12.2012

Atchley, R., Klee, D., Memmott, T., Goodrich, E., Wahbeh, H., and Oken, B. (2016). Event-related potentials correlates of mindfulness meditation competence. *Neuroscience* 320, 83–92. doi: 10.1016/j.neuroscience.2016.01.051

Ayusheeva, T., Tamozhnikov, S., Karpova, A., Saprygin, A., Borisova, N., and Savostyanov, A. (2018). "The effect of 5-HTTLPR on EEG reactions

among Yakuts and Russians during the recognition of emotionally colored verbal stimuli," in *Proceedings of the 11th International Conference Bioinformatics of Genome Regulation and Structure \Systems Biology, BGRS\SB 2018* 124–126 (Novosibirsk, Russia: IEEE). doi: 10.1109/CSGB.2018.8544755

Biedermann, B., de Lissa, P., Mahajan, Y., Polito, V., Badcock, N., Connors, M. H., et al. (2016). Meditation and auditory attention: an ERP study of meditators and non-meditators. *Int. J. Psychophysiol.* 109, 63–70. doi: 10.1016/j.ijpsycho.2016.09.016

Brefczynski-Lewis, J. A., Lutz, A., Schaefer, H. S., Levinson, D. B., and Davidson, R. J. (2007). Neural correlates of attentional expertise in long-term meditation practitioners. *Proc. Natl. Acad. Sci. U S A* 104, 11483–11488. doi: 10.1073/pnas.0606552104

Buss, A. H., and Perry, M. (1992). The aggression questionnaire. *J. Pers. Soc. Psychol.* 63, 452–459. doi: 10.1037/0022-3514.63.3.452

Cahn, B. R., and Polich, J. (2006). Meditation states and traits: EEG, ERP, and neuroimaging studies. *Psychol. Bull.* 132, 180–211. doi: 10.1037/0033-2909. 132.2.180

Cahn, B. R., and Polich, J. (2009). Meditation (Vipassana) and the P3a event-related brain potential. *Int. J. Psychophysiol.* 72, 51–60. doi: 10.1016/j.ijpsycho. 2008.03.013

Chan, D., and Woollacott, M. (2007). Effects of level of meditation experience on attentional focus: is the efficiency of executive or orientation networks improved? *J. Altern. Complement. Med.* 13, 651–657. doi: 10.1089/acm. 2007.7022

Chiesa, A., Calati, R., and Serretti, A. (2011). Does mindfulness training improve cognitive abilities? A systematic review of neuropsychological findings. *Clin. Psychol. Rev.* 31, 449–464. doi: 10.1016/j.cpr.2010.11.003

Cross, S. E., Hardin, E. E., and Gercek-Swing, B. (2011). The what, how, why, and where of self-construal. *Pers. Soc. Psychol. Rev.* 15, 142–179. doi: 10.1177/1088868310373752

Cross, S. E., and Madson, L. (1997). Models of the self: self-construals and gender. *Psychol. Bull.* 122, 5–37. doi: 10.1037/0033-2909.122.1.5

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Denny, B. T., Kober, H., Wager, T. D., and Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J. Cogn. Neurosci.* 24, 1742–1752. doi: 10.1162/jocn_a_00233

Fossati, P., Hevenor, S. J., Graham, S. J., Grady, C., Keightley, M. L., Craik, F., et al. (2003). In search of the emotional self: an fMRI study using positive and negative emotional words. *Am. J. Psychiatry* 160, 1938–1945. doi: 10.1176/appi. ajp.160.11.1938

Gard, T., Hölzel, B. K., Sack, A. T., Hempel, H., Lazar, S. W., Vaitl, D., et al. (2012). Pain attenuation through mindfulness is associated with decreased cognitive control and increased sensory processing in the brain. *Cereb. Cortex* 22, 2692–2702. doi: 10.1093/cercor/bhr352

Goldin, P. R., and Gross, J. J. (2010). Effects of mindfulness-based stress reduction (MBSR) on emotion regulation in social anxiety disorder. *Emotion* 10, 83–91. doi: 10.1037/a0018441

Hanin, Y. L. (1976). *Quick Guide to the Application of the Scale of Reactive and Personal Anxiety by Charles D. Spielberger.* Leningrad: LNIIFK.

Heider, F. (1958). *The Psychology of Interpersonal Relations.* Hoboken, NJ: John Wiley and Sons Inc.

Heinze, H. J., Münte, T. F., Kutas, M., Butler, S. R., Näätänen, R., Nuwer, M. R., et al. (1999). Cognitive event-related potentials. The international federation of clinical neurophysiology. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 52, 91–95.

Hölzel, B. K., Ott, U., Hempel, H., Hackl, A., Wolf, K., Stark, R., et al. (2007). Differential engagement of anterior cingulate and adjacent medial frontal cortex in adept meditators and non-meditators. *Neurosci. Lett.* 421, 16–21. doi: 10.1016/j.neulet.2007.04.074

Kakumanu, R. J., Nair, A. K., Sasidharan, A., John, J. P., Mehrotra, S., Panth, R., et al. (2019). State-trait influence of Vipassana meditation practice on P3 EEG dynamics. *Prog. Brain Res.* 244, 115–136. doi: 10.1016/bs.pbr.2018.10.027

Kasala, E. R., Bodduluru, L. N., Maneti, Y., and Thipparaboina, R. (2014). Effect of meditation on neurophysiological changes in stress mediated depression. *Complement. Ther. Clin. Pract.* 20, 74–80. doi: 10.1016/j.ctcp.2013.10.001

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., and Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *J. Cogn. Neurosci.* 14, 785–794. doi: 10.1162/08989290260138672

Knyazev, G. G., Mitrofanova, L. G., and Bocharov, A. V. (2010). Validization of Russian version of Goldberg's "Big five factor markers" inventory. *Psychol. Zh.* 31, 100–110.

MacCoon, D. G., MacLean, K. A., Davidson, R. J., Saron, C. D., and Lutz, A. (2014). No sustained attention differences in a longitudinal randomized trial comparing mindfulness based stress reduction versus active control. *PLoS One* 9:e97551. doi: 10.1371/journal.pone.0097551

Marchand, W. R. (2012). Mindfulness-based stress reduction, mindfulness-based cognitive therapy, and Zen meditation for depression, anxiety, pain, and psychological distress. *J. Psychiatr. Pract.* 18, 233–252. doi: 10.1097/01.pra. 0000416014.53215.86

Markus, H. R., and Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychol. Rev.* 98, 224–253. doi: 10.1037/0033-295x.98.2.224

Mishara, A. L., Lysaker, P. H., and Schwartz, M. A. (2014). Self-disturbances in schizophrenia: history, phenomenology, and relevant findings from research on metacognition. *Schizophr. Bull.* 40, 5–12. doi: 10.1093/schbul/ sbt169

Moore, A., and Malinowski, P. (2009). Meditation, mindfulness and cognitive flexibility. *Conscious. Cogn.* 18, 176–186. doi: 10.1016/j.concog.2008. 12.008

Moran, J. M., Macrae, C. N., Heatherton, T. F., Wyland, C. L., and Kelley, W. M. (2006). Neuroanatomical evidence for distinct cognitive and affective components of self. *J. Cogn. Neurosci.* 18, 1586–1594. doi: 10.1162/jocn.2006. 18.9.1586

Murray, R. J., Schaer, M., and Debbané, M. (2012). Degrees of separation: a quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self- and other-reflection. *Neurosci. Biobehav. Rev.* 36, 1043–1059. doi: 10.1016/j.neubiorev.2011.12.013

Nelson, B., Thompson, A., and Yung, A. R. (2012). Basic self-disturbance predicts psychosis onset in the ultra high risk for psychosis "prodromal" population. *Schizophr. Bull.* 38, 1277–1287. doi: 10.1093/schbul/sbs007

Pfeifer, J. H., Masten, C. L., Borofsky, L. A., Dapretto, M., Fuligni, A. J., and Lieberman, M. D. (2009). Neural correlates of direct and reflected self-appraisals in adolescents and adults: when social perspective-taking informs self-perception. *Child. Dev.* 80, 1016–1038. doi: 10.1111/j.1467-8624.2009. 01314.x

Posner, M. I., Rothbart, M., Sheese, B., and Tang, Y. Y. (2007). The anterior cingulate gyrus and the mechanism of self-regulation. *Cogn. Affect. Behav. Neurosci.* 7, 391–395. doi: 10.3758/cabn.7.4.391

Qin, P., and Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *Neuroimage* 57, 1221–1233. doi: 10.1016/j.neuroimage. 2011.05.028

Sass, L. A., and Parnas, J. (2003). Schizophrenia, consciousness, and the self. *Schizophr. Bull.* 29, 427–444. doi: 10.1016/s1053-8100(03) 00071-0

Satto, W., and Aoki, S. (2006). Right hemispheric dominance in processing of unconscious negative emotion. *Brain Cong.* 62, 261–266. doi: 10.1016/j.bandc. 2006.06.006

Slagter, H. A., Lutz, A., Greischar, L. L., Francis, A. D., Nieuwenhuis, S., Davis, J. M., et al. (2007). Mental training affects distribution of limited brain resources. *PLoS Biol.* 5:e138. doi: 10.1371/journal.pbio.0050138

Spielberger, C. D., Gorsuch, R. L., and Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Sridharan, D., Levitin, D. J., and Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc. Natl. Acad. Sci. U S A* 105, 12569–12574. doi: 10.1073/pnas. 0800005105

Symons, C. S., and Johnson, B. T. (1997). The self-reference effect in memory: a meta-analysis. *Psychol. Bull.* 121, 371–394. doi: 10.1037/0033-2909. 121.3.371

Tang, Y. Y., Lu, Q., Fan, M., Yang, Y., and Posner, M. I. (2012a). Mechanisms of white matter changes induced by meditation. *Proc. Natl. Acad. Sci. U S A* 109, 10570–10574. doi: 10.1073/pnas.1207817109

Tang, Y. Y., Rothbart, M. K., and Posner, M. I. (2012b). Neural correlates of establishing, maintaining, and switching brain states. *Trends Cogn. Sci.* 16, 330–337. doi: 10.1016/j.tics.2012.05.001

Tang, Y. Y., Lu, Q., Geng, X., Stein, E. A., Yang, Y., and Posner, M. I. (2010). Short-term meditation induces white matter changes in the anterior cingulate. *Proc. Natl. Acad. Sci. U S A* 107, 15649–15652. doi: 10.1073/pnas.10110 43107

Tang, Y. Y., Ma, Y., Wang, J., Fan, Y., Feng, S., Lu, Q., et al. (2007). Short-term meditation training improves attention and self-regulation. *Proc. Natl. Acad. Sci. U S A* 104, 17152–17156. doi: 10.1073/pnas.0707678104

Tang, Y. Y., and Posner, M. I. (2014). Training brain networks and states. *Trends Cogn. Sci.* 18, 345–350. doi: 10.1016/j.tics.2014.04.002

Tang, Y. Y., and Tang, R. (2014). Ventral-subgenual anterior cingulate cortex and self transcendence. *Front. Psychol.* 4:1000. doi: 10.3389/fpsyg.2013. 01000

Tang, Y. Y., Tang, R., and Posner, M. I. (2013). Brief meditation training induces smoking reduction. *Proc. Natl. Acad. Sci. U S A* 110, 13971–13975. doi: 10.1073/pnas.1311887110

Telles, S., Singh, D., Naveen, K. V., Pailoor, S., Singh, N., and Pathak, S. (2019). P300 and heart rate variability recorded simultaneously in meditation. *Clin. EEG Neurosci.* 50, 161–171. doi: 10.1177/1550059418790717

van den Hurk, P. A., Giommi, F., Gielen, S. C., Speckens, A. E. M., and Barendregt, H. P. (2010). Greater efficiency in attentional processing related to mindfulness meditation. *Q. J. Exp. Psychol.* 63, 1168–1180. doi: 10.1080/17470210903249365

van Leeuwen, S., Willer, N. G., and Melloni, L. (2009). Age effects on attentional blink performance in meditation. *Conscious. Cogn.* 18, 593–599. doi: 10.1016/j.concog.2009.05.001

van Veen, V., and Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiol. Behav.* 77, 477–482. doi: 10.1016/s0031-9384(02)00930-7

Wang, G., Mao, L., Ma, Y., Yang, X., Cao, J., Liu, X., et al. (2012). Neural representations of close others in collectivistic brains. *Soc. Cogn. Affect. Neurosci.* 7, 222–229. doi: 10.1093/scan/nsr002

Wenk-Sormaz, H. (2005). Meditation can reduce habitual responding. *Altern. Ther. Health Med.* 11, 42–58.

Yenikolopov, S. N., and Tsibulsky, N. P. (2007). Psychometric analysis of the Russian-language version of the Questionnaire of Diagnostics of Aggression A. Buss and M. Perry. *Psychol. J.* 1,115–124.

Zhu, Y., Zhang, L., Fan, J., and Han, S. (2007). Neural basis of cultural influence on self-representation. *Neuroimage* 34, 1310–1316. doi: 10.1016/j.neuroimage.2006.08.047

frontiers
in Human Neuroscience

Check for
updates

Cognitive Fitness Framework: Towards Assessing, Training and Augmenting Individual-Difference Factors Underpinning High-Performance Cognition

Eugene Aidman [1,2,3]*

[1]Land Division, Defence Science & Technology Group, Edinburgh, SA, Australia, [2]School of Psychology, The University of Sydney, Sydney, NSW, Australia, [3]School of Biomedical Sciences & Pharmacy, University of Newcastle, Newcastle, NSW, Australia

The aim of this article is to introduce the concept of Cognitive Fitness (CF), identify its key ingredients underpinning both real-time task performance and career longevity in high-risk occupations, and to canvas a holistic framework for their assessment, training, and augmentation. CF as a capacity to deploy neurocognitive resources, knowledge and skills to meet the demands of operational task performance, is likely to be multi-faceted and differentially malleable. A taxonomy of CF constructs derived from Cognitive Readiness (CR) and Mental fitness (MF) literature maps into phases of operational cycle from foundational to advanced, mission-ready and recovery. Foundational cognitive attributes, such as attention, executive control and co-action, were hypothesized to be trainable at the initial Cognitive Gym phase. More advanced training targets at the CR phase included stress and arousal regulation, adaptability, teamwork, situation awareness (including detection, sense-making and prediction) and decision making (de-biasing and confidence calibration). The mission-ready training phase is focused on tolerances (to sleep loss, monotony, pain, frustration, uncertainty) and resistance (to distraction, deception or manipulation). Operational Augmentation phase relies on support tools such as decision aids and fatigue countermeasures, while the Recovery phase employs reflexive (e.g., mindfulness), and restorative practices (e.g., nutrition and sleep hygiene). The periodization of cognitive training in this cycle is hypothesized to optimize both real-time cognitive performance and the resilience that enables life-long thriving. One of the most promising avenues of validating this hypothesis is by developing an expert consensus on the key CF ingredients and their relative importance in high-performance settings.

Keywords: cognitive fitness, task performance, operational readiness cycle, RDoC domains cognitive functioning, measurement, trainability

INTRODUCTION

Performance psychology is a rapidly expanding field that is of growing significance to a wide range of occupations—from competitive sport and performing arts to first responder and military professions. These user groups share a common focus on striving for superior performance in challenging tasks under stressful conditions, and on effective recovery to enable repeat performance across the lifespan. The physical and psychological factors contributing to task performance are tightly interconnected and go beyond mere "wellness" (i.e., the absence of pathology). They include, apart from knowledge and skills, a range of "capacity" factors, such as strength, endurance, and flexibility, that is best summarized by the concept of "fitness." This article introduces the concept of Cognitive Fitness (CF) as a capacity to deploy neurocognitive resources underpinning the execution of goal-directed action and proposes a hypothetical set of its ingredients. Similar to physical fitness (PF), CF enables the application of knowledge, skills and attitudes (KSA) in generating task performance. The ingredients of PF are well established, with robust measurement protocols for muscular strength, aerobic/anaerobic endurance and range of motion/joint flexibility (Jeffreys and Moody, 2016), as well as validated training interventions such as strength and conditioning, cardiovascular fitness, or high-intensity interval training.

The meaning of *psychological fitness* is less clear, with this widely used term referring to diverse characteristics ranging from "character strengths and assets" (Cornum et al., 2011; Vie et al., 2016) to "resources that provide protection against the development of mental disorders" (Wesemann et al., 2018). The concept of *mental fitness* (MF) has emerged in the mental health and positive psychology literature (McCarthy, 1964; Seligman, 2008) to promote a positive and proactive notion of mental health. The MF literature is focused on identifying protective factors, such as cognitive flexibility, implicated both in the prevention of mental illness and in the promotion of flourishing (Keyes, 2007; Robinson et al., 2015). MF is also critical in the world of work, especially in high-stakes occupations where cognitive lapses can undermine the performance of complex socio-technical systems, while individual's and teams' superior capacity to sense, think, decide and act is widely seen as conferring critical performance advantages (Baker and Phillips, 2000; Bowers and Cannon-Bowers, 2014; Fletcher and Wind, 2014; Herzog and Deuster, 2014; Ahn and Cox, 2016; Bogga, 2017).

The growing literature on *cognitive readiness* (CR; Foster, 1996; Morrison and Fletcher, 2002; Grier, 2011; O'Neil et al., 2014) has developed extensive conceptual models of factors contributing to sustained professional performance in complex, dynamic, and unpredictable environments (Bolstad et al., 2006; Grier, 2012). In particular, CR has been construed as a comprehensive set of predictors—both distal and proximal—of cognitive performance by the military personnel in complex missions facing agile, near-peer opposition (Kluge and Burkolter, 2013; Sotos, 2019). The key components of CR include: (1) trainable skills, knowledge and attitudes (KSAs); (2) dynamic

functional states; and (3) stable, trait-like characteristics ranging from cognitive ability to working memory and learning styles (Grier, 2012; Mason and McQuade, 2013). CF (Aidman, 2017) corresponds to the cognitive element of this latter CR component. Despite the lack of causal cognitive mechanisms explained by the CR construct (Crameri et al., 2019) it has been instrumental in stimulating the development of measurement frameworks for assessing individuals' and teams' fluctuating capacity for operational task performance (Fatkin and Patton, 2008; Grier et al., 2012) and evaluating training interventions to improve it (Kluge and Burkolter, 2013; Peña and Brody, 2016). The explanatory power of the CR construct can be enhanced by establishing its connections to causal factors, such as arousal regulation, discomfort tolerance and inhibitory control, underpinning an individual's performance in cognitively demanding tasks. These fundamental, biologically traceable dimensions of cognitive functioning have been well established in clinical neuroscience (Cuthbert and Insel, 2013; Appelbaum, 2017). Their incremental predictive validity for the assessment of psychopathology (e.g., Yücel et al., 2019) and mental health in nonclinical populations (Carcone and Ruocco, 2017) indicates that these broad domains of cognitive functioning can potentially underpin CR as well. The concept of CF (Aidman, 2017) was introduced to examine these causal connections between CR and cognitive factors underpinning mental health, to bridge the gap between the CR and MF literature, and to develop a more tractable and systemic approach to the assessment and training of high-performance cognition. CF is focused on the "capacity" component of CR—as distinct from its "knowledge, skill and expertise" (KSE) components. The aim of the current article is to develop a working definition of CF, by drawing on the "why," the "what" and the "how-to" questions from the cognitive training and readiness literatures, and connecting them to the broad, biologically traceable domains of cognitive functioning developed in clinical neuroscience. A hypothesized set of constituent elements of CF is then articulated in a Cognitive Fitness Framework (CF2), followed by a discussion of its potential applications in the areas of assessment, training periodization and operational augmentation. Finally, a research agenda is suggested to improve CF2 and the measurement tools supporting it.

COGNITIVE TRAINING: FROM "HOW TO" TO "WHAT" AND "WHY"

MF has been defined as a set of malleable attributes that can develop through regular practice—analogous to physical training (Robinson et al., 2015). How trainable are the elements of CF is an empirical question. The growing literature on cognitive training has accumulated promising evidence of its effectiveness in areas such as visuospatial training in developmental populations (Boccia et al., 2017) and memory training for older adults (Gross et al., 2012; although see Redick, 2019; for a broader analysis of the latter).

More specifically, the occupational performance literature has accumulated promising evidence of the cognitive attributes that are considered both trainable and capable of producing

reliable performance gains for the end-user—be it an athlete (Morris and Summers, 2004; Fadde and Zaichkowsky, 2018), police officer (Page et al., 2016), first-responder (Joyce et al., 2019), or a warfighter (Adler et al., 2015; Cooper and Fry, 2018; Blacker et al., 2019).

Despite the inconsistent evidence for the effectiveness of cognitive training (Walton et al., 2018; Redick, 2019) and substantial gaps in conceptual integration of cognitive attributes relevant to high-stakes performance applications, core psychological skills such as goal setting, imagery, attention and stress/arousal regulation have been shown to improve with systematic, deliberate practice (Zaichkowsky and Peterson, 2018) characterized by immediate performance feedback and gradual improvement through repetition (Ericsson, 2008). However, the literature is not clear on the parameters of such practice, what it should target and when—or even on how complete this skillset is. Where do you start? What should you train first, second, last? In what combinations, doses, frequency, with what recovery times? At what phase of your training cycle? If you have to choose, which ones are more important? What are the skill fade rates and refresher training requirements? This systemic picture seems missing here, with the resulting symptoms of slow progress, such as persistently low rates of transfer of cognitive training to untrained tasks (Redick, 2019). At the same time, the field is flooded with a plethora of technological inventions for cognitive training and augmentation that are more focused on demonstrating the diverse new technologies than on what cognitive faculties are worth training for (with the attendant questions about the limits of their trainability), and what the best methods of training them are in a holistic and pragmatic approach that accounts for how these faculties develop and fade, improve and decline through existing real-life processes of maturation, aging, training and education, medical treatment, et cetera.

Addressing the questions of trainability would open up a realistic, evidence-based consideration of those cognitive attributes that are either unlikely to improve through training (and thus should be selected for) or fluid/cyclical in nature (and thus need to be monitored and augmented). Unpacking the neurocognitive mechanisms underlying task performance is critical to assessing attribute trainability. This mechanistic analysis requires the increasingly relevant evidence from neurosciences about factors impacting cognitive functioning—from genetics to social interactions.

DOMAINS OF COGNITIVE FUNCTIONING

With a ground-swell in the mental health literature suggesting that mental illness is not a category, CF2 suggests that neither is high-performance. Both are natural consequences of the varying levels of psychological functioning (including cognitive, affective and motivational) ranging from deficit to norm, and further to high or gifted performance. Broad expert consensus exists on key domains of cognitive functioning that underpin mental health (Morris and Cuthbert, 2012; Yücel et al., 2019). The deficiency of

categorical diagnostic systems is well known—DSM-5 and ICD are being challenged by the Research Domain Criteria (RDoC; Cuthbert and Insel, 2013) framework with its broad dimensions of cognitive functioning (Appelbaum, 2017; Clark et al., 2017).

RDoC defines "major domains for the study of mental illness and validate them using optimal genetic, neuroscientific, physiological, behavioral, and self-report measures" (Morris and Cuthbert, 2012). The long-term goals of RDoC were to validate tasks for use in clinical trials, identify new targets for treatment development, and provide a pathway by which research findings can be translated into changes in clinical decision making. RDoC identified broad higher-level domains of functioning that comprise multiple sub-dimensional constructs, reflecting state-of-the-art knowledge about major systems of cognition, motivation, and social behavior. In its present form, the RDoC Matrix contains five broad Domains cognitive functioning that are differentiated into 23 main Constructs (shown in brackets here) which are further divided into Subconstructs (Clark et al., 2017):

1. Negative Valence Systems (fear, anxiety, sustained threat, loss, frustration).
2. Positive Valence Systems (approach motivation, initial and sustained responsiveness to reward, reward learning, and habit).
3. Cognitive Systems (attention, perception, working and declarative memory, language, and cognitive control).
4. Social Processing Systems (affiliation/attachment, communication, self/other-perception).
5. Regulatory Systems (arousal, circadian rhythms, sleep and wakefulness).

The full list of RDoC constructs is regularly updated at https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/constructs/rdoc-matrix.shtml. The constructs are being continuously revised and refined, with the overall goal of improving measurement validity and treatment efficacy. Expert consensus frameworks have become a best-practice standard, they are known to stimulate research discoveries and accelerate translational pathways by estimating the relevance of primary RDoC constructs (and their sub-dimensions) to specific application domains such as substance and behavioral addictions (Yücel et al., 2019).

RDoC has informed the development of reliable and valid measures across a range of units of analysis for each construct—from genes and cells to neurocircuits, whole-body physiology (e.g., heart-rate or event-related potentials), behavior and subjective experience (e.g., Passell et al., 2019). These measures have enabled and inspired studies to determine the full range of variation along with these measurement constructs, from deficit to norm and characterizing both clinical and nonclinical populations (Carcone and Ruocco, 2017). Extending this range to the well-adjusted functioning and high-performance domains is an important next step, given that nonclinical populations have been under-represented in the current RDoC-driven research. This would require developing an expert consensus on the relative importance of

primary RDoC constructs and their sub-dimensions to various high-performance applications. Consequently, CF2 is aimed at building on RDoC foundational evidence in order to define major domains for the study of CF and develop guidelines for assessing them using an optimal mix of biomarker, physiological, behavioral, and self-report measures (Aidman, 2017).

COGNITIVE FITNESS FRAMEWORK

Integrating the evidence about the mechanisms of cognitive deficit and psychopathology (the RDoC literature) with what is relevant to high performance by well-adjusted individuals who are motivated to excel, remains incomplete. While the exact composition of RDoC domains relevant to work performance is awaiting full articulation through consensus studies similar to Robinson et al. (2015) and Yücel et al. (2019), their preliminary scoping can be informed by the widely recognized "psychological skillset" established in performance psychology (Zaichkowsky and Peterson, 2018). In exercise sciences, the ingredients of PF are well-established and include strength/power, endurance, agility and flexibility (Jeffreys and Moody, 2016). The corresponding features of cognitive performance include focus intensity (Sherlin et al., 2013) for strength, attention span and mental effort tolerance (Aidman et al., 2002, 2016; Aidman, 2005) for endurance, task shifting (Genet and Siemer, 2011), cognitive flexibility and creativity (Palmiero et al., 2019) for flexibility, and adaptability (Chandra and Leong, 2016; Zhang et al., 2019) and self-regulation (Schunk and Greene, 2017) for agility. Research evidence accumulated in sport psychology and other high-performance contexts, points to the same core domains of cognitive functioning, while their relative importance may depend on the specifics of task and mission profiles under consideration.

In particular, the MF resource index (Robinson et al., 2015) populates the same three categories—strength, endurance, and flexibility—with a set of positive psychology constructs such as self-efficacy (for strength), acceptance (for flexibility) and resilience (for endurance). These allocations are metaphorical—they "employ metaphor" (Robinson et al., 2015, p. 56) to create constructs that are similar to the well-understood components of PF. As a result, their neuro-psychological bases remain unclear, and the question of their measurement is left wide open. The extensive set of CR constructs (Cosenzo et al., 2007; Grier, 2011, 2012; O'Neil et al., 2014) is focused on higher-order abilities, such as decision making, problem-solving and metacognition, it also contains some underlying cognitive capacity constructs of agility, speed of processing and memory capacity (for review, see Crameri et al., 2019).

Based on the literature summarized above, CF can be defined as a "multi-faceted and differentially malleable capacity to deploy neurocognitive resources, knowledge, and skills to meet the demands of operational task performance, and to sustain this performance throughout a career- and life-long application."

The key aspects of this working definition are:

1. CF entails multiple capacity factors that are different from knowledge and skills.

2. Operational task performance is core to defining the composition and relative importance of these factors.
3. The malleability of individual CF factors is to be established empirically.
4. Sustained performance implies resilience, well-being and longevity as by-products of fitness.

Several factors impacting task performance and career longevity in high stakes occupations can be considered for inclusion in the CF set. **Table 1** summaries these operationally relevant constructs and notionally allocates them to the phases of the operational readiness cycle: from the foundational (cognitive gym) to advanced (readiness), mission-ready (operational tolerances), and recovery phase.

In particular, cognitive primaries, such as *attention*, *executive control*, and *co-action*, that underpin most cognitive skills, may potentially be trainable through the gold-standard "isolate—overload—recover" regimes at the initial, Cognitive Gym phase (Temby et al., 2015). More advanced training targets at the CR phase include *stress management* (arousal regulation and "getting in The Zone on cue"), *adaptability*, *teamwork*, *situation awareness* (including detection, sense-making and prediction) and *decision making* (de-biasing and confidence calibration). The mission-ready training phase is focused on *tolerances* (from pain and sleep loss to monotony, frustration and uncertainty) and *resistances* (to distraction, deception or manipulation). These capacities get further enhanced through operational support (with tools such as *decision aids* and *fatigue countermeasures*). The recovery phase completes the cycle, with its role widely recognized by expert consensus (Kellmann et al., 2018; Reardon et al., 2019), employing both reflexive (e.g., mindfulness) and restorative practices (e.g., healthy eating, hydration and sleep hygiene) and relying on social support. The full cycle reinforces the cognitive fundamentals that are known for their contribution to both real-time cognitive performance under challenging conditions (Blank et al., 2014; Crameri et al., 2019) and the resilience that enables career longevity and life-long thriving (Seligman, 2008; Cornum et al., 2011).

PERIODIZATION OF THE COGNITIVE FITNESS CYCLE

Some of the most important questions about the CF are centered around the sequential periodization of cognitive training. **Figure 1** shows a hypothesized sequence in which these elements can be addressed in the training cycle across the CF2. The following questions seem worth addressing in future research:

1. What are the pattern-forming factors of CF? The factors that can drive change (both positive and negative) in multiple other factors. They should be trained first and foremost.
2. On what basis can training of some CF constructs be prioritized over the others?
3. What are the design principles for periodization of CF training to distinct micro-, meso- and macro-cycles to address the daily, weekly, monthly and yearly timescales (e.g., DeWiggins et al., 2010)?

TABLE 1 | Phases of cognitive fitness (CF) cycle: from training to operational performance and recovery.

Phase	The domain of cognitive functioning	Target constructs	Examples of training/development objectives[a]
Cognitive gym: Foundational training	Cognitive fitness:	Self-awareness	Stress symptoms detection
	Trainable cognitive primaries	Attention	Focus endurance
		Task switching	Dual-tasking
		Impulse control	Response override
		Co-action	Action mirroring
			Handover execution
Advanced cognitive training	Cognitive skills	Controlled response	Effortless concentration
		Energy management	Arousal regulation
			Progressive muscle relaxation
			Resonant frequency breathing
		Situation awareness	Perceptual acuity (detection)
			Sense-making (interpretation)
			Anticipatory skills (prediction)
		Decision making	Pattern recognition
			Bias detection and mitigation
			Confidence calibration
		Adaptability	Change detection
			Cognitive flexibility
		Teamwork	Shared mental models
			Non-verbal communication
Mission-ready training	Tolerance and resistance	Tolerances	Generalized discomfort tolerance
			Pain tolerance
			Alertness upregulation (drowsiness tolerance)
			Mental effort tolerance
			Monotony tolerance
			Frustration tolerance
			Ambiguity tolerance
			Startle/surprise tolerance
		Resistances	Distractor resistance
			Susceptibility to deception
			Resistance to manipulation
		Task resilience	Error detection
			Performance recovery
Operational augmentation	Operational task performance	Cognitive state	Alertness monitoring[b]
		Cognitive workload	Fatigue countermeasures[b]
		Decision making	Adaptive decision aids[b]
		Equipment use	Operator state-aware autonomy[b]
Recovery	Cognitive recovery	Sleep recovery	Sleep/wake cycle management
		Reflexive practices	Mindfulness and meditation
		Nutrition	Healthy eating habits[c]
			Hydration[c]
		Connectedness	Interactional competence
			Social support[c]

Notes: [a]The list is not exhaustive and is subject to validation through expert consensus. These training objectives are not directly linked to operational task performance. Similar to strength and conditioning in physical training, the products of CF training feed into the subsequent cognitive skills training, and only through this skill training—into operational performance. [b]Given the nature of the Operational Augmentation phase, performance augmentation tools are listed here instead of training objectives. [c]The recovery phase is focused on the development of habits and practices that promote cognitive fitness and, as such, they are applicable to all other phases of the Cognitive Fitness Cycle.

4. How long is the optimal training cycle for a given CF attribute? Or how can this cycle be tailored for an individual?

5. Points of insertion into a holistic training cycle—the following options can be considered:

   (a) CF training can be combined with existing training modalities that address physical fitness, technical skill acquisition or tactical skill application (Ward et al., 2017).

   (b) CF training can be combined with mental/psychological skills training—the traditional province of sport psychology where self-management skills get added or overlaid to the already acquired fitness, technical and tactical skills (Birrer and Morgan, 2010).

   (c) CF training can be a stand-alone modality—a largely neglected area where cognitive capacity training becomes part of the foundational holistic fitness.

It is safe to assume that the selection of insertion points will be guided by the task, context and resources available. In addition, their relative effectiveness can be compared directly (Röthlin et al., 2016).

FIGURE 1 | Cognitive fitness framework (CF2).
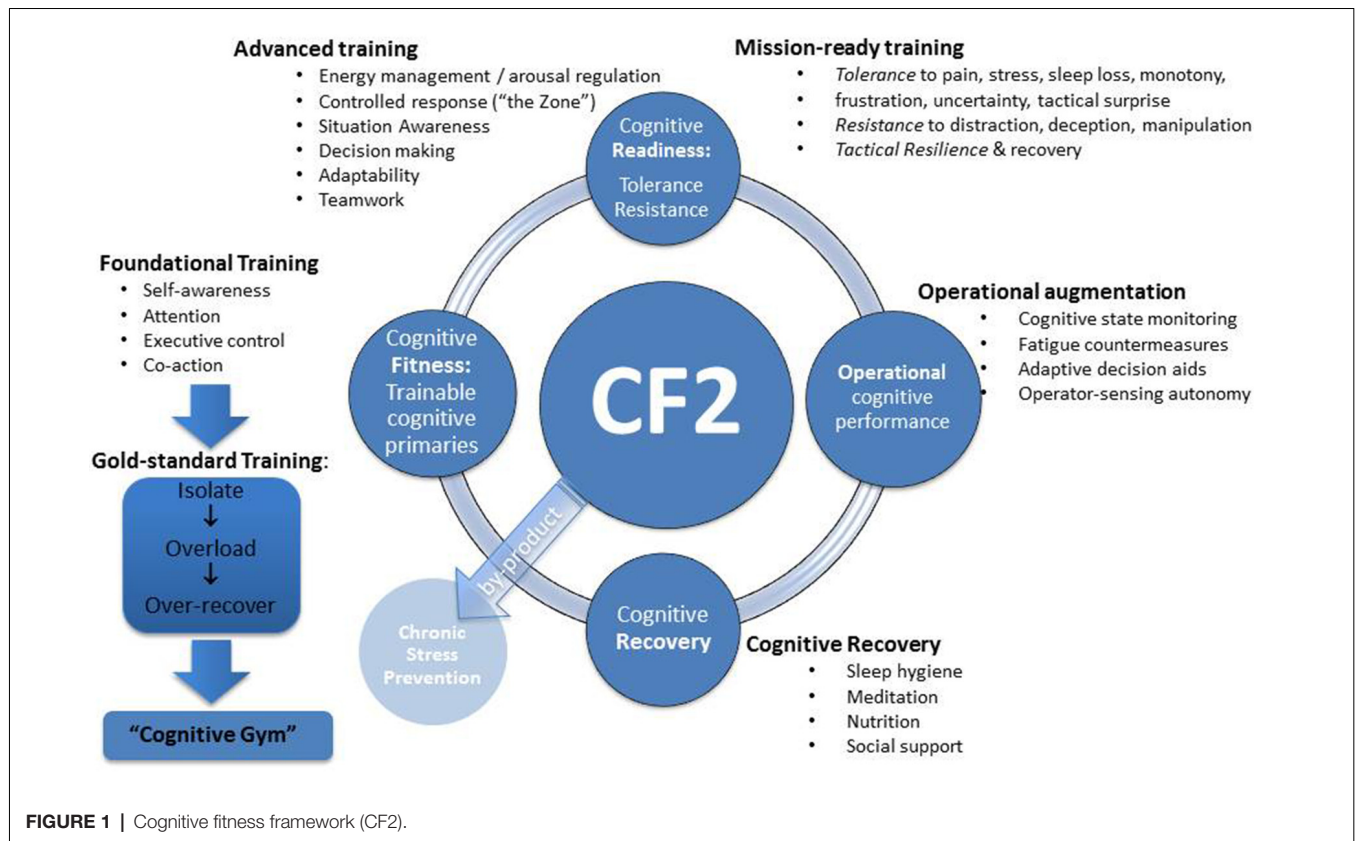
APPLICATIONS

The CF2 is a step towards a consensus on the selection of cognitive attributes to train, the limits of their trainability, and the methods of assessing them in the 21st-century workforce. It helps to map out various lines of research effort and see where individual projects fit. For example, current research on foundational cognitive training is progressing under the construct of Cognitive Gym (Temby et al., 2015; Jarvis et al., 2019). On the other hand, the emerging work on team decision making has been driven by research evidence on the so-called *c*-factor (collective intelligence, or team "smarts"). This *c*-factor has surprisingly little connection to the individual team members IQ, and instead has been linked to their interactional competence and team diversity (Woolley et al., 2010; Blanchard et al., 2018).

One of the core predictors of fitness and performance is executive functioning (EF)—a primary cognitive capacity underpinning self-discipline, attentional focus and impulse control. Its known predictive links include BMI and cardiovascular health (Schlam et al., 2013), learning outcomes in academic and occupational training settings, injury incidence and overtraining, job performance and resilience, post-traumatic stress and other mental health vulnerabilities (Moffitt et al., 2011). Even a modest gain in EF capacity (either through selection or training) is known to drive population-wide gains in health, learning achievement in education and training (Moffitt et al., 2013), and productivity/work safety (including injuries).

Twin studies indicate heavy genetic influences on EF (for review, see Friedman and Miyake, 2017), while longitudinal research shows substantial intra-individual variability (Moffitt et al., 2011). Direct estimation of EF heritability (vs. trainability) would go a long way towards informing wide-ranging investment decisions about selection and training programs. Adding this cognitive mediator to the original predictor set would make both performance and health prediction models more comprehensive and holistic.

CONCLUSIONS AND FUTURE DIRECTIONS

The construct of CF developed here is a means to address the gap between the CR, RDoC and MF literatures by offering a unifying framework to integrate the multitude of biologically traceable factors underpinning individuals' performance in cognitively demanding tasks, to assess their trainability and inform the development of methods to improve them through training and augmentation.

The CF2 is a working hypothesis mapping out the research agenda to identify and measure key attributes of CF, underpinning both real-time cognitive performance under challenging conditions and the resilience that enables career longevity and life-long thriving. CF2 also offers a hypothesized sequence for cognitive training in a CF training cycle. As a hypothesis, CF2 requires testing and validation. One of the most promising validation avenues is through the development of an

expert consensus on the key CF ingredients and their relative importance in high-performance settings. The Delphi method utilized in expert consensus studies of cognitive functioning in mental health (see Yücel et al., 2019) appears a good fit for validating the CF2 hypothesis. Once the relative importance of CF constructs is confirmed through expert consensus, training protocol evaluation studies can inform the selection of training methods that are best suited for each CF construct, including the formulation of training objectives to complement the training targets for each training phase in CF2.

The next challenge is extending the range of measurement of the assessment tools measuring CF constructs to cover both cognitive deficit and gifted performance and to employ best-practice measurement protocols to improve the reliability, validity and utility of these assessment tools. These improvements in the measurement of CF constructs are critical to stimulating the design and development of the environments and protocols to improve CF, and to developing fieldable technologies to protect and enhance cognitive performance.

REFERENCES

Adler, A. B., Bliese, P. D., Pickering, M. A., Hammermeister, J., Williams, J., Harada, C., et al. (2015). Mental skills training with basic combat training soldiers: a group-randomized trial. *J. Appl. Psychol.* 100, 1752–1764. doi: 10.1037/apl0000021

Ahn, B., and Cox, M. F. (2016). Knowledge, skills and attributes of graduate student and postdoctoral mentors in undergraduate research settings. *J. Eng. Educat.* 105, 605–629. doi: 10.1002/jee.20129

Aidman, E. (2017). The cognitive fitness framework: a roadmap for systematic, evidence-based mental skills training and performance enhancement. *J. Sci. Med. Sport* 20, S50–S52. doi: 10.1016/j.jsams.2017. 09.159

Aidman, E. (2005). "Measuring subjective components of fatigue: the role of effort perception," in *Promoting Health and Performance for Life: Proceedings of the ISSP 11th World Congress of Sport Psychology*, eds T. Morris, P. Terry, S. Gordon, S. Hanrahan, L. Ievleva, G. Kolt, et al. (Sydney: International Society of Sport Psychology).

Aidman, E., Crampton, J., Brown, T., and Bond, J. (2016). "Mental skills training meets future neuro-technologies: enhancing cognitive performance through systematic intervention," in *Paper Presented at the 2016 APS Congress*, Melbourne, VIC, Australia.

Aidman, E., Galanis, G., Manton, J., Vozzo, A., and Bonner, M. (2002). Evaluating human systems in military training. *Aust. J. Psychol.* 54, 168–173. doi: 10.1080/00049530412331312754

Appelbaum, P. (2017). Moving toward the future in the diagnosis of mental disorders. *Psychol. Sci. Public Interest* 18, 67–71. doi: 10.1177/1529100617727267

Baker, H. K., and Phillips, A. L. (2000). Knowledge, skills and attributes needed to succeed in financial management: evidence from entry-level and mid-level practitioners. *Financ. Pract. Educ.* 10, 189–200.

Birrer, D., and Morgan, G. (2010). Psychological skills training as a way to enhance an athlete's performance in high-intensity sports. *Scand. J. Med. Sci. Sports* 80, 78–87. doi: 10.1111/j.1600-0838.2010.01188.x

Blacker, K. J., Hamilton, J., Roush, G., Pettijohn, K. A., and Biggs, A. T. (2019). Cognitive training for military application: a review of the literature and practical guide. *J. Cogn. Enhancement* 3, 30–51. doi: 10.1007/s41465-018-0076-1

Blanchard, M., Kleitman, S., Jackson, S., Rizvandi, N., and Aidman, E. (2018). "Dynamic scenario-based cognitive assessment: preliminary findings and implications for cognition and metacognition," in *Paper Presented at DHSS-2018*, Perth, Western Australia.

Blank, M., Sylvia, B. G., Garza, J. G., and Wade, B. R. (2014). Sapper-athlete-warrior program: an integrated approach to periodized warrior fitness. *J. Sport Psychol. Action* 5, 73–87. doi: 10.1080/21520704.2014. 892912

Boccia, M., Rosella, M., Vecchione, F., Tanzilli, A., Palermo, L., D'Amico, S., et al. (2017). Enhancing allocentric spatial recall in pre-schoolers through navigational training programme. *Front. Neurosci.* 11:574. doi: 10.3389/fnins. 2017.00574

Bogga, D. (2017). "Training mental fitness," in *Human Dimension*, ed. L. Sinclair (Canberra, ACT: Centre for Army Learning), 18–27.

Bolstad, C. A., Cuevas, H. M., Babbitt, B. A., Semple, C. A., and Vestewig, R. E. (2006). "Predicting cognitive readiness of military health teams," in *16th World Congress of the International Ergonomics Association*, Maastricht, Netherlands.

Bowers, C., and Cannon-Bowers, J. (2014). "Cognitive readiness for complex team performance," in *Teaching and Measuring Cognitive Readiness*, eds H. F. O'Neil, R. S. Perez, and E. L. Baker (New York, NY: Springer), 301–323.

Carcone, D., and Ruocco, A. C. (2017). Six years of research on the national institute of mental health's research domain criteria (RDoC) initiative: a systematic review. *Front. Cell. Neurosci.* 11:46. doi: 10.3389/fncel.2017. 00046

Chandra, S., and Leong, F. T. (2016). A diversified portfolio model of adaptability. *Am. Psychol.* 71, 847–862. doi: 10.1037/a0040367

Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W., and Reed, G. (2017). Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5 and the national institute of mental health's research domain criteria (RDoC). *Psychol. Sci. Public Interest* 18, 72–145. doi: 10.1177/1529100617727266

Cooper, R., and Fry, C. (2018). *Aviation Non-Technical Skills Guidebook*. Canberra, ACT: Defence Flight Safety Bureau.

Cornum, R., Matthews, M. D., and Seligman, M. E. P. (2011). Comprehensive soldier fitness: building resilience in a challenging institutional context. *Am. Psychol.* 66, 4–9. doi: 10.1037/a0021420

Cosenzo, K. A., Fatkin, L. T., and Patton, D. J. (2007). Ready or not: enhancing operational effectiveness through use of readiness measures. *Aviat. Space Environ. Med.* 78, B96–B106.

Crameri, L., Hettiarachchi, I., and Hanoun, S. (2019). A review of individual operational cognitive readiness: theory development and future directions. *Hum. Factors* doi: 10.1177/0018720819868409 [Epub ahead of print].

Cuthbert, B. N., and Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med.* 11:126. doi: 10.1186/1741-70 15-11-126

DATA AVAILABILITY STATEMENT

No datasets were generated or analyzed for this study.

AUTHOR CONTRIBUTIONS

EA developed the conceptual framework for the article, drafted, revised, and approved the manuscript.

FUNDING

ACKNOWLEDGMENTS

DeWiggins, S., Hite, B., and Alston, V. (2010). Personal performance plan: application of mental skills training to real-world military tasks. *J. Appl. Sport Psychol.* 22, 458–473. doi: 10.1080/10413200.2010.500606

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: a general overview. *Acad. Emerg. Med.* 15, 988–994. doi: 10.1111/j.1553-2712.2008.00227.x

Fadde, P. J., and Zaichkowsky, L. (2018). Training perceptual-cognitive skills in sports using technology. *J. Sport Psychol. Action* 9, 239–248. doi: 10.1080/21520704.2018.1509162

Fatkin, L. T., and Patton, D. (2008). "Mitigating the effects of stress through cognitive readiness," in *Performance Under Stress*, eds P. A. Hancock and J. L. Szalma (Aldershot, UK: Ashgate Publishing Limited), 209–229.

Fletcher, J. D., and Wind, A. P. (2014). "The evolving definition of cognitive readiness for military operations," in *Teaching and Measuring Cognitive Readiness*, eds H. F. O'Neil, R. S. Perez and E. L. Baker (New York, NY: Springer), 24–52.

Foster, G. D. (1996). Military readiness: a case of rhetorical overkill. *Rusi J.* 141, 28–29. doi: 10.1080/03071849608446028

Friedman, N. P., and Miyake, A. (2017). Unity and diversity of executive functions: individual differences as a window on cognitive structure. *Cortex* 86, 186–204. doi: 10.1016/j.cortex.2016.04.023

Genet, J. J., and Siemer, M. (2011). Flexible control in processing affective and non-affective material predicts individual differences in trait resilience. *Cogn. Emot.* 25, 380–388. doi: 10.1080/02699931.2010.491647

Grier, R. A. (2011). Cognitive readiness at the tactical level: a review of measures. *Proc. Hum. Factors Ergon. Soc.* 55, 404–408. doi: 10.1037/e578902012-083

Grier, R. A. (2012). Military cognitive readiness at the operational and strategic level: a theoretical model for measurement development. *J. Cogn. Eng. Decis. Mak.* 6, 358–392. doi: 10.1177/1555343412444606

Grier, R. A., Fletcher, J. D., and Morrison, J. E. (2012). Defining and measuring military cognitive readiness. *Proc. Hum. Factors Ergon. Soc.* 56, 435–437. doi: 10.1177/1071181312561098

Gross, A. L., Parisi, J. M., Spira, A. P., Kueider, A. M., Ko, J. Y., Saczynski, J. S., et al. (2012). Memory training interventions for older adults: a meta-analysis. *Aging Ment. Health* 16, 722–734. doi: 10.1080/13607863.2012.667783

Herzog, T. P., and Deuster, P. A. (2014). Performance psychology as a key component of human performance optimization. *J. Spec. Oper. Med.* 14, 99–105.

Jarvis, A., Temby, P., Hoggan, B., and Ma-Wyatt, A. (2019). "Multiple object tracking ability, working memory capacity and individual differences as predictors of complex task performance," in *Paper Presented at Defence Human Science Symposium*, Canberra, Australia.

Jeffreys, I., and Moody, J. (Eds). (2016). *Strength and Conditioning for Sports Performance.* London: Routledge.

Joyce, S., Shand, F., Lal, T. J., Mott, B., Bryant, R. A., and Harvey, S. B. (2019). Resilience@ Work mindfulness program: results from a cluster randomized controlled trial with first responders. *J. Med. Internet Res.* 21:e12894. doi: 10.2196/12894

Kellmann, M., Bertollo, M., Bosquet, L., Brink, M., Coutts, A. J., Duffield, R., et al. (2018). Recovery and performance in sport: consensus statement. *Int. J. Sports Physiol. Perform.* 13, 240–245. doi: 10.1123/ijspp.2017-0759

Keyes, L. M. (2007). Promoting and protecting mental health as flourishing: a complementary strategy for improving national mental health. *Am. Psychol.* 62, 95–108. doi: 10.1037/0003-066x.62.2.95

Kluge, A., and Burkolter, D. (2013). Enhancing research on training for cognitive readiness: research issues and experimental designs. *J. Cogn. Eng. Decis. Mak.* 7, 96–118. doi: 10.1177/1555343412446483

Mason, P. A., and McQuade, J. S. (2013). Preface to the special issue on exploring cognitive readiness in complex operational environments: advances in theory and practice, part III. *J. Cogn. Eng. Decis. Mak.* 7, 66–68. doi: 10.1177/1555343412456025

McCarthy, D. (1964). Mental fitness. *Am. Psychol.* 19, 201–202.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., et al. (2011). A gradient of childhood self-control predicts health, wealth and public safety. *Proc. Natl. Acad. Sci.* 108, 2693–2698. doi: 10.1073/pnas.1010076108

Moffitt, T. E., Poulton, R., and Caspi, A. (2013). Lifelong impact of early self-control. *Am. Sci.* 101:352. doi: 10.1511/2013.104.352

Morris, T. E., and Summers, J. E. (2004). *Sport Psychology: Theory, Applications and Issues.* Brisbane, QLD: John Wiley & Sons.

Morris, S. E., and Cuthbert, B. N. (2012). Research domain criteria: cognitive systems, neural circuits and dimensions of behavior. *Dialogues Clin. Neurosci.* 14, 29–37.

Morrison, J. E., and Fletcher, J. D. (2002). *Cognitive Readiness (P-3735).* Alexandria, VA: Institute for Defense Analyses.

O'Neil, H. F., Perez, R. S., and Baker, E. L. (2014). *Teaching and Measuring Cognitive Readiness.* New York, NY: Springer.

Page, J. W., Asken, M. J., Zwemer, C. F., and Guido, M. (2016). Brief mental skills training improves memory and performance in high stress police cadet training. *J. Police Crim. Psychol.* 31, 122–126. doi: 10.1007/s11896-015-9171-8

Palmiero, M., Giulianella, L., Guariglia, P., Boccia, M., D'Amico, S., and Piccardi, L. (2019). The dancers' visuospatial body map explains their enhanced divergence in the production of motor forms: evidence in the early development. *Front. Psychol.* 10:1274. doi: 10.3389/fpsyg.2019.01274

Passell, E., Dillon, D. G., Baker, J. T., Vogel, S. C., Scheuer, L. S., Mirin, N. L., et al. (2019). *Digital Cognitive Assessment: Results from the TestMyBrain NIMH Research Domain Criteria (RDoC) Field Test Battery Report.* National Institute of Mental Health Report. Available online at: psyarxiv.com. Accessed August 21, 2019.

Peña, J., and Brody, N. (2016). Increasing cognitive readiness through computer and videogame-based training. *Commun. Res.* 43, 571–592. doi: 10.1177/0093650214554614

Reardon, C. L., Hainline, B., Aron, C. M., Baron, D., Baum, A. L., Bindra, A., et al. (2019). Mental health in elite athletes: international olympic committee consensus statement (2019). *Br. J. Sports Med.* 53, 667–699. doi: 10.1136/bjsports-2019-100715

Redick, T. S. (2019). The hype cycle of working memory training. *Curr. Dir. Psychol. Sci.* 28, 423–429. doi: 10.1177/0963721419848668

Robinson, P., Oades, L. G., and Caputi, P. (2015). Conceptualising and measuring mental fitness: a delphi study. *Int. J. Wellbeing* 5, 53–73. doi: 10.5502/ijw.v5i1.4

Röthlin, P., Birrer, D., Horvath, S., and grosse Holtforth, M. (2016). Psychological skills training and a mindfulness-based intervention to enhance functional athletic performance: design of a randomized controlled trial using ambulatory assessment. *BMC Psychol.* 4:39. doi: 10.1186/s40359-016-0147-y

Schlam, T. R., Wilson, N. L., Shoda, Y., Mischel, W., and Ayduk, O. (2013). Preschoolers' delay of gratification predicts their body mass 30 years later. *J. Pediatr.* 162, 90–93. doi: 10.1016/j.jpeds.2012.06.049

Seligman, M. E. P. (2008). Positive health. *Appl. Psychol.Int. Rev.* 57, 3–18.

Sherlin, L. H., Larson, N. C., and Sherlin, R. M. (2013). Developing a performance brain training™ approach for baseball: a process analysis with descriptive data. *Appl. Psychophysiol. Biofeedback* 38, 29–44. doi: 10.1007/s10484-012-9205-2

Schunk, D. H., and Greene, J. A. (2017). *Handbook of Self-Regulation of Learning and Performance.* New York, NY: Routledge.

Sotos, J. G. (2019). Physical fitness programs don't fit today's fight. *U.S. Naval Inst. Proc.* 145:1396.

Temby, P., Fidock, J., Oostergo, M., and Shillabeer, C. (2015). "Enhancing cognitive fitness in the Australian army—building the cognitive gym," in *Paper Presented at Defence Human Sciences Symposium*, Adelaide, Australia.

Vie, L. L., Scheier, L. M., Lester, P. B., and Seligman, M. E. P. (2016). Initial validation of the U.S. army global assessment tool. *Mil. Psychol.* 28, 468–487. doi: 10.1037/mil0000141

Walton, C. C., Keegan, R. J., Martin, M., and Hallock, H. (2018). The potential role for cognitive training in sport: more research needed. *Front. Psychol.* 9:1121. doi: 10.3389/fpsyg.2018.01121

Ward, N., Paul, E., Watson, P., Cooke, G. E., Hillman, C. H., Cohen, N. J., et al. (2017). Enhanced learning through multimodal training: evidence from a comprehensive cognitive, physical fitness, and neuroscience intervention. *Sci. Rep.* 7:5808. doi: 10.1038/s41598-017-06237-5

Wesemann, U., Willmund, G. D., Ungerer, J., Kreim, G., Zimmermann, P. L., Bühler, A., et al. (2018). Assessing psychological fitness in the military—development of an effective and economic screening instrument. *Mil. Med.* 183, e261–e269. doi: 10.1093/milmed/usy021

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., and Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science* 330, 686–688. doi: 10.1126/science.1193147

Yücel, M., Oldenhof, E., Ahmed, S. H., Belin, D., Billieux, J., Bowden-Jones, H., et al. (2019). A transdiagnostic dimensional approach towards a neuropsychological assessment for addiction: an international Delphi consensus study. *Addiction* 114, 1095–1109. doi: 10.1111/add.14424

Zaichkowsky, L., and Peterson, D. (2018). *The Playmaker's Advantage: How to Raise Your Mental Game to the Next Level.* New York, NY: Simon and Schuster.

Zhang, L., Aidman, E., Burns, B., and Kleitman, S. (2019). "Measuring individual differences in adaptability: integrating self-report and performance-based assessment," in *Paper Presented at Defence Human Science Symposium,* Canberra, Australia.

Investigating the "Flow" Experience: Key Conceptual and Operational Issues

Sami Abuhamdeh*

Department of Psychology, Istanbul Şehir University, Istanbul, Turkey

The "flow" experience (Csikszentmihalyi, 1975) has been the focus of a large body of empirical work spanning more than four decades. Nevertheless, advancement in understanding – beyond what Csikszentmihalyi uncovered during his initial breakthrough in 1975 – has been modest. In this conceptual analysis, it is argued that progress within the field has been impeded by a lack of consistency in how flow is operationalized, and that this inconsistency in part reflects an underlying confusion regarding what flow is. Flow operationalizations from papers published within the past 5 years are reviewed. Across the 42 reviewed studies, flow was operationalized in 24 distinct ways. Three specific points of inconsistency are then highlighted: (1) inconsistences in operationalizing flow as a continuous versus discrete construct, (2) inconsistencies in operationalizing flow as inherently enjoyable (i.e., "autotelic") or not, and (3) inconsistencies in operationalizing flow as dependent on versus distinct from the task characteristics proposed to elicit it (i.e., the conditions/antecedents). After tracing the origins of these discrepancies, the author argues that, in the interest of conceptual intelligibility, flow should be conceptualized and operationalized exclusively as a discrete, highly enjoyable, "optimal" state of consciousness, and that this state should be clearly distinguished from the conditions proposed to elicit it. He suggests that more mundane instances of goal-directed engagement are better conceived and operationalized as variations in task involvement rather than variations in flow. Additional ways to achieve greater conceptual and operational consistency within the field are suggested.

Keywords: flow, enjoyment, task involvement, intrinsic motivation, critical review

INVESTIGATING THE "FLOW" EXPERIENCE: KEY CONCEPTUAL AND OPERATIONAL ISSUES

Csikszentmihalyi (1975) introduced the concept of "flow" 42 years ago in his groundbreaking book *Beyond Boredom and Anxiety*. The concept of flow was not entirely new – the experience itself held much in common with Maslow's (1964) conception of "peak experience," as well as accounts of ecstatic experiences by Laski (1961). However, Csikszentmihalyi's approach was appreciably more

FIGURE 1 | The characteristics and conditions of flow (from Nakamura and Csikszentmihalyi, 2002).

Characteristics of the flow state:

- Intense and focused concentration on what one is doing in the present moment
- Merging of action and awareness
- Loss of reflective self-consciousness
- A sense that one can control one's actions
- Distortion of temporal experience
- Experience of the activity as intrinsically rewarding

The conditions (i.e. antecedents) of flow:

- Perceived challenges, or opportunities for action, that neither overmatch nor underutilize existing skills
- Clear proximal goals
- Immediate feedback about the progress that is being made

systematic and empirically driven than previous approaches. Within a few years, flow was the focus of hundreds of empirical studies from a diversity of fields including educational psychology, recreation and leisure sciences, game design, and many others.

Over the years, many predictors and consequences of "flow"[1] have been identified (e.g., Jackson and Roberts, 1992; Csikszentmihalyi et al., 1993; Jackson et al., 2001; Demerouti, 2006; Schüler, 2007; Stavrou et al., 2007; Engeser and Rheinberg, 2008; Fullagar and Kelloway, 2009; Nielsen and Cleal, 2010; Bakker et al., 2011; Rodríguez-Sánchez et al., 2011; Seger and Potts, 2012; Coffey et al., 2016). But what have we learned about flow itself – about the state of *optimal experience* – since Csikszentmihalyi introduced the concept in 1975? Here, the view is sobering. The conceptualization introduced in 1975 remains essentially unchanged. Furthermore, fundamental questions persist. [For example, although flow is conceptualized as a multifaceted construct (**Figure 1**), very little is known regarding its latent structure – the causal relations among its proposed components, the relative contribution of each component to the overall flow experience, etc.]. Indeed, and perhaps most alarming, after almost 42 years of research, there appears to be significant disagreement among researchers regarding what flow actually is and how to measure it. This last point can best be appreciated by first reviewing the many different ways in which flow has been operationalized in the literature.

A REVIEW OF FLOW OPERATIONALIZATIONS IN THE PSYCHOLOGICAL LITERATURE

Within any field of science, the consensual operationalization of central constructs is a *sine qua non* for progress. When this is lacking, results across studies cannot be compared, and the potential for progress in the field is severely undermined. To examine the degree of consistency with which flow has been operationalized within the psychological literature, a review was conducted, limited to publications from the past 5 years[2]. A PsychINFO search yielded the 42 publications listed in **Table 1** (see the **Appendix** for the specific inclusion criteria used to select these publications). As shown in the first column, across the 42 reviewed studies, flow was operationalized in 24 distinct ways. Furthermore, the differences between these operationalizations were often considerable, so that the meaning of "flow" often changed dramatically from one study to the next.

The fourth, fifth, and sixth columns of **Table 1** indicate three key ways in which the operationalizations differed. Column 4 indicates whether flow was operationalized as a continuous versus discrete construct in each study. Column 5 indicates whether flow was operationalized as enjoyable (i.e., "autotelic") or not. Column 6 indicates whether flow was operationalized using one or more of its proposed antecedents (i.e., clear goals, immediate feedback, and a balance of challenge and skill).

In the remainder of this conceptual analysis, I elaborate the nature of the three issues highlighted in **Table 1** and attempt to trace their origins. Based on my reading of Csikszentmihalyi's conceptualization of flow, I suggest that most operationalizations of flow currently found in the literature miss the mark. I argue that flow should be conceptualized and operationalized exclusively as a state of optimal experience – that is, as a discrete, highly rewarding state of consciousness – and that the potential for progress in our understanding of flow largely depends on it.

THE THREE ISSUES

Issue 1: Is Flow a Discrete or Continuous Construct?

Many psychological constructs, such as happiness, anxiety, and self-efficacy, represent continuous (i.e., spectrum and dimensional) constructs. At any given moment, your happiness may be very low, very high, or anything in between. Other psychological constructs, such as euphoria, fury, and the "suicidal mode" (Rudd, 2000), represent discrete (i.e., categorical and taxonic) constructs. Although it may be possible to locate them on a continuum, they are not applicable to its full range. Occasionally it is not entirely clear whether a construct is continuous or discrete. When this happens in the realm of science, fierce debate usually ensues in an attempt to resolve the conflict. An example of this can be found in the field of abnormal

---

[1] Here I put "flow" in quotes because, as will be shown, most studies of flow haven't operationalized flow as conceptualized by Csikszentmihalyi – as a (discrete) state of optimal experience.

TABLE 1 | Flow operationalizations in the psychological literature from the past 5 years.

Flow operationalization	Source	Continuous or discrete?	Enjoyment included?[1]	Flow condition(s) included?
Flow Short Scale (Rheinberg et al., 2003) and related scales[2]	Baumann et al., 2016 Barros et al., 2018 Brom et al., 2017 Harris et al., 2017a,b Hermann and Vollmeyer, 2016 Schattke et al., 2014	Continuous	No	Yes (partly)
Four items from the absorption subscale of the Flow Short Scale	Rivkin et al., 2018	Continuous	No	No
Flow Short Scale with three additional items measuring "autotelic experience"	Rankin et al., 2019	Continuous	Yes	Yes (partly)
Flow State Scale (Jackson and Marsh, 1996) and related scales	Borovay et al., 2019 Beltrán et al., 2018 Forkosh and Drake, 2017 Harmat et al., 2015 Joo et al., 2015 Kaye et al., 2018 Marston et al., 2016	Continuous	Yes	Yes (partly)
As above	Kawabata and Evans, 2016	Discrete	Yes	Yes (partly)
17 of the 36 items in the Flow State Scale	Lin et al., 2019	Continuous	Yes	Yes (partly)
3 of 9 subscales from Flow State Scale	Matthews, 2015	Continuous	No	No
Core Flow Scale (Martin and Jackson, 2008)	Kocjan and Avsec, 2017	Continuous	Yes	No
"3-Channel" flow model (Csikszentmihalyi, 1975)	Huskey et al., 2018 Chen and Sun, 2016 Sun et al., 2017	Discrete	No	Yes (fully)
"Quadrant" flow model (Massimini and Carli, 1988)	Ilies et al., 2017 Sather et al., 2017	Discrete	No	Yes (fully)
Three items measuring interest, enjoyment, and absorption	Bricteux et al., 2017	Continuous	Yes	No
Three items measuring absorption	Vanwesenbeeck et al., 2016	Continuous	No	No
Ten items measuring interest, attention, and control	Cho, 2018	Continuous	Yes	No
Eight items intended to measure conditions and experience of flow	Wanzer et al., 2018	Continuous	Yes	Yes (partly)
Ps presented with description of flow and asked how much their own experience emulated it	Kennedy et al., 2014 Vuorre and Metcalfe, 2016	Continuous	Yes	No
Eight of nine items previously used by Hektner et al. (2007)	Kulkarni et al., 2016	Continuous	Yes	Yes (partly)
A 28 item flow scale (Chou and Ting, 2003)	Soutter and Hitchens, 2016	Continuous	Yes	No
Eleven items taken from Kwak et al. (2014).	Brailovskaia et al., 2018	Continuous	Yes	No
Flow Scale for Games (Kiili, 2006)	Hou, 2015	Continuous	Yes	Yes (partly)
Flow subscale of game engagement questionnaire (Poels et al., 2007)	Dixon et al., 2019	Continuous	No	No
Flow subscale of game engagement questionnaire (Brockmyer et al., 2009)	Smith et al., 2017	Continuous	Yes	No
Three questions prefaced by description of flow (Novak et al., 2000)	Rodríguez-Ardura and Meseguer-Artola, 2017	Continuous	No	No
An 8-item flow scale (Waterman et al., 2003)	Bonaiuto et al., 2016 Mao et al., 2016	Continuous	No	Yes (partly)
Flow questionnaire (Csikszentmihalyi and Csikszentmihalyi, 1988)	Lavoie and Main, 2019	Discrete	Yes	No

[1]Operationalizations which included one or more items measuring "autotelic experience" (i.e. intrinsically motivating), but did not include items measuring "enjoyment" specifically, were nevertheless classified as having an enjoyment component, given that intrinsic motivation implies enjoyment. [2]The meaning of "related scales": We did not distinguish between long versus short versions of scales, nor did we distinguish between older versus newer versions of scales, nor did we distinguish between original versus translated versions of scales. They were all considered to be versions of the same scale and are not differentiated in the table.

FIGURE 2 | The quadrant model of flow. Challenge and skill scores represent within-person *z*-scores.
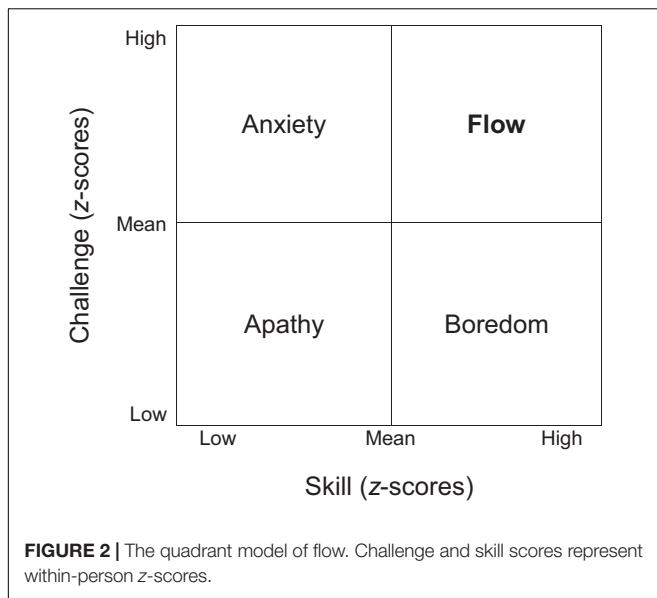
psychology, where the designation of psychological disorders as continuous versus discrete has been hotly contested.

Looking at Column 4 of **Table 1**, we can see that in a majority of the studies flow was operationalized as a continuous construct, applicable to the full range of participants' experience in varying degrees. For example, the Flow State Scale-2 (Jackson and Eklund, 2002) composed of items intended to tap the six experiential characteristics of flow, as well as the three conditions (**Figure 1**), asks participants to indicate the extent to which the items characterize their experience in a just-completed activity on a 5-point Likert scale, ranging from 1 ("strongly agree") to 5 ("strongly disagree"). Responses to the items are usually averaged to compute a single "flow" score for each and every observation.

A few studies, in contrast, operationalized flow as a discrete construct. For example, two studies which used the experience sampling method (Csikszentmihalyi et al., 1977) used a "quadrant" approach popularized earlier by Csikszentmihalyi and his colleagues (e.g., Csikszentmihalyi and Csikszentmihalyi, 1988; Massimini and Carli, 1988; **Figure 2**). Using this approach, flow is operationalized as any observation in which both perceived challenge and perceived skill are both "high" (i.e., above the person's average).

So is flow a continuous construct which exists in greater or lesser degrees across the full range of human experience (like happiness, for example)? Or is it a discrete state that is sometimes experienced, but usually not? In the preface to *Beyond Boredom and Anxiety*, Csikszentmihalyi described flow as such:

"On the rare occasions that it happens, we feel a sense of exhilaration, a deep sense of enjoyment that is long cherished and that becomes a landmark in memory for what life should be like. This is what we mean by "optimal experience." (p. ii)

Also from the preface:

"From their accounts of what it felt like to do what they were doing, I developed a theory of optimal experience based on the concept of

flow – the state in which people are so involved in an activity that nothing seems to matter; the experience itself is so enjoyable they will do it even at great cost, for the sheer sake of doing it." (p. iv)

As is evident from the passages above (and many others), Csikszentmihalyi conceptualized flow as an "optimal" state of consciousness, one that usually occurs relatively rarely in life. You can be in flow, or not in flow. When you are not in flow, Csikszentmihalyi referred to these states in his work as "non-flow" states (e.g., Csikszentmihalyi, 1975; Csikszentmihalyi and LeFevre, 1989).

Csikszentmihalyi and Csikszentmihalyi (1988) created the Flow Questionnaire as a first attempt to operationalize flow (see Moneta, 2012). Participants are presented with first-hand accounts of what it feels like to be in flow, and then are asked a series of questions including "Have you ever felt similar experiences?" and "If yes, what activities where you engaged in when you had such experiences?" Thus, the Flow Questionnaire operationalizes flow as a discrete construct. Csikszentmihalyi and his colleagues have also used the "quadrant model" (**Figure 2**) to classify states of consciousness as either flow or non-flow states (i.e., anxiety, apathy, boredom/relaxation) (e.g., Csikszentmihalyi and LeFevre, 1989; Shernoff et al., 2003). This measurement method, too, operationalizes flow as a discrete construct.

Given that Csikszentmihalyi and his colleagues have conceptualized and operationalized flow as a discrete construct, it may be surprising to learn that a significant majority of the studies conducted within the past 5 years operationalized flow as a continuous construct (**Table 1**). How did this come to be? To address this question, it is necessary to appreciate the difficulty of capturing flow. Flow is described as occurring rarely in regular life (Csikszentmihalyi, 1975, 1990). The rarity with which flow is experienced presents a serious problem for the flow researcher, as statistical power is strongly dependent on having a large sample size. The difficulty of capturing flow is compounded in the psychological laboratory, where participants engage in what is typically an unfamiliar task in an inherently evaluative context. Both of these attributes – the unfamiliarity of the task and the evaluative nature of the context – are likely to work against the (already slim) likelihood of flow being experienced by a study participant, given that (1) flow appears more likely to be experienced by individuals who have developed considerable skill in the activity at hand (Jackson and Csikszentmihalyi, 1999; Rheinberg, 2008; Marin and Bhattacharya, 2013; Cohen and Bodner, 2019) and (2) performance anxiety is not conducive to flow (Csikszentmihalyi, 1975; Fullagar et al., 2013).

One strategy to deal with this "problem of low N" is to reformulate flow from a discrete state of consciousness to one experienced in varying degrees across the full spectrum of conscious experience. Using this approach, any state of consciousness can be classified along a flow continuum, with one end being very low flow and the other end being very high flow (e.g., Jackson and Marsh, 1996; Rheinberg et al., 2003). By doing this, all observations collected in a given study may be included in statistical analyses and contribute toward calculated effects. But reformulating flow in this manner alters the concept in a fundamental way. Flow is by definition an optimal experience,

and so designating all other experiences as variations in flow (low flow, moderate flow, etc.) diminishes the intelligibility of the construct. "Low flow" is a contradiction in terms, just as "mild rage" and "moderate ecstasy" are, given that level of intensity is built into the construct.

Besides the conceptual confusion that results from operationalizing flow as a construct applicable to the full range of conscious experience, there is a second reason to avoid operationalizing flow in this manner. When the concept of flow is extended to apply to the full range of experience, it has questionable discriminant validity over pre-existing constructs in surrounding fields. Within the field of intrinsic motivation, dozens of studies have examined a state-level construct called *task involvement* (e.g., Harackiewicz et al., 1987; Elliot and Harackiewicz, 1994, 1996; Tauer and Harackiewicz, 2004; Abuhamdeh and Csikszentmihalyi, 2012a), which represents the degree to which an individual concentrates on and becomes absorbed in an activity. Research on task involvement predates the first operationalizations of flow as a continuous construct, and appears to have been influenced by Csikszentmihalyi's work on optimal experience (Harackiewicz and Sansone, 1991). If flow is reformulated as a continuous construct, how do we know associated findings are not redundant with what has already been found with respect to task involvement? What is presented as a new contribution to the psychological literature may in fact be old news.

In reality it seems unlikely that there is a sharp boundary between flow and non-flow experiential states. Such thresholds appear to be exceedingly rare when it comes to states of consciousness, even extraordinary ones such as flow. Nevertheless, because flow is conceptualized as an "optimal" experience, it should be operationalized as such. Or else it shouldn't be called "flow."

Issue 2: Is Flow Inherently Enjoyable?

In the preface to *Beyond Boredom and Anxiety* (1975), Csikszentmihalyi described the purpose of his research:

"The goal was to focus on people who were having peak experiences, who were intrinsically motivated, and who were involved in play as well as real life activities, in order to find out whether I could detect similarities in their experiences, their motivation, and the situations that produce enjoyment." (p. xiii)

From this passage, and many others, it is clear that Csikszentmihalyi conceptualized flow as an *enjoyable* experience. Indeed, it was the enjoyable nature of flow, and the positive implications this enjoyment had for motivation, that positioned it as a vehicle for skill development and personal growth (i.e., greater "complexity") (Csikszentmihalyi and Rathunde, 1998). Csikszentmihalyi hasn't veered from this initial conception. In more recent work by Csikszentmihalyi and his colleagues, the enjoyable, "autotelic" (i.e., intrinsically rewarding) nature of flow has been consistently emphasized (e.g., Nakamura and Csikszentmihalyi, 2009; Nakamura et al., 2019).

Despite Csikszentmihalyi's conceptualization of flow as a form of enjoyment, it is quite common for flow researchers to exclude enjoyment (or "autotelic experience") from their operationalizations of flow, as shown in **Table 1**. Of the 42 reviewed studies, 17 of them did not include enjoyment (or autotelic experience or intrinsic motivation) in their operationalizations. How did this come to be? Why is flow being operationalized by some flow researchers without an enjoyment component? In reviewing the history of this issue I identified several likely sources (Abuhamdeh, in press).

Source #1: Martin Seligman

Beginning in his bestselling book *Authentic Happiness* (2002), Seligman (2011) began asserting that "it is the absence of emotion, of any kind of consciousness, that is at the heart of flow." (p. 111). Seligman (2011)'s reasoning for this is expressed in many places, including his modestly titled follow-up book *Flourish: A Visionary New Understanding of Happiness and Well-being* (2011), in which he wrote: "I believe that the concentrated attention that flow requires uses up all the cognitive and emotional resources that make up thought and feeling." (p. 11).

Judging by how often he has been cited, flow researchers have taken Seligman's views on flow very seriously. But his assertion that flow is devoid of emotion is in direct conflict with Csikszentmihalyi's conceptualization of flow as a form of enjoyment (given that enjoyment is an emotion). Furthermore, the notion that the intensive allocation of cognitive resources to a task prevents emotions from being experienced is at odds with contemporary emotion theory and research. Perhaps the most complete account of how emotions are elicited is provided by appraisal theories of emotion (Arnold, 1960; Lazarus, 1966; Scherer, 1984; Smith and Ellsworth, 1985; Frijda, 1986; Oatley and Johnson-Laird, 1987). Among appraisal theorists, there is consensus that appraisals do not always require conscious intervention (Ellsworth and Scherer, 2003; Moors, 2010). In fact it is generally presumed that appraisal processes usually occur automatically (Smith and Kirby, 2001; Moors, 2010). Appraisals must be fast and efficient given that changes in the environment can occur very quickly (Lazarus, 2001). Thus, like other automatic processes, they need not consume significant attentional resources.

Appraisal theorists also agree that with increasing practice there is greater automatization of appraisal processes (Moors et al., 2013). This has particular relevance for flow because flow appears to be more commonly experienced by individuals who are quite skilled in the activity they are engaged in (and thus have logged many hours of practice) (Csikszentmihalyi, 1975; Dietrich, 2004; Marin and Bhattacharya, 2013; Cohen and Bodner, 2019). Therefore, it seems especially likely that any appraisal processes that may occur during flow are mostly or fully automatic.

Source #2: A Failure to Differentiate Between *Experiencing* Emotions and One's *Awareness and Labeling* of These Emotions

One defining feature of flow is an absence of self-awareness. Flow researchers have sometimes assumed that this absence of self-awareness during flow prevents the experience of emotion during flow. For example, from a recent paper (Kyriazos et al., 2018): "Flow-ers seem to be almost beyond experiencing emotions,

probably due to the absence of self-awareness..." But self-awareness is not a precondition for the *experience* of emotions, only the *recognition* and *labeling* of them. This is why non-human mammals who lack a sense of self are nevertheless capable of experiencing emotions (Panksepp, 2005). Similarly, among humans, those younger than 7 months (and who therefore have not yet developed a sense of self) are nevertheless able to experience a wide range of emotions (Izard et al., 1995). The only emotions not in the repertoire of these children appear to be the so-called "self-conscious emotions" (e.g., pride, shame, and guilt), which young children first appear capable of experiencing between the ages of 2.5 and 3 years (Lewis, 2008). Indeed, even children who lack a cerebral cortex are capable of experiencing emotions (Merker, 2007).

Source #3: Csikszentmihalyi's Confusing Usage of the Word "Pleasure" in His Work

In his book Flow (1990), Csikszentmihalyi wrote, "None of these [flow] experiences may be particularly pleasurable at the time they are taking place, but afterward we think back on them and say, "That really was fun" and wish they would happen again." This statement may seem to imply that the experience of flow itself may not be particularly enjoyable. However, to properly interpret this passage it is necessary to understand Csikszentmihalyi's unusual usage of the word "pleasure" in his work, and the sharp distinction he draws between pleasure and enjoyment. Csikszentmihalyi (1990) considers pleasurable experiences to be those that satisfy biological needs, such as eating and sleeping (p. 45). According to Csikszentmihalyi, the experience of pleasure is derived from "restorative homeostatic experiences." Thus an artist who stayed up all night feverishly working on a painting, foregoing both food and rest, did not have a "pleasurable" experience according to Csikszentmihalyi's usage, because the behavior did not satisfy any biological needs (in fact it was in conflict with them). But this should not be misinterpreted as implying that the artist did not enjoy him/herself.

Issue 3: Should Flow Be Partly or Fully Operationalized Using Its Proposed Antecedents?

Csikszentmihalyi and his colleagues make a clear distinction between the conditions of flow and the experience of flow itself (**Figure 1**). Yet if we refer once again to **Table 1**, we see that a large number of studies ignored this distinction by operationalizing flow using *both* the experiential elements of the flow state and one or more of the conditions of flow. For example, in the Flow State Scale (Jackson and Marsh, 1996), some items measure the experiential elements of flow (e.g., "I had total concentration") whereas others measure the proposed conditions (e.g., "my goals were clearly defined"). The items are then usually averaged by researchers to yield a single "flow" score.

Given the strong distinction Csikszentmihalyi and his colleagues make between the conditions proposed to elicit flow and the state of flow itself, why is this distinction routinely ignored in empirical work? One explanation may be found in Csikszentmihalyi's earlier work. Though for the past several

years Csikszentmihalyi and his colleagues have drawn a sharp distinction, this was not always the case. In *Beyond Boredom and Anxiety* (1975), for example, Csikszentmihalyi himself grouped the conditions of flow with the experiential elements by including all of them under the heading "Elements of the flow experience" (p. 38). And this continued for several years. In *Flow* (1990), he included both the conditions of flow and the experiential elements under the general heading "The elements of enjoyment." (p. x). It wasn't until approximately 20 years ago that Csikszentmihalyi and his colleagues began consistently differentiating the conditions from the experience.

Additionally, it should be noted that Csikszentmihalyi and his colleagues themselves sometimes operationalized flow based solely on the ratio of challenges and skills (e.g., Massimini and Carli, 1988; Csikszentmihalyi and LeFevre, 1989; Stein et al., 1995; Shernoff et al., 2003; Asakawa, 2004). Indeed, before the current popularity of flow scales, this was the most common way to operationalize flow. This likely served to further reinforce the idea that flow and the conditions that elicit it are one and the same.

So how to proceed? It has been argued that the primary objective of any scientific endeavor is to provide causal explanations (e.g., Popper, 1957; Shadish et al., 2002). Thus the conceptual distinction Csikszentmihalyi and his colleagues make between the conditions of flow and the state itself is an important one. Indeed, much of what distinguished Csikszentmihalyi's initial work on flow from previous work on peak experiences was that he attempted to not only describe the experience, but to explain it by identifying the conditions which elicited it. This is why Csikszentmihalyi's work on flow is sometimes referred to as a "model" or "theory." Without distinguishing cause from effect, however, it is neither.

That the distinction should be consistently made is supported by empirical findings, too. "Flow" (as measured by the Flow Short Scale, Rheinberg et al., 2003) is not always optimized by a balance of challenges and skills, which suggests that inferring flow based on this condition is not a safe bet (Engeser and Rheinberg, 2008). Indeed, the relationship between challenge and enjoyment appears to be very unstable across both activity and person (Abuhamdeh and Csikszentmihalyi, 2009, 2012b). This variation helps account for why the variance in subjective experience explained by challenge-skill ratios across all daily activities tends to be low (Ellis et al., 1994).

As can be seen in **Table 1**, most of the commonly used flow scales conflate the conditions and the experience. One notable exception among them, however, is the 10-item Core Flow Scale (Martin and Jackson, 2008), used in one of the 42 studies. The aim of the scale, as described by the authors, is "to assess the central subjective (phenomenological) experience of flow." Because this scale does not conflate the conditions of flow with the experience of flow, it may be the best option among the current fleet of validated scales. However when using this scale, or any other which purports to measure the components of flow, it is advisable to allow the weighting of the components to vary freely rather than the usual custom of assuming they are equal and taking their average, since the relative contribution of each component

to the overall experience of flow in specific contexts is unknown (see Jackson and Marsh, 1996).

TWO REMAINING QUESTIONS

The preceding discussion raises two specific questions which deserve to be addressed here.

Question 1: If Flow Is to Be Operationalized as a Discrete Construct, Where Should the Boundary Between "Flow" and "Non-flow" Be Set?

This is clearly a difficult question to answer satisfactorily.[3] A sharp boundary or threshold is unlikely to exist. Individuals who describe their optimal experiences do not commonly report a sudden transition point between flow and non-flow. This therefore presents a dilemma for the flow researcher, as any delineation of a cutoff would necessarily involve a degree of arbitrariness. Nevertheless, to remain true to flow's conceptualization as a discrete state, a boundary must be set.

Previous attempts to distinguish flow from non-flow have varied considerably in approach. The most common approach has been to classify experience based on challenge–skill ratios (such as the quadrant model shown in **Figure 2**). However, this approach infers flow based solely on a single proposed condition (the balance of challenge and skill), which, as previously discussed, is not warranted. Furthermore, dividing experience in such a manner often results in 25% or more of all daily experiences being designated as "flow" experiences (e.g., Csikszentmihalyi and LeFevre, 1989; Hektner and Asakawa, 2000).

Rather than the researchers deciding which experiences qualify as flow experiences, an alternative strategy has been to have the participants decide for themselves. Indeed, this is how Csikszentmihalyi initially began measuring flow experiences (see Moneta, 2012). In the Flow Questionnaire (Csikszentmihalyi and Csikszentmihalyi, 1988) respondents are first provided with a description of a flow experience, and then are asked to indicate whether they have ever experienced flow. If so, various follow-up questions about these experiences are then asked. Similar measures which tap single flow experiences have since been created (e.g., Novak et al., 2003). These measures appear to come closest to operationalizing flow as it is conceptualized – as a discrete, optimal state of consciousness. Unfortunately, they are not commonly used. Out of the 42 studies listed in **Table 1**, only one used such a measure.

Kawabata and Evans (2016), noting the inability of most commonly used flow scales to differentiate flow experiences from non-flow experiences (e.g., the Flow State Scale, Jackson and Marsh, 1996; the Flow Short Scale, Rheinberg et al.,

2003), proposed a remedy. They first administered one of the more popular flow scales to participants (the Flow State Scale-2; Jackson and Eklund, 2002) immediately following physical activity of some sort (e.g., physical education class and training session). They then used latent class analysis to divide participants into four groups based on the participants flow scores. Kawabata and Evans noted that the participants in the two groups with the highest item-averages both had average scores greater than 3 (the midpoint of the 5-point scale), and on this basis they proposed that the participants in the two groups experienced flow. This constituted 54% of the sample. Though the sensibility of the criterion used in this case to delineate a cutoff appears dubious and resulted in a suspiciously high number of participants who were deemed to have experienced flow, the study represents the first serious attempt to rectify what is a major limitation of most flow scales.

Although no sharp boundary between "flow" and "non-flow" is likely to exist, this does not mean that a cutoff cannot be based on sensible criteria. This may seem contradictory, but such cut-offs are routinely designated for practical reasons in other fields, with success (for example in the medical sciences for high blood pressure, obesity, etc., as well as in clinical psychology for the assessment of psychological disorders). Taxonomic analytic techniques (Meehl, 1995; De Boeck et al., 2005; Ruscio et al., 2006) appear especially well-suited for identifying potential cut-off points. As one possibility, previous factor analyses based on data derived from flow scales indicate that two of the proposed components of flow – a lack of self-consciousness and a merging of action and awareness – load poorly on a higher-order "flow" factor (see Swann et al., 2018), even though these two features were commonly mentioned features of flow in Csikszentmihalyi's early interviews. One possible explanation for this is that these two features only become experientially salient at very high levels of involvement, which may have been underrepresented in the factor-analytic studies. If this is the case, the implied inflection point would offer a sound basis for a cut-off. More generally, taxonomic analytic techniques should help clarify whether flow represents a difference in quality of experience versus simply a difference in degree.

Question 2: What About "Sub-Optimal" Experiences? Does the Flow Model Have No Relevance for Them?

In this conceptual analysis I've argued that flow should be operationalized as Csikszentmihalyi conceptualized it: as an exceptional, "optimal" experience. But what about less intense, "non-flow" states of goal-directed engagement? Does the flow model have no relevance when it comes to these much more common states? Clearly it does. There is evidence that all three of the proposed antecedents of flow (clear goals, immediate feedback, and optimal challenges), in at least some situations, promote enjoyment (Harter, 1978; Reser and Scherl, 1988; Abuhamdeh and Csikszentmihalyi, 2012b; Pratt et al., 2016). But the fact that the conditions of flow have relevance for these states should not prompt researchers to automatically label these states as flow, as doing so obfuscates the meaning of flow.

---

[3]The difficulty this presents is one reason why, in my own empirical work on goal-directed engagement, despite my longstanding interest in flow, I've resisted operationalizing flow altogether, instead opting to measure experience in a more piecemeal fashion using lower-level constructs that can be meaningfully applied to the full range of conscious experience (e.g., Abuhamdeh and Csikszentmihalyi, 2012a; Abuhamdeh et al., 2015).

It is interesting to note that Csikszentmihalyi himself recognized the relevance of the flow model for less intense states than flow. He introduced the concept of "micro-flow" to help account for such experiences (Csikszentmihalyi, 1975). However, the introduction of another discrete construct (with all the accompanying operational dilemmas) to account for less intense states at this point seems unnecessary. Two pre-existing constructs in the motivation literature, mostly ignored by flow researchers, appear very capable of capturing such states. Crucially, both of them are continuous constructs that can be applied meaningfully to the full range of conscious experience.

Construct #1: Task Involvement

Flow has been described as being composed of cognitive, emotional, and motivational components (e.g., Delle Fave and Massimini, 2005). In terms of its cognitive aspect, the defining feature of flow is intense attentional focus on the task at hand (Nakamura and Csikszentmihalyi, 2002). It is this deep attentional involvement that appears to underlie several of the other characteristics of flow including the merging of action and awareness and the absence of self-consciousness (Dietrich, 2004; Csikszentmihalyi et al., 2005; Kawabata and Mallett, 2011).

Task involvement, as previously described, represents the degree to which an individual concentrates on and becomes absorbed in an activity (Elliot and Harackiewicz, 1994). Operationalizations usually include items that measure both absorption and concentration. The task involvement construct nicely captures the central cognitive feature of flow. In contrast to flow, however, task involvement is a purely cognitive phenomenon representing the degree of attentional involvement in an activity; it is not inherently enjoyable and motivating in concept, though it often predicts both (Abuhamdeh and Csikszentmihalyi, 2012a).

Construct #2: Intrinsic Motivation

Because of the enjoyable nature of flow, it is "autotelic," meaning it motivates the person who experiences it to continue doing what he/she is doing. The meaning of autotelic and intrinsic motivation are synonymous. Intrinsic motivation, as conceptualized and operationalized within the motivation literature, captures both the emotional and (therefore) motivational properties of flow, yet, in contrast, is applicable to the full range of conscious experience.

The standard way to measure intrinsic motivation is by asking participants how *enjoyable* and *interesting* the activity they are (or were) engaged in is. The measurement of both enjoyment and interest is important, because interest appears to be a positive emotion distinct from enjoyment (Tomkins, 1962; Izard, 1977; Panksepp, 1998; Silvia, 2008). This view is backed by empirical findings which indicate that interest and enjoyment, in at least some contexts, have different antecedents, as well as different trajectories in response to performance feedback (Reeve, 1989; Egloff et al., 2003).

In sum, the conditions of flow have implications for a much wider array of states than just flow. The constructs *task involvement* and *intrinsic motivation* appear particularly well-suited for capturing these states. The incorporation of these constructs into empirical investigations of goal-directed engagement has the added benefit of allowing the associated research findings to be more easily assimilated into the surrounding motivation literature.

SUMMARY AND CONCLUSION

Almost 50 years ago, Csikszentmihalyi (1975) began a program of research with the aim of understanding the common experiential characteristics of so-called "optimal experiences," as well as the conditions which promote these experiences. To this end, he asked hundreds of rock climbers, chess players, artists, etc. to describe what their best moments felt like. Based on this research, Csikszentmihalyi developed the concept of "flow."

Since that time, hundreds of empirical studies have been conducted in an attempt to further understand flow. Yet if we survey the ways in which flow has been operationalized in these studies, we are forced to reckon with an unsettling fact: a consensual operationalization of flow has yet to be established. Across studies, operationalizations vary considerably, so that the meaning of flow from one study to the next often changes drastically.

In this conceptual analysis, I've highlighted three key inconsistencies found in flow operationalizations: (1) inconsistences in operationalizing flow as a discrete versus continuous construct, (2) inconsistencies in operationalizing flow as inherently enjoyable (i.e., autotelic) or not, and (3) inconsistencies in operationalizing flow as dependent on versus distinct from the task characteristics proposed to elicit it (i.e., the conditions/antecedents). I've argued that these inconsistencies are born out of conceptual misunderstandings, as well as the methodological difficulties inherent in operationalizing optimal experience.

The lack of a standard operationalization of flow does not bode well for the field. It is only by adopting a standard operationalization that questions about the nature of flow (e.g., is the distortion of time a consistent component of optimal experience?) as well as flow's relation to other constructs (e.g., what is the relationship between flow and performance?) can be addressed. It is only by the consistent application of a standard operationalization that a period of "normal science" (Kuhn, 1962) may ensue.[4]

Given that a standard operationalization of flow is needed, whose conceptualization of flow should it be based on? A tacit assumption made throughout this paper is that Csikszentmihalyi's conceptualization of flow is the only valid conceptualization. The reasoning for this is as follows: Unlike

---

[4]Swann et al. (2018) recently assessed the current state of flow research in sport and exercise psychology, using Kuhn's (1962) model of scientific development as a guide. Their provocative thesis was that flow research, following a long period of "normal science," is now approaching a "crisis point." However in Kuhn's (1962) scheme, "normal science" represents the practice of working within a firmly established research paradigm, characterized by, among other things, uniform conceptualizations and standard operationalizations. As shown in the current paper, flow research cannot be characterized as such. At least from a methodological standpoint, the current state of the field seems to have more in common with the preceding stage in Kuhn's (1962) scheme – what he referred to as the "pre-paradigm" stage. Indeed, in his famous book, Kuhn (1962) himself seemed to imply that all of the social sciences are pre-paradigmatic (p. 161).

most psychological constructs, which are generic in their nature (e.g., euphoria, misery, anxiety, etc.), we put "flow" in quotes (or italicize it, or write it with a capital F) because it is a proper noun, a term coined by a specific psychologist to represent his particular conceptualization of optimal experience. In other words, the term flow comes with Csikszentmihalyi's conceptualization "pre-installed." His conceptualization is therefore the default conceptualization, and this is true regardless of its merits.[5]

---

[5] By the same token, if I formulated a conceptualization of ecstatic love which I called Glow, and other researchers, inspired by my work on Glow, wished to investigate it, they would need to operationalize Glow as I conceptualized it (as a state of ecstatic love) in order to make any claims about Glow based on their subsequent findings.

Of course, once this conceptualization is operationalized in a valid and consistent manner, and systematically tested and evaluated, it may turn out that Csikszentmihalyi's conceptualization of optimal experience should be modified or updated in one or more ways. In this case, a revised conceptualization would be warranted. This would be a positive development, a sign of progress.

AUTHOR CONTRIBUTIONS

REFERENCES

Abuhamdeh, S., and Csikszentmihalyi, M. (2009). Intrinsic and extrinsic motivational orientations in the competitive context: an examination of person–situation interactions. *J. Pers.* 77, 1615–1635. doi: 10.1111/j.1467-6494.2009.00594.x

Abuhamdeh, S., and Csikszentmihalyi, M. (2012a). Attentional involvement and intrinsic motivation. *Motiv. Emot.* 36, 257–267. doi: 10.1007/s11031-011-9252-7

Abuhamdeh, S., and Csikszentmihalyi, M. (2012b). The importance of challenge for the enjoyment of intrinsically motivated, goal-directed activities. *Personal. Soc. Psychol. Bull.* 38, 317–330. doi: 10.1177/0146167211427147

Abuhamdeh, S., Csikszentmihalyi, M., and Jalal, B. (2015). Enjoying the possibility of defeat: outcome uncertainty, suspense, and intrinsic motivation. *Motiv. Emot.* 39, 1–10. doi: 10.1007/s11031-014-9425-2

Abuhamdeh, S. A. (in press). "On the Relationship Between Flow and Enjoyment," in *Advances in Flow Research*, 2nd Edn, eds S. Engeser and C. Peifer (New York, NY: Springer).

Arnold, M. B. (1960). *Emotion and Personality*. New York, NY: Columbia University Press.

Asakawa, K. (2004). Flow experience and autotelic personality in japanese college students: how do they experience challenges in daily life? *J. Happ. Stud.* 5, 123–154. doi: 10.1023/B:JOHS.0000035915.97836.89

Bakker, A. B., Oerlemans, W., Demerouti, E., Slot, B. B., and Ali, D. K. (2011). Flow and performance: a study among talented Dutch soccer players. *Psychol. Sport Exerc.* 12, 442–450. doi: 10.1016/j.psychsport.2011.02.003

Barros, M. F. S., Araújo-Moreira, F. M., Trevelin, L. C., and Radel, R. (2018). Flow experience and the mobilization of attentional resources. *Cogn.Affect. Behav. Neurosci.* 18, 810–823. doi: 10.3758/s13415-018-0606-4

Baumann, N., Lürig, C., and Engeser, S. (2016). Flow and enjoyment beyond skill-demand balance: the role of game pacing curves and personality. *Motiv. and Emot.* 40, 507–519. doi: 10.1007/s11031-016-9549-7

Beltrán, H. C., Reigal, R. E., Uribe, S. F., Reyes, F. V., and Chirosa Ríos, L. J. (2018). Motivación autodeterminada y estado de flow en un programa extraescolar de Small Sided Games = Self-determined motivation and state of flow in an extracurricular program of Small Sided Games. *Anal. Psicol.* 34, 391–397.

Bonaiuto, M., Mao, Y., Roberts, S., Psalti, A., Ariccio, S., Cancellieri, U. G., et al. (2016). Optimal experience and personal growth: flow and the consolidation of place identity. *Front. Psychol.* 7:1654. doi: 10.3389/fpsyg.2016.01654

Borovay, L. A., Shore, B. M., Caccese, C., Yang, E., and Hua, O. (2019). Flow, achievement level, and inquiry-based learning. *J. Adv. Acade.* 30, 74–106. doi: 10.1177/1932202X18809659

Brailovskaia, J., Rohmann, E., Bierhoff, H.-W., and Margraf, J. (2018). The brave blue world: facebook flow and facebook addiction disorder (FAD). *PLoS One* 13:e0201484. doi: 10.1371/journal.pone.0201484

Bricteux, C., Navarro, J., Ceja, L., and Fuerst, G. (2017). Interest as a moderator in the relationship between challenge/skills balance and flow at work: an analysis

at within-individual level. *J. Happ. Stud.* 18, 861–880. doi: 10.1007/s10902-016-9755-8

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., and Pidruzny, J. N. (2009). The development of the game engagement questionnaire: a measure of engagement in video game-playing. *J. Exp. Soc. Psychol.* 45, 624–634. doi: 10.1016/j.jesp.2009.02.016

Brom, C., Děchtěrenko, F., Frollová, N., Stárková, T., Bromová, E., and D'Mello, S. K. (2017). Enjoyment or involvement? Affective-motivational mediation during learning from a complex computerized simulation. *Comput. Educ.* 114, 236–254. doi: 10.1016/j.compedu.2017.07.001

Chen, L.-X., and Sun, C.-T. (2016). Self-regulation influence on game play flow state. *Comput. Hum Behav.* 54, 341–350. doi: 10.1016/j.chb.2015.08.020

Cho, M. (2018). Task complexity and modality: exploring learners' experience from the perspective of flow. *Modern Lang. J.* 102, 162–180. doi: 10.1111/modl.12460

Chou, T.-J., and Ting, C.-C. (2003). The role of flow experience in cyber-game addiction. *Cyberpsychol. Behav.* 6, 663–675. doi: 10.1089/109493103322725469

Coffey, J. K., Wray-Lake, L., Mashek, D., and Branand, B. (2016). A multi-study examination of well-being theory in college and community samples. *J. Happ. Stud.* 17, 187–211. doi: 10.1007/s10902-014-9590-8

Cohen, S., and Bodner, E. (2019). The relationship between flow and music performance anxiety amongst professional classical orchestral musicians. *Psychol. Music* 47, 420–435. doi: 10.1177/0305735618754689

Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*. San Francisco, CA: Jossey-Bass Publishers.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York, NY: Harper & Row.

Csikszentmihalyi, M., Abuhamdeh, S., and Nakamura, J. (2005). *Flow. In Handbook of Competence and Motivation*. Guilford Publications, 598–608.

Csikszentmihalyi, M., and Csikszentmihalyi, I. S. (eds) (1988). *Optimal Experience: Psychological Studies of Flow in Consciousness*. New York, NY: Cambridge University Press.

Csikszentmihalyi, M., Larson, R., and Prescott, S. (1977). The ecology of adolescent activity and experience. *J. Youth Adoles.* 6, 281–294. doi: 10.1007/BF02138940

Csikszentmihalyi, M., and LeFevre, J. (1989). Optimal experience in work and leisure. *J. Pers. Soc. Psychol.* 56, 815–822. doi: 10.1037//0022-3514.56.5.815

Csikszentmihalyi, M., and Rathunde, K. (1998). "The development of the person: an experiential perspective on the ontogenesis of psychological complexity," in *Handbook of Child Psychology: Theoretical Models of Human Development*, 5th Edn, eds W. Damon, and R. M. Lerner, (Hoboken, NJ: John Wiley & Sons Inc), 635–684.

Csikszentmihalyi, M., Rathunde, K. R., Whalen, S., and Wong, M. (1993). *Talented Teenagers: The Roots of Success and Failure*. New York, NY: Cambridge University Press.

De Boeck, P., Wilson, M., and Acton, G. S. (2005). A conceptual and psychometric framework for distinguishing categories and dimensions. *Psychol. Rev.* 112, 129–158. doi: 10.1037/0033-295X.112.1.129

Delle Fave, A., and Massimini, F. (2005). The investigation of optimal experience and apathy: developmental and psychosocial implications. *Eur. Psychol.* 10, 264–274. doi: 10.1027/1016-9040.10.4.264

Demerouti, E. (2006). Job characteristics, flow and performance: the moderating role of conscientiousness. *J. Occup. Health Psychol.* 11, 266–280. doi: 10.1037/1076-8998.11.3.266

Dietrich, A. (2004). Neurocognitive mechanisms underlying the experience of flow. *Conscious Cogn.* 13, 746–761. doi: 10.1016/j.concog.2004.07.002

Dixon, M. J., Gutierrez, J., Stange, M., Larche, C. J., Graydon, C., Vintan, S., et al. (2019). Mindfulness problems and depression symptoms in everyday life predict dark flow during slots play: implications for gambling as a form of escape. *Psychol. Addict. Behav.* 33, 81–90. doi: 10.1037/adb0000435

Egloff, B., Schmukle, S. C., Burns, L. R., Kohlmann, C.-W., and Hock, M. (2003). Facets of dynamic positive affect: differentiating joy, interest, and activation in the positive and negative affect schedule (PANAS). *J. Pers. Soc. Psychol.* 85, 528–540. doi: 10.1037/0022-3514.85.3.528

Elliot, A. J., and Harackiewicz, J. M. (1994). Goal setting, achievement orientation, and intrinsic motivation: a mediational analysis. *J. Pers. Soc. Psychol.* 66, 968–980. doi: 10.1037/0022-3514.66.5.968

Elliot, A. J., and Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: a mediational analysis. *J. Pers. Soc. Psychol.* 70, 461–475. doi: 10.1037/0022-3514.70.3.461

Ellis, G. D., Voelkl, J. E., and Morris, C. (1994). Measurement and analysis issues with explanation of variance in daily experience using the flow model. *J. Leisure Res.* 26, 337–356. doi: 10.1080/00222216.1994.11969966

Ellsworth, P. C., and Scherer, K. R. (2003). "Appraisal processes in emotion," in *In Handbook of Affective Sciences*, eds R. J. Davidson, K. R. Sherer, and H. H. Goldsmith, (Oxfor: Oxford University Press), 572–595.

Engeser, S., and Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motiv. Emot.* 32, 158–172. doi: 10.1007/s11031-008-9102-4

Forkosh, J., and Drake, J. E. (2017). Coloring versus drawing: effects of cognitive demand on mood repair, flow, and enjoyment. *Art Ther.* 34, 75–82. doi: 10.1080/07421656.2017.1327272

Frijda, N. H. (1986). *The Emotions*. Paris: Editions de la Maison des Sciences de l'Homme.

Fullagar, C. J., and Kelloway, E. K. (2009). Flow at work: an experience sampling approach. *J. Occup. Organ. Psychol.* 82, 595–615. doi: 10.1348/096317908X357903

Fullagar, C. J., Knight, P. A., and Sovern, H. S. (2013). Challenge/skill balance. Flow, and Performance Anxiety. *Appl. Psychol.* 62, 236–259. doi: 10.1111/j.1464-0597.2012.00494.x

Harackiewicz, J. M., Abrahams, S., and Wageman, R. (1987). Performance evaluation and intrinsic motivation: the effects of evaluative focus, rewards, and achievement orientation. *J. Pers. Soc. Psychol.* 53, 1015–1023. doi: 10.1037/0022-3514.53.6.1015

Harackiewicz, J. M., and Sansone, C. (1991). "Goals and intrinsic motivation: you can get there from here," in *Advances in Motivation and Achievement: Goals and Self Regulatory Processes*, Vol. 7, eds M. L. Maehr, and P. R. Pintrich, (Greenwich, CT: JAI Press), 21–49.

Harmat, L., de Manzano, Ö, Theorell, T., Högman, L., Fischer, H., and Ullén, F. (2015). Physiological correlates of the flow experience during computer game playing. *Int. J. Psychophysiol.* 97, 1–7. doi: 10.1016/j.ijpsycho.2015.05.001

Harris, D. J., Vine, S. J., and Wilson, M. R. (2017a). Flow and quiet eye: the role of attentional control in flow experience. *Cogn. Process.* 18, 343–347. doi: 10.1007/s10339-017-0794-9

Harris, D. J., Vine, S. J., and Wilson, M. R. (2017b). Is flow really effortless? The complex role of effortful attention. *Sport Exerc.Perform. Psychol.* 6, 103–114. doi: 10.1037/spy0000083

Harter, S. (1978). Pleasure derived from challenge and the effects of receiving grades on children's difficulty level choices. *Child Dev.* 49, 788–799. doi: 10.2307/1128249

Hektner, J. M., and Asakawa, K. (2000). "Learning to like challenges," in *Becoming Adult: How teenagers Prepare for the World of Work*, eds M. Csikszentmihalyi, and B. Schneider, (New York, NY: Basic books), 95–112.

Hektner, J. M., Schmidt, J. A., and Csikszentmihalyi, M. (2007). *Experience Sampling Method: Measuring the Quality of Everyday Life*. Thousand Oaks, CA: Sage Publications, Inc.

Hermann, J. M., and Vollmeyer, R. (2016). 'Girls should cook, rather than kick!'—Female soccer players under stereotype threat. *Psychol. Sport Exerc.* 26, 94–101. doi: 10.1016/j.psychsport.2016.06.010

Hou, H.-T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: a video-based process exploration. *Comput. Hum. Behav.* 48, 424–435. doi: 10.1016/j.chb.2015.02.010

Huskey, R., Wilcox, S., and Weber, R. (2018). Network neuroscience reveals distinct neuromarkers of flow during media use. *J. Commun.* 68, 872–895. doi: 10.1093/joc/jqy043

Ilies, R., Wagner, D., Wilson, K., Ceja, L., Johnson, M., DeRue, S., et al. (2017). Flow at work and basic psychological needs: effects on well-being. *Appl. Psychol.* 66, 3–24. doi: 10.1111/apps.12075

Izard, C. E. (1977). *Human Emotions*. Berlin: Springer.

Izard, C. E., Fantauzzo, C. A., Castle, J. M., Haynes, O. M., Rayias, M. F., and Putnam, P. H. (1995). The ontogeny and significance of infants' facial expressions in the first 9 months of life. *Dev. Psychol.* 31, 997–1013. doi: 10.1037/0012-1649.31.6.997

Jackson, S. A., and Csikszentmihalyi, M. (1999). *Flow in Sports: The Keys to Optimal Experiences and Performances*. Champaign, IL: Human Kinetics Books.

Jackson, S. A., and Eklund, R. C. (2002). Assessing flow in physical activity: the flow state scale-2 and dispositional flow scale-2. *J. Sport Exerc. Psychol.* 24, 133–150. doi: 10.1123/jsep.24.2.133

Jackson, S. A., and Marsh, H. (1996). Development and validation of a scale to measure optimal experience: the flow state scale. *J. Sport Exerc. Psychol.* 18, 17–35. doi: 10.1123/jsep.18.1.17

Jackson, S. A., and Roberts, G. C. (1992). Positive performance states of athletes: toward a conceptual understanding of peak performance. *Sport Psychol.* 6, 156–171. doi: 10.1123/tsp.6.2.156

Jackson, S. A., Thomas, P. R., Marsh, H. W., and Smethurst, C. J. (2001). Relationships between flow, self-concept, psychological skills, and performance. *J. Appl. Sport Psychol.* 13, 129–153. doi: 10.1080/104132001753149865

Joo, Y. J., Oh, E., and Kim, S. M. (2015). Motivation, instructional design, flow, and academic achievement at a Korean online university: a structural equation modeling study. *J. Comput. Higher Educ.* 27, 28–46. doi: 10.1007/s12528-015-9090-9

Kawabata, M., and Evans, R. (2016). How to classify who experienced flow from who did not based on the flow state scale-2 scores: a pilot study of latent class factor analysis. *Sport Psychol.* 30, 267–275. doi: 10.1123/tsp.2014-0053

Kawabata, M., and Mallett, C. (2011). Flow experience in physical activity: examination of the internal structure of flow from a process-related perspective. *Mot. Emot.* 35, 393–402. doi: 10.1007/s11031-011-9221-1

Kaye, L. K., Monk, R. L., Wall, H. J., Hamlin, I., and Qureshi, A. W. (2018). The effect of flow and context on in-vivo positive mood in digital gaming. *Int. J. Hum. Comput. Stud.* 110, 45–52. doi: 10.1016/j.ijhcs.2017.10.005

Kennedy, P., Miele, D. B., and Metcalfe, J. (2014). The cognitive antecedents and motivational consequences of the feeling of being in the zone. *Conscious. Cogn.* 30, 48–61. doi: 10.1016/j.concog.2014.07.007

Kiili, K. (2006). Evaluations of an experiential gaming model. *Hum. Technol.* 2, 187–201. doi: 10.17011/ht/urn.2006518

Kocjan, G. Z., and Avsec, A. (2017). Bringing the psychology of situations into flow research: personality and situation characteristics as predictors of flow. *Psihol. Teme* 26, 195–210. doi: 10.31820/pt.26.1.9

Kuhn, T. S. (1962). *The structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kulkarni, A., Anderson, W., Sanders, M. A., Newbold, J., and Martin, L. L. (2016). Manipulated flow reduces downstream defensiveness. *J. Posit. Psychol.* 11, 26–36. doi: 10.1080/17439760.2015.1015157

Kwak, K., Choi, S. K., and Lee, B. (2014). SNS flow, SNS self-disclosure and post hoc interpersonal relations change: focused on korean facebook user. *Comput. Hum. Behav.* 31, 294–304. doi: 10.1016/j.chb.2013.10.046

Kyriazos, T. A., Stalikas, A., Prassa, K., Galanakis, M., Flora, K., and Chatzilia, V. (2018). The flow short scale (FSS) dimensionality and What MIMIC shows on heterogeneity and invariance. *Psychology* 09:1357. doi: 10.4236/psych.2018.96083

Laski, M. (1961). *Ecstasy: A Study of Some Secular and Religious Experiences*. Bloomington, IN: Indiana University Press.

Lavoie, R. V., and Main, K. J. (2019). When losing money and time feels good: the paradoxical role of flow in gambling. *J. Gambl. Issues* 41, 53–72.

Lazarus, R. S. (1966). *Psychological Stress and the Coping Process.* New York, NY: McGraw-Hill.

Lazarus, R. S. (2001). "relational meaning and discrete emotions," in *Series in Affective Science. Appraisal Processes in Emotion: Theory, Methods, Research.* eds S.R. Klaus, S. Angela, and J. Tom, (New York, NY: Oxford University Press), 37–67.

Lewis, M. (2008). "Self-conscious emotions: embarrassment, pride, shame, and guilt," in *Handbook of emotions*, 3rd Edn. eds M. Lewis, and J. M. Haviland-Jones (New York, NY: The Guilford Press), 742–756.

Lin, C.-W., Wang, K.-Y., Chang, S.-H., and Lin, J.-A. (2019). Investigating the development of brand loyalty in brand communities from a positive psychology perspective. *J. Bus Res.* 99, 446–455. doi: 10.1016/j.jbusres.2017.08.033

Mao, Y., Roberts, S., Pagliaro, S., Csikszentmihalyi, M., and Bonaiuto, M. (2016). Optimal experience and optimal identity: a multinational study of the associations between flow and social identity. *Front Psychol.* 7:67. doi: 10.3389/fpsyg.2016.00067

Marin, M. M., and Bhattacharya, J. (2013). Getting into the musical zone: trait emotional intelligence and amount of practice predict flow in pianists. *Front. Psychol.* 4:853. doi: 10.3389/fpsyg.2013.00853

Marston, H. R., Kroll, M., Fink, D., and Gschwind, Y. J. (2016). Flow experience of older adults using the iStoppFalls exergame. *Games Cult.*11, 201–222. doi: 10.1177/1555412015605219

Martin, A. J., and Jackson, S. A. (2008). Brief approaches to assessing task absorption and enhanced subjective experience: examining 'short' and 'core' flow in diverse performance domains. *Motiv. Emot.* 32, 141–157. doi: 10.1007/s11031-008-9094-0

Maslow, A. H. (1964). *Religions, Values, and Peak-Experiences.* Ohio State University Press.

Massimini, F., and Carli, M. (1988). "The Systematic Assessment of Flow in Daily Experience," in *Optimal experience: Psychological studies of flow in consciousness.* eds M. Csikszentmihalyi, and I. S. Csikszentmihalyi, (New York, NY: Cambridge University Press), 266–287. doi: 10.1017/cbo9780511621956.016

Matthews, N. L. (2015). Too good to care: the effect of skill on hostility and aggression following violent video game play. *Comput. Hum. Behav.* 48, 219–225. doi: 10.1016/j.chb.2015.01.059

Meehl, P. E. (1995). Bootstraps taxometrics: solving the classification problem in psychopathology. *Am. Psychol.* 50, 266–275. doi: 10.1037/0003-066X.50.4.266

Merker, B. (2007). Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav. Brain Sci.* 30, 63–81. discussion 81-134,doi: 10.1017/S0140525X07000891

Moneta, G. B. (2012). "On the measurement and conceptualization of flow," in *Advances in Flow Research*, ed. S. Engeser, (New York, NY: Springer).

Moors, A. (2010). Automatic constructive appraisal as a candidate cause of emotion. *Emot. Rev.* 2, 139–156. doi: 10.1177/1754073909351755

Moors, A., Ellsworth, P. C., Scherer, K. R., and Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emot. Rev.* 5, 119–124. doi: 10.1177/1754073912468165

Nakamura, J., and Csikszentmihalyi, M. (2002). "The concept of flow," in *Handbook of Positive Psychology*, eds C. R. Snyder and S. J. Lopez (Oxford: Oxford University Press), 89–105.

Nakamura, J., and Csikszentmihalyi, M. (2009). "Flow theory and research," in *Oxford Library of Psychology. Oxford handbook of positive psychology*, 2nd Edn. eds S. J. Lopez and C. R. Snyder (New York, NY: Oxford University Press), 195–206.

Nakamura, J., Tse, D. C. K., and Shankland, S. (2019). "Flow," in *The Oxford Handbook of Human Motivation*, ed. R. M. Ryan, (Oxford: Oxford University Press), 567.

Nielsen, K., and Cleal, B. (2010). Predicting flow at work: investigating the activities and job characteristics that predict flow states at work. *J. Occup. Health Psychol.* 15, 180–190. doi: 10.1037/a0018893

Novak, T. P., Hoffman, D. L., and Duhachek, A. (2003). The influence of goal-directed and experiential activities on online flow experiences. *J. Consum. Psychol.* 13, 3–16. doi: 10.1207/s15327663jcp13-1%262_01

Novak, T. P., Hoffman, D. L., and Yung, Y.-F. (2000). Measuring the customer experience in online environments: a structural modeling approach. *Mark. Sci.* 19, 22–42. doi: 10.1287/mksc.19.1.22.15184

Oatley, K., and Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cogni. Emot.* 1, 29–50. doi: 10.1080/02699938708408362

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions.* Oxford: Oxford University Press.

Panksepp, J. (2005). Affective consciousness: core emotional feelings in animals and humans. *Conscious. Cogn.* 14, 30–80. doi: 10.1016/j.concog.2004.10.004

Poels, K., de Kort, Y. A. W., and IJsselsteijn, W. A. (2007). *D3.3: Game Experience Questionnaire: Development of a Self-Report Measure to Assess the Psychological Impact of Digital Games.* Avaliable at: https://research.tue.nl/en/publications/d33-game-experience-questionnaire-development-of-a-self-report-me

Popper, K. (1957). The aim of science. *Ratio* 1, 24–35.

Pratt, J. A., Chen, L., and Cole, C. (2016). The influence of goal clarity, curiosity, and enjoyment on intention to code. *Behav. IT* 35, 1091–1101. doi: 10.1080/0144929x.2016.1171399

Rankin, K., Walsh, L. C., and Sweeny, K. (2019). A better distraction: exploring the benefits of flow during uncertain waiting periods. *Emotion* 19, 818–828. doi: 10.1037/emo0000479

Reeve, J. (1989). The interest-enjoyment distinction in intrinsic motivation. *Motiv. Emot.* 13, 83–103. doi: 10.1007/BF00992956

Reser, J. P., and Scherl, L. M. (1988). Clear and unambiguous feedback: a transactional and motivational analysis of environmental challenge and self-encounter. *J. Environ. Psychol.* 8, 269–286. doi: 10.1016/S0272-4944(88)80034-3

Rheinberg, F. (2008). "Intrinsic motivation and flow-experience," in *Motivation and action*, eds H. Heckhausen, and J. Heckhausen, (Cambridge: Cambridge University Press), 323–348. doi: 10.1017/cbo9780511499821.014

Rheinberg, F., Vollmeyer, R., and Engeser, S. (2003). "The Assessment of Flow Experience," in *Diagnosis of Motivation and Self-Concept*, eds J. Stiensmeier-Pelster, and F. Rheinberg, (Göttingen: Hogrefe).

Rivkin, W., Diestel, S., and Schmidt, K.-H. (2018). Which daily experiences can foster well-being at work? A diary study on the interplay between flow experiences, affective commitment, and self-control demands. *J. Occup. Health Psychol.* 23, 99–111. doi: 10.1037/ocp0000039

Rodríguez-Sánchez, A., Salanova, M., Cifre, E., and Schaufeli, W. B. (2011). When good is good: a virtuous circle of self-efficacy and flow at work among teachers. *Rev. e Psicol. Soci.* 26, 427–441. doi: 10.1174/021347411797361257

Rodríguez-Ardura, I., and Meseguer-Artola, A. (2017). Flow in e−learning: What drives it and why it matters. *Br. J. Educ,. Technol.* 48, 899–915. doi: 10.1111/bjet.12480

Rudd, M. D. (2000). The suicidal mode: a cognitive-behavioral model of suicidality. *Suicide LifeThreat. Behav.* 30, 18–33. doi: 10.1111/j.1943-278X.2000.tb01062.x

Ruscio, J. I. P., Haslam, N., and Ruscio, A. (2006). *Introduction to the Taxometric Method: A Practical Guide.* Mahwah, NJ: Lawrence Erlbaum & Associates.

Sather, T. W., Howe, T., Nelson, N. W., and Lagerwey, M. (2017). Optimizing the experience of flow for adults with aphasia: a focus on environmental factors. *Top. Lang. Disord.* 37, 25–37.

Schattke, K., Brandstätter, V., Taylor, G., and Kehr, H. M. (2014). Flow on the rocks: motive-incentive congruence enhances flow in rock climbing. *Int. J. Sport Psychol.* 45, 603–620.

Scherer, K. R. (1984). Emotion as a multicomponent process: a model and some cross-cultural data. *Rev. Personal. Soc. Psychol.* 5, 37–63.

Schüler, J. (2007). Arousal of flow experience in a learning setting and its effects on exam performance and affect. - *Z. Padagog. Psychol.* 21, 217–227. doi: 10.1024/1010-0652.21.3.217

Seger, J., and Potts, R. (2012). Personality correlates of psychological flow states in videogame play. *Curr. Psychol.*31, 103–121. doi: 10.1007/s12144-012-9134-5

Seligman, M. E. P. (2002). *Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment.* New York, NY: Free Press.

Seligman, M. E. P. (2011). *Flourish: A Visionary New Understanding of Happiness and Well-Being.* New York, NY: Free Press.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston, MA: Houghton, Mifflin and Company.

Shernoff, D. J., Csikszentmihalyi, M., Shneider, B., and Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *Sch. Psychol. Q.* 18, 158–176. doi: 10.1521/scpq.18.2.158.21860

Silvia, P. J. (2008). Interest—The curious emotion. *Curr. Dir. Psychol. Sci.* 17, 57–60. doi: 10.1111/j.1467-8721.2008.00548.x

Smith, C. A., and Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *J. Pers. Soc. Psychol.* 48, 813–838. doi: 10.1037/0022-3514.48.4.813

Smith, C. A., and Kirby, L. D. (2001). "Affect and cognitive appraisal processes," in *Handbook of Affect and Social Cognition*, ed. P. Forgas, (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 75–92.

Smith, L. J., Gradisar, M., King, D. L., and Short, M. (2017). Intrinsic and extrinsic predictors of video-gaming behaviour and adolescent bedtimes: the relationship between flow states, self-perceived risk-taking, device accessibility, parental regulation of media and bedtime. *Sleep Med* 30, 64–70. doi: 10.1016/j.sleep.2016.01.009

Soutter, A. R. B., and Hitchens, M. (2016). The relationship between character identification and flow state within video games. *Comput. Hum. Behav.* 55(Part B), 1030–1038. doi: 10.1016/j.chb.2015.11.012

Stavrou, N. A., Jackson, S. A., Zervas, Y., and Karteroliotis, K. (2007). Flow experience and athletes' performance with reference to the orthogonal model of flow. *Sport Psychol.* 21, 438–457. doi: 10.1123/tsp.21.4.438

Stein, G. L., Kimiecik, J. C., Daniels, J., and Jackson, S. A. (1995). Psychological antecedents of flow in recreational sport. *Personal. Soc. Psychol. Bull.* 21, 125–135. doi: 10.1177/0146167295212003

Sun, J. C.-Y., Kuo, C.-Y., Hou, H.-T., and Lin, Y.-Y. (2017). Exploring learners' sequential behavioral patterns, flow experience, and learning performance in an anti-phishing educational game. *J. Educ. Technol. Soc.* 20, 45–60.

Swann, C., Piggott, D., Schweickle, M., and Vella, S. A. (2018). A review of scientific progress in flow in sport and exercise: normal science. Crisis, and a Progressive Shift. *J. Appl. Sport Psychol.* 30, 249–271. doi: 10.1080/10413200.2018.1443525

Tauer, J. M., and Harackiewicz, J. M. (2004). The effects of cooperation and competition on intrinsic motivation and performance. *J. Pers. Soc. Psychol.* 86, 849–861. doi: 10.1037/0022-3514.86.6.849

Tomkins, S. S. (1962). *Affect, Imagery, Consciousness: Vol. I. The Positive Affects.* Oxford: Springer.

Vanwesenbeeck, I., Ponnet, K., and Walrave, M. (2016). Go with the flow: how children's persuasion knowledge is associated with their state of flow and emotions during advergame play. *J. Consum. Behav.* 15, 38–47. doi: 10.1002/cb.1529

Vuorre, M., and Metcalfe, J. (2016). The relation between the sense of agency and the experience of flow. *Conscious. Cogn.* 43, 133–142. doi: 10.1016/j.concog.2016.06.001

Wanzer, D. L., Finley, K. P., Zarian, S., and Cortez, N. (2018). Experiencing flow while viewing art: Development of the aesthetic experience questionnaire. *Psychol Aesthet.Creat.Arts* 14, 113–124. doi: 10.1037/aca0000203

Waterman, A. S., Schwartz, S. J., Goldbacher, E., Green, H., Miller, C., and Philip, S. (2003). Predicting the subjective experience of intrinsic motivation: the roles of self-determination, the balance of challenges and skills, and self-realization values. *Personal. Soc. Psychol. Bull.* 29, 1447–1458. doi: 10.1177/0146167203256907

APPENDIX

How the Publications in Table 1 Were Selected

An "advanced search" in PsycINFO specified the following parameters:

(1) Publication date: July 2014 to July 2019.
(2) Publication type: Peer-reviewed journals.
(3) Subject: Major heading: Flow (consciousness state).

This yielded 111 publications. Forty-one of these publications did not include a flow operationalization, and were therefore not included in the review. Of the remaining 70 publications, those which included one or more of the following features were also not included in the review:

(1) Publications in which flow was operationalized as a trait-level construct (e.g., "flow proneness") rather than a state-level construct.
(2) Publications in which flow was operationalized as "collective flow."
(3) Publications in which the flow operationalization was not clearly described.
(4) Publications in which flow was operationalized in two or more distinct ways.
(5) Publications not in English.
(6) Validation studies.

This process yielded the 42 publications shown in **Table 1**. Although not exhaustive (given the inclusion criteria above), the listing is intended to be adequately representative of the operationalizations found in the psychological literature.

Investigating Flow State and Cardiac Pre-ejection Period During Electronic Gaming Machine Use

W. Spencer Murch[1]*, Mario A. Ferrari[1], Brooke M. McDonald[1,2] and Luke Clark[1]

[1] Centre for Gambling Research at UBC, Department of Psychology, The University of British Columbia, Vancouver, BC, Canada, [2] Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada

Flow activities (e.g. sports and gaming) have been associated with positive affect and prolonged engagement. In the gambling field, modern electronic gaming machines (EGMs, including modern slot machines) have drawn concern as a potentially flow-inducing activity that may be associated with gambling-related harms. Current research has heavily relied on self-reported flow, and further insights may be afforded by physiological methods. We present data from three separate experiments in which self-reported gambling flow and cardiac pre-ejection period (PEP; a measure of sympathetic nervous system arousal) were examined. Male undergraduate participants gambled on a genuine EGM in a laboratory setting for a period of at least 15 min, and completed the Flow subscale of the game experience questionnaire (GEQ). Aggregated data were analyzed using multilevel regression. Although EGM gambling was not associated with significant changes in PEP across participants, we found that self-reported flow states were associated with significant decreases in PEP during the first five minutes of EGM use. Thus, participants who experienced flow showed a greater sympathetic nervous system response to the onset of gambling. Though these effects were consistent in experiments 1 and 2, in experiment 3 the effect was inverted during the same time window. We conclude that flow during EGM gambling appears to be associated with early changes in sympathetic nervous system activity, but stress that more research is needed to characterize boundary conditions and moderating factors.

Keywords: flow, immersion, heart rate, gambling, slot machine, electronic gaming machine, impedance cardiography, pre-ejection period

INTRODUCTION

"Although it is possible to flow while engaged in any activity, some situations appear to be designed almost exclusively so as to provide the experience of flow."

–Mihaly Csikszentmihalyi

Flow – the trance-like experience of extreme focus on a task or activity – is often described in the context of leisure activities such as rock climbing, chess, or art (Jackson and Marsh, 1996; Csikszentmihalyi, 2014; Stavrou et al., 2015). The experience is typically associated with increases in positive affect (Asakawa, 2004; Rogatko, 2009; Csikszentmihalyi, 2014; Murch et al., 2017).

According to Flow Theory, some activities are particularly adept at eliciting and sustaining the flow experience, and gambling has been proposed as one such "flow activity" (Csikszentmihalyi, 2014, p. 140, 146). A negative implication is that businesses may offer products specifically designed to encourage flow, capitalizing on prolonged or more frequent participation by individuals who are seeking to escape from stress or low mood (Schüll, 2012; Dixon et al., 2018).

Researchers first became interested in the flow-related aspects of gambling in the 1980s. Jacobs (1986) proposed that gambling activities can provide a pleasurable, trance-like sensation that reduces gamblers' self-awareness. He suggested that this state of absorption in gambling was akin to the clinical symptom of dissociation, although in modern accounts it appears more characteristic of non-pathological or "normative" dissociation (Butler, 2006; Thomson and Jaque, 2012). Jacobs posited that this state of absorption could contribute to gambling addictions, and that these experiences could be addictive in-and-of-themselves (Jacobs, 1986, 1988). One study aimed to directly compare Csikszentmihalyi's and Jacobs' constructs in samples of student athletes and problem gamblers (Wanner et al., 2006). Results showed that the problem gambler group endorsed every item on both the Flow Trait Scale, and Jacobs' "Dissociation Questionnaire" (Jacobs, 1986; Jackson and Marsh, 1996). In an earlier analysis of data in the current study, we reported high internal consistency between items on the Dissociation Questionnaire (which includes feelings of being "in a trance," and losing track of time; Jacobs, 1986), and the Flow subscale of the GEQ ("I felt completely absorbed," "I forgot everything around me"; Poels and De Kort, 2007; IJsselsteijn et al., 2013; Murch and Clark, 2019), again suggesting considerable overlap between Flow Theory and Jacobs' absorption construct (but see Murch et al., 2019).

The susceptibility of regular gamblers to experiencing gambling flow is reliably associated with symptoms of disordered gambling; the Dissociation Questionnaire has been repeatedly correlated with measures of problem gambling (Kofoed et al., 1997; Diskin and Hodgins, 1999; Wanner et al., 2006; Noseworthy and Finlay, 2009; Hopley and Nicki, 2010; Cartmill et al., 2015; Murch et al., 2017, 2019; Dixon et al., 2018). In two experiments, gamblers were asked to monitor an area off-screen at the same time as they gambled on an electronic gaming machine (EGM; including modern slot machines), providing a response when target shapes appeared off-screen (Diskin and Hodgins, 1999; Murch et al., 2017). In both studies, levels of problematic gambling were associated with reduced detection of peripheral targets while gambling. This effect is consistent with the "attentional narrowing" mechanism proposed in Flow Theory (Csikszentmihalyi, 2014, p. 139). A more granular investigation of gambling flow found that specific flow experiences may have protective (losing track of time, autotelic experiences) or aggravating (senses of concentration and control) effects on gambling harms (Trivedi and Teichert, 2017; see also Palomäki and Laakasuo, 2016).

Some forms of gambling may be especially good at eliciting flow. EGMs are disproportionately associated with problem gambling (Breen and Zimmerman, 2002; MacLaren, 2015;

Binde et al., 2017; Gainsbury et al., 2019). Recent accounts of EGM gambling have argued that these devices may be designed to maximize "time on device," conceivably via flow experiences (Schüll, 2012, p. 74). In an Australian survey of gamblers who endorsed feeling "in a trance" while gambling, 79% of respondents had been using an EGM at the time (Office for Problem Gambling, 2006). Several scholars have proposed that absorption in EGM gambling may be an effective (though ultimately maladaptive) coping strategy for those seeking to avoid symptoms of depression, anxiety, or stress (Schüll, 2012; Dixon et al., 2018, 2019).

Current research relies heavily on self-report measures of flow, which can be susceptible to disruption (e.g. by introducing a secondary task; Murch et al., 2017). Psychophysiological methods may provide alternative markers for investigating the gambling flow phenomenon more covertly. Past examinations of EGM use have suggested a role for both the sympathetic (Anderson and Brown, 1984; Griffiths, 1993; Coventry and Constable, 1999; Coventry and Hudson, 2001) and parasympathetic nervous systems (Murch et al., 2017; Murch and Clark, 2019). However, little evidence exists for a link between gambling flow and physiological measures. In two experiments, we found no significant relationships between EGM flow and respiratory sinus arrhythmia, a cardiac marker of parasympathetic nervous system tone (Murch et al., 2017; Murch and Clark, 2019).

The present study evaluated the relationship between gambling flow and sympathetic nervous system arousal, indexed by cardiac pre-ejection period (PEP). PEP is an impedance cardiography-derived metric, which approximates the interval between onset of the electrical signal that stimulates left ventricular contraction (QRS complex) and opening of the aortic valve (commencement of blood efflux from the left ventricle into the aorta). In human studies and animal models, PEP has demonstrated excellent validity as an inverse measure of sympathetic arousal (Cacioppo et al., 2007, pp. 461, 619). In human studies, PEP has been observed to decrease (indicating sympathetic arousal) in response to anger, disgust, and fear emotional induction, and increase in response to happiness, sadness, and amusement (Kreibig, 2010). PEP is also sensitive to reward anticipation and delivery: PEP decreased when participants anticipated social reward (Brinkmann and Franzen, 2017), and was linearly related to reward size in delayed-match-to-sample tasks (Richter and Gendolla, 2009; Brinkmann and Franzen, 2013).

We report data from three laboratory experiments, in which self-reported flow and PEP data were collected for an EGM gambling session that lasted at least 15 min. We first hypothesized that PEP would decrease (relative to baseline) in response to EGM gambling, indicating sympathetic nervous system arousal associated with the gambling activity. We divided the gambling sessions into 5-min blocks to test the time-course of this response, as the effects of gambling on PEP may not be uniform across a gambling session. Our second and primary hypothesis proposed that EGM-related changes in PEP would interact significantly with participants' flow ratings.

MATERIALS AND METHODS

Participants

The studies included in these analyses were approved by UBC's Behavioural Research Ethics Board. Participants were recruited to three experiments conducted between 2015 and 2018 ($N_1 = 121$, age $M = 21.25$, $SD = 2.91$; $N_2 = 80$, age $M = 20.55$, $SD = 2.37$; $N_3 = 106$, age $M = 20.80$, $SD = 2.39$, **Figure 1**). Primary analyses for Studies 1 and 2 are already published (Ferrari et al., 2018; Murch and Clark, 2019), without the measures of PEP. Study 3 has not been submitted for peer-reviewed publication (Murch, 2016). Study 1 was primarily interested in examining testosterone change in relation to EGM gambling. Study 2 looked at levels of flow and heart rate variability during EGM gambling with differing bet strategies tested within-subjects. Study 3 examined gambling immersion using a social manipulation, in which participants who provided psychophysiological data were, in some cases, tested alongside other participants seated at adjacent EGMs. Participants in Studies 2 and 3 gambled while an experimenter seated behind them monitored the physiological recording. Participants in Study 1 gambled without anyone else in the room. All participants were male undergraduate students, at least 19 years of age, who responded to an online advertisement posted by the psychology department. Most participants were compensated with partial course credit, though some participants in Study 1 were paid $15 CAD instead. Participants were included only if they were not high-risk problem gamblers (i.e. problem gambling severity index score <8, see below), had no allergies to gels or adhesives, and no current prescriptions for psychotropic or cardiac medications.

Questionnaires

Participants completed the Problem Gambling Severity Index (PGSI), which probes past-year problem gambling symptoms (Ferris and Wynne, 2001). Responses to the nine items were rated on a 4-point Likert scale ranging from "Never" (0), to "Almost always" (3), and a total score was obtained.

After gambling on the EGMs, participants completed the Flow subscale of the GEQ ("I felt completely absorbed," "I forgot everything around me"; Poels and De Kort, 2007; IJsselsteijn et al., 2013). In Study 2, participants completed this questionnaire after each of four 5-min gambling blocks. Responses were given on a 5-point Likert scale ranging from "Not at all" (0), to "Extremely" (4). Scores for the two items were averaged, and scores were standardized within each study. Past analysis of these items in Study 2 indicated relatively high reliability estimates (Cronbach's $\alpha = 0.80$; Murch and Clark, 2019).

Procedure

After providing written consent, participants completed the PGSI. Individuals scoring greater than seven (indicating high risk problem gambling), on this measure were excluded from the gambling task, and instead proceeded straight to debriefing. The lab was then cleared of any additional participants and participants providing physiological data were asked to remove their shirt for electrode placement. For the impedance signal, we applied eight Ag/AgCl electrodes (Vermed, Buffalo, NY, United States); four were applied laterally on the neck and four were applied laterally on the chest below the armpit (e.g. Gramzow et al., 2008). For an electrocardiogram, we then applied three electrodes to the upper left pectoral, upper right pectoral

FIGURE 1 | Inclusion flowchart.
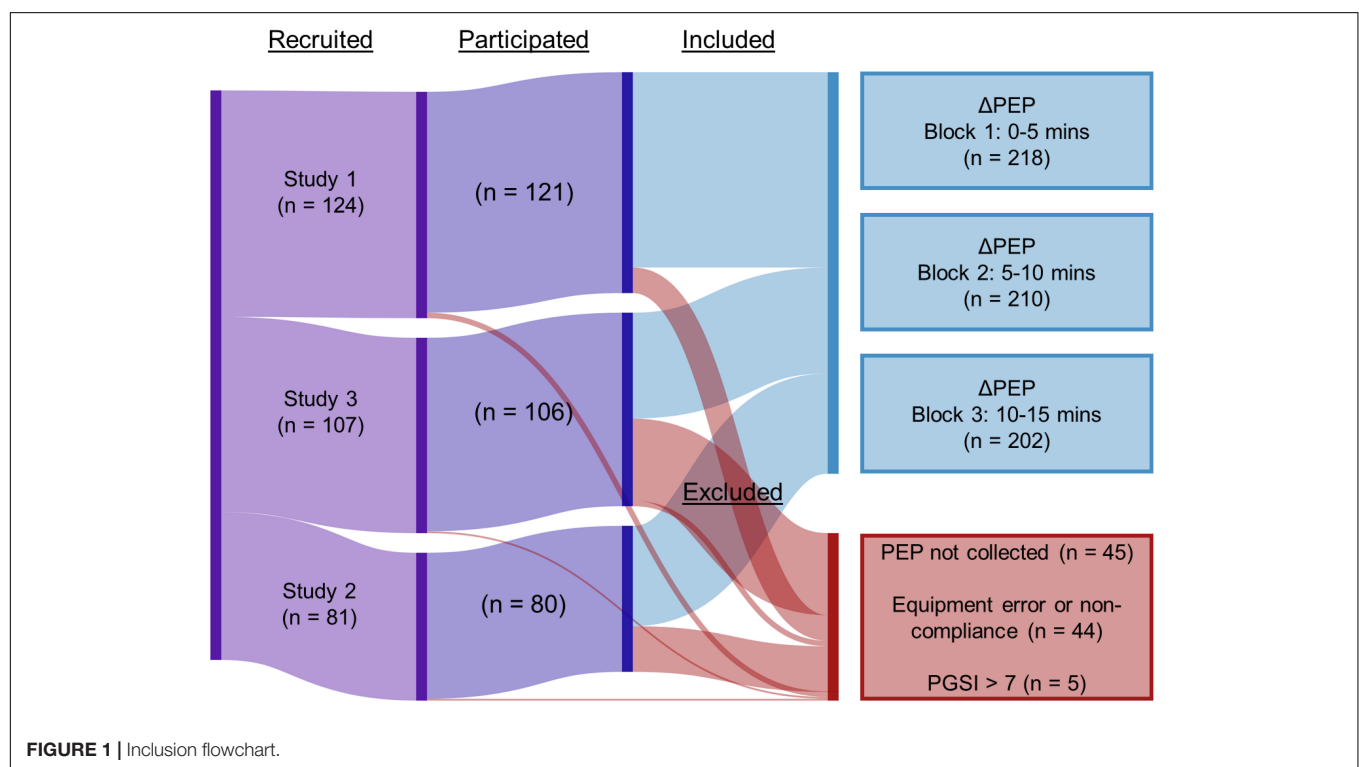
and lower left abdomen. The data were relayed wirelessly to the RSPEC-R and NICO-R modules of a Biopac MP150 system (BIOPAC Systems, Inc., Goleta, CA, United States) recording at 1,000 Hz. Participants then put their shirts on and provided a 5-min baseline recording in a seated position. In Study 1, participants' baseline recording was obtained at the same time as they provided a saliva sample via passive drool into a small vial. Participants in Studies 2 and 3 were instructed to close their eyes during the baseline recording, but did not provide a saliva sample.

In each study, participants gambled on a genuine EGM for at least 15 min. Each EGM was a modern, multi-line device (see Dixon et al., 2014), set on a one cent denomination (i.e. if betting on a single line, each spin would cost $0.01). In Study 1, participants gambled continuously on the EGM "Dragon's Fire" (Scientific Games Co., Las Vegas, NV). Study 2 consisted of four 5-min gambling blocks on the EGM "Buffalo Spirit," (Scientific Games Co., Las Vegas, NV). Study 2 participants completed the GEQ Flow questionnaire after each block. This introduced a break lasting approximately 1-min between blocks. Participants who provided PEP data in Study 3 gambled on "Double Diamond," or "Triple Diamond," (IGT, Las Vegas, NV, United States). Participants in all studies were provided $40–60 CAD (equivalent to 4,000–6,000 in-game credits) to use on the machine. Each study constrained participants' betting strategies in some way. In Studies 1 and 3, a multi-line bet strategy was set, at $0.40 and $0.20, respectively, to ensure frequent reinforcement (Livingstone and Woolley, 2008; Murch et al., 2017). In Study 2, bet strategies were systematically manipulated, from one credit bet on one payline (i.e. a $0.01 bet, the minimum), to five credits bet on each of 20 paylines (i.e. $1.00 per spin). Each study involved a cash bonus incentive: participants in Study 1 were paid a $10 bonus if they finished the session in profit (i.e. over 4,000 credits), whereas participants in Studies 2 and 3 received a variable bonus from $2 to $12 based on their remaining credits.

Processing and Analyses

Pre-ejection period is defined as the latency between Q-wave onset in an electrocardiogram, which reflects the onset of the electrical signal prompting left ventricular contraction, and the upward inflection point (B) in the derived impedance signal, dZ/dt (Cacioppo et al., 2007). In order to address our time-course hypotheses and retain comparability to baseline recordings, PEP data were partitioned into 5-min blocks. As 15 min was the shortest session length, we extracted the first three blocks (0–5, 5–10, and 10–15 min) from each study. Physiological data were visually inspected for artifacts. Blocks were excluded in cases where either the participant had run out of credit and stopped gambling, or serious artifacts precluded an accurate extraction of PEP. PEP extraction was completed using the PEP algorithm in Acqknowledge 4.4 (BIOPAC Systems, Inc., Goleta, CA, United States). Complete or partial PEP data was available for 218 participants across the three studies (**Figure 1**). Baseline PEP scores were subtracted from PEP scores for each block. This array of difference scores represents the change in PEP from baseline to each block, the dependent variable "ΔPEP."

We performed three linear multilevel regression models with maximum likelihood estimation to predict ΔPEP given the block in which it was recorded, and the self-reported flow state score associated with that block. Participants in Study 2 gave flow ratings for each of the three blocks separately, while participants in Studies 1 and 3 gave a single flow rating for all blocks after the session was completed. ΔPEP blocks were nested within participants and studies, and we examined indices of model fit (AIC and BIC) to determine that these factors should be modeled as random effects (Field, 2012). Block and study number were dummy-coded. Since ΔPEP was calculated by subtracting baseline PEP levels, a value that reflects no task-related change is necessarily equal to zero, and as such the model intercept was suppressed. Models 1 and 2 directly address our hypotheses. Model 3 was included in order to explore the simple main effects of block and study on the relationship between flow and PEP. This allowed us to investigate whether any effects observed in Model 2 appeared heterogeneously across different experimental contexts.

Model 1: ΔPEP predicted by block number.

Model 2: ΔPEP predicted by block number and block-by-flow interaction terms.

Model 3: ΔPEP predicted by block number and block-by-flow-by-study interaction terms.

Analyses were performed in JASP, and R version 3.5.2, using the "nlme" package (Fox and Weisberg, 2011; R Core Team, 2018; JASP Team, 2019; Pinheiro et al., 2019). To assess the underlying assumptions of linearity and homoscedasticity, we calculated variance inflation factors and visually inspected the distributions of fitted and residual values at the levels of the factors and random effects. We were satisfied that the models did not violate the underlying assumptions of the analyses. These data and analyses have been publicly archived[1].

RESULTS

The overall mean PEP during baseline blocks was 106.00 ms (SD = 21.79 ms). Task-related PEP levels were comparable during Block 1 (mean = 105.98 ms, SD = 22.17 ms), Block 2 (mean = 105.32 ms, SD = 19.87 ms), and Block 3 (mean = 106.47 ms, SD = 20.36 ms). The mean GEQ Flow score was 1.62 (SD = 1.12) in Study 1, 1.14 (SD = 0.72) in Study 2, and 1.21 (SD = 1.02) in Study 3, indicating mild-to-moderate levels of flow in the three experiments. A one-way ANOVA indicated that the average GEQ Flow scores differed significantly between the three studies [$F(2,197.62)$ = 6.93, p = 0.001; Welch correction employed due to unequal variances], with higher scores in Study 1 than Study 2 ($p_{Bonferroni}$ = 0.004), and Study 3 ($p_{Bonferroni}$ = 0.007), but not between Studies 2 and 3 ($p_{Bonferroni}$ > 0.99).

Regression Model Results

Model 1: Overall, there was no significant change in PEP relative to baseline levels [Block 1: B = 0.51, $t(408)$ = 0.51, p = 0.61; Block 2: B = −0.37, $t(408)$ = −0.36, p = 0.72; Block 3: B = 0.61, $t(408)$ = 0.60, p = 0.55].

---

[1]https://dataverse.scholarsportal.info/dataset.xhtml?persistentId=doi:10.5683/SP2/JFR1B3

Model 2: ΔPEP again did not differ significantly from baseline for Block 1 [$B = 0.70$, $t(405) = 0.70$, $p = 0.48$], Block 2 [$B = -0.22$, $t(405) = -0.22$, $p = 0.83$], or Block 3 [$B = 0.72$, $t(405) = 0.71$, $p = 0.48$]. The block-by-flow interaction term was significant for Block 1 [$B = -1.89$, $t(405) = -2.13$, $p = 0.03$], but not for Block 2 [$B = -0.87$, $t(405) = -0.96$, $p = 0.34$], or Block 3 [$B = -0.50$, $t(405) = -0.56$, $p = 0.58$]. The model fit was not significantly improved over Model 1 [$\chi^2(3) = 5.24$, $p = 0.16$].

Model 3: ΔPEP did not differ significantly from zero for Block 1 ($p = 0.89$, **Table 1**), Block 2 ($p = 0.56$), or Block 3 ($p = 0.69$). In Study 1, ΔPEP interacted significantly with flow during Block 1 ($p = 0.01$, **Figure 2**), but not during Block 2 ($p = 0.11$), or Block 3 ($p = 0.33$). In Study 2, ΔPEP interacted significantly with flow during Block 1 ($p = 0.02$), but not during Block 2 ($p = 0.40$), or Block 3 ($p = 0.25$). Lastly, in Study 3, ΔPEP interacted significantly with flow during Block 1 ($p = 0.02$), but not during Block 2 ($p = 0.10$), or Block 3 ($p = 0.10$). Notably, the direction of the Block 1 effect differed from those observed in Studies 1 and 2. The model fit was significantly improved over Model 2 [$\chi^2(6) = 16.03$, $p = 0.01$].

DISCUSSION

We tested cardiac PEP as a potential sympathetic nervous system marker of flow while undergraduate students gambled on authentic EGMs situated in a laboratory environment. We examined whether PEP changes were associated with EGM use, the stability of these levels over time, and their associations with self-reported flow, using multilevel regression models that accounted for the nested data structure. We did not observe significant change in PEP from the pre-task baseline to gambling. When we examined the interaction between task block and flow on PEP during gambling, we found that self-reported flow was associated with decreases in PEP (indicating *increased*
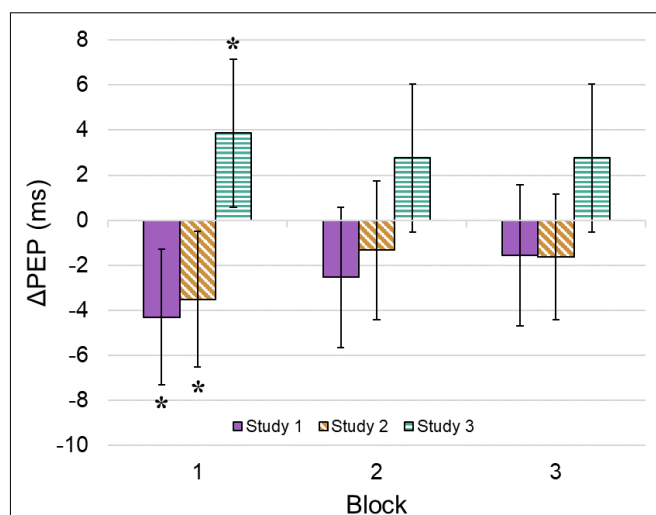
FIGURE 2 | Block-by-flow-by-study interactions. Plotted values represent interaction effect coefficients summarized in **Table 1** [i.e. data points represent predicted change in PEP from baseline for a participant whose immersion score was one standard deviation (SD) above the mean]. For example, the mean flow rating in Study 3 was 1.21 (*SD* = 1.02). A participant who gave a flow rating of 2.23 (+1 SD) in Study 3 is expected to have a 3.87 ms increase in PEP during Block 1 compared to their baseline level. A participant who gave a flow rating of 0.19 (−1 SD) in Study 3 is expected to have a 3.87 ms decrease in PEP during Block 1. Bars represent the 95% confidence interval. *$p < 0.05$.

TABLE 1 | Predicted ΔPEP from baseline, Model 3.

Predictor	B	SE (B)	t(399)	p
Main effects				
Block 1	0.14	1.00	0.14	0.89
Block 2	−0.58	1.01	−0.58	0.56
Block 3	0.40	1.01	0.40	0.69
Interactions				
Block 1 × Flow × Study 1	−4.29	1.54	−2.79	0.01
Block 2 × Flow × Study 1	− 2.53	1.59	−1.59	0.11
Block 3 × Flow × Study 1	−1.55	1.60	−0.97	0.33
Block 1 × Flow × Study 2	−3.51	1.54	−2.28	0.02
Block 2 × Flow × Study 2	−1.32	1.57	−0.84	0.40
Block 3 × Flow × Study 2	−1.64	1.42	−1.15	0.25
Block 1 × Flow × Study 3	3.87	1.67	2.31	0.02
Block 2 × Flow × Study 3	2.76	1.67	1.65	0.10
Block 3 × Flow × Study 3	2.77	1.67	1.65	0.10

Flow scores have been standardized. Block and Study factors represent dummy-codes. Values in column B are unstandardized coefficients. Values in column SE (B) represent the standard error for the coefficient in that row.

sympathetic nervous system activity) during Block 1 (the first 5 min of gambling). When we explored this interaction within the three studies, we found opposing relationships between block and flow on ΔPEP. Studies 1 and 2 showed results consistent with Model 2: higher self-reported flow states during gambling were associated with greater decreases in PEP during Block 1 (but not Blocks 2 or 3). In Study 3, flow was associated with increased PEP (i.e. reduced sympathetic activity) and again, this effect was only statistically significant during Block 1. Taking these results together, it appears that early physiological responses to EGM use were related to increases in participants' subsequent flow ratings. We have thus found tentative support for an association between subjective flow and fluctuations in sympathetic nervous system activity. Crucially, however, the direction of this effect may depend on particular aspects of the task procedure.

It is worth speculating on why the observed interactions with flow were limited to the first 5 min of gambling. As our flow ratings were taken at the end of the session in Studies 1 and 3, this firstly indicates that participants' early experiences of the EGM are particularly important in accounting for variability in later flow ratings (a kind of primacy effect). These results further indicate that the initial physiological response to EGM use is an important factor in determining whether the session produces flow overall. Perhaps early experiences that produce physiological change increase the likelihood that gamblers will experience flow. In future research, it would be fruitful to take multiple flow measurements within a prolonged EGM

gambling session, to characterize the subjective time-course, although such designs are challenging due to the potential for distractors to impair flow.

One possible explanation for the opposing results across the three studies is the social manipulation present in Study 3. Participants in that experiment were made aware that they may be gambling alongside other participants, and this may have impacted either their physiological response or experience of flow while gambling. Further, the gambling sessions in Study 1 (which saw the largest effect at Block 1) were conducted without an experimenter present in the room (in order to minimize any observer effects on risk-taking; Rockloff and Dyer, 2007; Rockloff et al., 2011). Thus participants' physiological responses to the gambling task may have been moderated by these social factors, either of researchers or other participants. Alternatively, our effects could be related to participants in each study employing different betting strategies. This necessarily affected the rate of reinforcement in these studies and may have also had an impact on self-reported flow state (Murch and Clark, 2019). Consistent with past findings, we found the highest levels of flow in Study 1, which employed a 40-line bet strategy to achieve high rates of reinforcement (Livingstone and Woolley, 2008; Templeton et al., 2015). Study 3 employed a smaller, 20-line strategy, and Study 2 compared several bet strategies that varied the number of lines bet, either 1, 5, or 20. Thus, if there is a real relationship between PEP and flow during EGM use, it may depend on additional factors that we could not systematically control in this aggregated analysis.

When not accounting for flow, we observed no significant change in PEP while gambling. Previous work has typically inferred sympathetic arousal from increases in mean heart rate during gambling, including on EGMs (Anderson and Brown, 1984; Griffiths, 1993; Coventry and Norman, 1997; Coventry and Constable, 1999; Coventry and Hudson, 2001). However, the physiology of heart rate change is complex, and affected by both branches of the autonomic nervous system (Cacioppo et al., 2007). Decreases in vagal tone while gambling (Murch et al., 2017; Murch and Clark, 2019), could potentially increase heart rate while sympathetic arousal remains constant, accounting for past results. A separate possibility is that heart rate effects did reflect sympathetic arousal in past experiments, but our laboratory environment or PEP measure may have lacked the sensitivity needed to detect a sympathetic response here.

Our findings are preliminary and intended to stimulate further enquiry; they have several important limitations. First, the three study protocols differed in numerous ways, and it is possible that methodological differences drove the disparate pattern of results. Second, the laboratory environment may have attenuated physiological reactivity. EGM gambling is regarded as an appetitive psychological challenge that involves intense audiovisual stimuli, motor actions and monetary outcomes, but responses to EGM use may differ based on whether the device is situated in a gambling venue, or in a laboratory environment (c.f. Anderson and Brown, 1984). Third, participants were convenience-sampled from an undergraduate population and were not regular EGM users. This potentially diminished both physiological responses to the EGM task, and the level of flow that was reported. Fourth, participants were men, because practical application of our PEP methods precluded the recruitment of women. Fifth, the GEQ Flow scale is unidimensional, focusing on absorption states, and other measures may provide insight into different aspects of the flow state (e.g. Jackson and Eklund, 2002). Finally, the block-by-flow-by-study analytic approach was exploratory, and the available data could not clarify why opposing effects were observed between the studies. Our preliminary conclusion is that cardiac sympathetic nervous system responses early in an EGM gambling session may affect subsequent ratings of flow for that session. However, follow-up studies should be undertaken in an attempt to replicate and clarify this effect.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by UBC Behavioural Research Ethics Board. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

WM collected and processed the data, aggregated the data across experiments, performed the analyses, and drafted the manuscript. MF and BM collected and processed the data, and contributed to writing. LC supervised the experiments, provided the testing facilities and research funding, and contributed to writing.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Anderson, G., and Brown, R. I. F. (1984). Real and laboratory gambling, sensation-seeking and arousal. *Br. J. Psychol.* 75, 401–410. doi: 10.1111/j.2044-8295.1984.tb01910.x

Asakawa, K. (2004). Flow experience and autotelic personality in Japanese college students: how do they experience challenges in daily life? *J. Happiness Stud.* 5, 123–154. doi: 10.1023/B:JOHS.0000035915.97836.89

Binde, P., Romild, U., and Volberg, R. A. (2017). Forms of gambling, gambling involvement and problem gambling: evidence from a Swedish population survey. *Int. Gambl. Stud.* 17, 490–507. doi: 10.1080/14459795.2017.1360928

Breen, R. B., and Zimmerman, M. (2002). Rapid onset of pathological gambling in machine gamblers. *J. Gambl. Stud.* 18, 31–43. doi: 10.1023/A:1014580112648

Brinkmann, K., and Franzen, J. (2013). Not everyone's heart contracts to reward: insensitivity to varying levels of reward in dysphoria. *Biol. Psychol.* 94, 263–271. doi: 10.1016/j.biopsycho.2013.07.003

Brinkmann, K., and Franzen, J. (2017). Blunted cardiovascular reactivity during social reward anticipation in subclinical depression. *Int. J. Psychophysiol.* 119, 119–126. doi: 10.1016/j.ijpsycho.2017.01.010

Butler, L. D. (2006). Normative dissociation. *Psychiatr. Clin. North Am.* 29, 45–62. doi: 10.1016/j.psc.2005.10.004

Cacioppo, J., Tassinary, L. G., and Berntson, G. G. (2007). *The Handbook of Psychophysiology*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511546396

Cartmill, T., Slatter, T., and Wilkie, B. (2015). The role of anxiety and dissociation in young Australian gamblers. *J. Gambl. Stud.* 31, 1215–1226. doi: 10.1007/s10899-014-9510-1

Coventry, K. R., and Constable, B. (1999). Physiological arousal and sensation-seeking in female fruit machine gamblers. *Addiction* 94, 425–430. doi: 10.1046/j.1360-0443.1999.94342512.x

Coventry, K. R., and Hudson, J. (2001). Gender differences, physiological arousal and the role of winning in fruit machine gamblers. *Addiction* 96, 871–879. doi: 10.1080/09652140020050997

Coventry, K. R., and Norman, A. C. (1997). Arousal, sensation seeking and frequency of gambling in off-course horse racing bettors. *Br. J. Psychol.* 88, 671–681. doi: 10.1111/j.2044-8295.1997.tb02664.x

Csikszentmihalyi, M. (2014). *Flow and the Foundations of Positive Psychology*. New York, NY: Springer. doi: 10.1007/978-94-017-9088-8

Diskin, K. M., and Hodgins, D. C. (1999). Narrowing of attention and dissociation in pathological video lottery gamblers. *J. Gambl. Stud.* 15, 17–28. doi: 10.1023/A:102306291

Dixon, M. J., Graydon, C., Harrigan, K. A., Wojtowicz, L., Siu, V., and Fugelsang, J. A. (2014). The allure of multi-line games in modern slot machines. *Addiction* 109, 1920–1928. doi: 10.1111/add.12675

Dixon, M. J., Gutierrez, J., Stange, M., Larche, C. J., Graydon, C., Vintan, S., et al. (2019). Mindfulness problems and depression symptoms in everyday life predict dark flow during slots play: implications for gambling as a form of escape. *Psychol. Addict. Behav.* 33, 81–90. doi: 10.1037/adb0000435

Dixon, M. J., Stange, M., Larche, C. J., Graydon, C., Fugelsang, J. A., and Harrigan, K. A. (2018). Dark flow, depression and multiline slot machine play. *J. Gambl. Stud.* 34, 73–84. doi: 10.1007/s10899-017-9695-1

Ferrari, M. A., Chan, M., Brown, P. N., and Clark, L. (2018). Slot machine gambling and testosterone: evidence for a "winner–loser" effect? *Psychol. Addict. Behav.* 32, 961–971. doi: 10.1037/adb0000425

Ferris, J., and Wynne, H. (2001). *The Canadian Problem Gambling Index: Final Report*. Ottawa: Canadian Centre on Substance Abuse. doi: 10.1007/s10899-010-9224-y

Field, A. (2012). *Discovering Statistics Using r*. London: Sage Publications Ltd.

Fox, J., and Weisberg, S. (2011). *An R Companion to Applied Regression*, 2nd Edn. Thousand Oaks, CA: Sage.

Gainsbury, S. M., Angus, D. J., and Blaszczynski, A. (2019). Isolating the impact of specific gambling activities and modes on problem gambling and psychological distress in internet gamblers. *BMC Public Health* 19:1372. doi: 10.1186/s12889-019-7738-5

Gramzow, R. H., Willard, G., and Mendes, W. B. (2008). Big tales and cool heads: academic exaggeration is related to cardiac vagal reactivity. *Emotion* 8, 138–144. doi: 10.1037/1528-3542.8.1.138

Griffiths, M. (1993). Tolerance in gambling: an objective measure using the psychophysiological analysis of male fruit machine gamblers. *Addict. Behav.* 18, 365–372. doi: 10.1016/0306-4603(93)90038-B

Hopley, A. A. B., and Nicki, R. M. (2010). Predictive factors of excessive online poker playing. *Cyberpsychol. Behav. Soc. Netw.* 13, 379–385. doi: 10.1089/cyber.2009.0223

IJsselsteijn, W. A., de Kort, Y. A. W., and Poels, K. (2013). *The Game Experience Questionnaire*. Eindhoven: Technische Universiteit Eindhoven.

Jackson, S. A., and Eklund, R. C. (2002). Assessing flow in physical activity: the flow state scale-2 and dispositional flow scale-2. *J. Sport Exerc. Psychol.* 24, 133–150.

Jackson, S. A., and Marsh, H. (1996). Development and validation of a scale to measure optimal experience: the flow state scale. *J. Sport Exerc. Psychol.* 18, 17–35. doi: 10.1123/jsep.18.1.17

Jacobs, D. F. (1986). A general theory of addictions: a new theoretical model. *J. Gambl. Behav.* 2, 15–31. doi: 10.1007/BF01019931

Jacobs, D. F. (1988). Evidence for a common dissociative-like reaction among addicts. *J. Gambl. Behav.* 4, 27–37. doi: 10.1007/BF01043526

JASP Team, (2019). *JASP (Version 0.10)*.

Kofoed, L., Morgan, T. J., Buchkowski, J., and Carr, R. (1997). Dissociative experiences scale and MMPI-2 scores in video poker gamblers, other gamblers, and alcoholic controls. *J. Nerv. Ment. Dis.* 185, 58–60. doi: 10.1097/00005053-199701000-00011

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biol. Psychol.* 84, 394–421. doi: 10.1016/j.biopsycho.2010.03.010

Livingstone, C., and Woolley, R. (2008). *The Revelance and Role of Gaming Machine Games and Game Features on the Play of Problem Gamblers*. Adelaide: Independent Gambling Authority. doi: 10.13140/RG.2.1.3070.2249

MacLaren, V. V. (2015). Video lottery is the most harmful form of gambling in Canada. *J. Gambl. Stud.* 32, 459–485. doi: 10.1007/s10899-015-9560-z

Murch, W. S. (2016). *Pay as you Flow: Measuring the Slot Machine Zone with Attentional Dual Tasks*. Vancouver: The University of British Columbia. doi: 10.14288/1.0307373

Murch, W. S., Chu, S. W. M., and Clark, L. (2017). Measuring the slot machine zone with attentional dual tasks and respiratory sinus arrhythmia. *Psychol. Addict. Behav.* 31, 375–384. doi: 10.1037/adb0000251

Murch, W. S., and Clark, L. (2019). Effects of bet size and multi-line play on immersion and respiratory sinus arrhythmia during electronic gaming machine use. *Addict. Behav.* 88, 67–72. doi: 10.1016/j.addbeh.2018.08.014

Murch, W. S., Limbrick-Oldfield, E. H., Ferrari, M. A., MacDonald, K. I., Fooken, J., Cherkasova, M. V., et al. (2019). Zoned in or zoned out? Investigating immersion in slot machine gambling using mobile eye tracking. *Addiction* doi: 10.1111/add.14899

Noseworthy, T. J., and Finlay, K. (2009). A comparison of ambient casino sound and music: effects on dissociation and on perceptions of elapsed time while playing slot machines. *J. Gambl. Stud.* 25, 331–342. doi: 10.1007/s10899-009-9136-x

Office for Problem Gambling, (2006). *Gambling Prevalence in South Australia*. Adelaide: Office for Problem Gambling.

Palomäki, J., and Laakasuo, M. (2016). Zone out and your money flows away; go with the flow and your money might stay. *Nord. Stud. Alcohol Drugs* 33, 111–113. doi: 10.1515/nsad-2016-0007

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team, (2019). *nlme: Linear and Nonlinear Mixed Effects Models*. Available at: https://cran.r-project.org/package=nlme (accessed January 13, 2020).

Poels, K., and De Kort, Y. (2007). "It is always a lot of fun!" exploring dimensions of digital game experience using focus group methodology," in *Proceedings of the International Academic Conference on the Future of Game Design and Technology: FuturePlay*, Toronto, 83–89.

R Core Team, (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Richter, M., and Gendolla, G. H. E. (2009). The heart contracts to reward: monetary incentives and preejection period. *Psychophysiology* 46, 451–457. doi: 10.1111/j.1469-8986.2009.00795.x

Rockloff, M. J., and Dyer, V. (2007). An experiment on the social facilitation of gambling behavior. *J. Gambl. Stud.* 23, 1–12. doi: 10.1007/s10899-006-9042-4

Rockloff, M. J., Greer, N., and Fay, C. (2011). The social contagion of gambling: how venue size contributes to player losses. *J. Gambl. Stud.* 27, 487–497. doi: 10.1007/s10899-010-9220-2

Rogatko, T. P. (2009). The influence of flow on positive affect in college students. *J. Happiness Stud.* 10, 133–148. doi: 10.1007/s10902-007-9069-y

Schüll, N. D. (2012). *Addiction by Design*. Princeton, NJ: Princeton University Press.

Stavrou, N. A. M., Psychountaki, M., Georgiadis, E., Karteroliotis, K., and Zervas, Y. (2015). Flow theory – goal orientation theory: positive experience is related to athlete's goal orientation. *Front. Psychol.* 6:1499. doi: 10.3389/fpsyg.2015.01499

Templeton, J. A., Dixon, M. J., Harrigan, K. A., and Fugelsang, J. A. (2015). Upping the reinforcement rate by playing the maximum lines in multi-line slot machine play. *J. Gambl. Stud.* 31, 949–964. doi: 10.1007/s10899-014-9446-5

Thomson, P., and Jaque, S. V. (2012). Dancing with the muses: dissociation and flow. *J. Trauma Dissociation* 13, 478–489. doi: 10.1080/15299732.2011.652345

Trivedi, R. H., and Teichert, T. (2017). The janus-faced role of gambling flow in addiction issues. *Cyberpsychol. Behav. Soc. Netw.* 20, 180–186. doi: 10.1089/cyber.2016.0453

Wanner, B., Ladouceur, R., Auclair, A. V., and Vitaro, F. (2006). Flow and dissociation: examination of mean levels, cross-links, and links to emotional well-being across sports and recreational and pathological gambling. *J. Gambl. Stud.* 22, 289–304. doi: 10.1007/s10899-006-9017-5

Flow States and Associated Changes in Spatial and Temporal Processing

Scott Sinnett[1]*, Joshua Jäger[2†], Sarah Morgana Singer[2†] and Roberta Antonini Philippe[3]

[1] Department of Psychology, University of Hawaii at Manoa, Honolulu, HI, United States, [2] Institute of Psychology, University of Bern, Bern, Switzerland, [3] Laboratoire PHASE, Institut des Sciences du Sport, Faculté des Sciences Sociales et Politiques, Université de Lausanne, Lausanne, Switzerland

Improved perception during high performance is a commonly reported phenomenon. However, it is difficult to determine whether these reported changes experienced during flow states reflect veridical changes in perceptual processing, or if instead are related to some form of memory or response bias. Flow is a state in which an individual experiences high focus and involvement in a specific task, and typically experiences a lack of distractibility, a disordered sense of time, great enjoyment, and increased levels of performance. The present pre-registered study investigated 27 athletes and musicians using a temporal order judgement (TOJ) task before and after a sports or music performance over three sessions. Participants' flow experiences were surveyed in order to measure how modulations of flow over successive performances potentially modulates spatiotemporal perception and processing. Hierarchical linear modeling showed a positive moderation of subjectively experienced flow and performance on post-measures of a TOJ task. Specifically, the higher the subjective flow experience of the sport or music performance was rated, the better the participant performed in the post-performance TOJ task compared to the pre-performance TOJ task. The findings of the present study provide a more comprehensive explanation of human perception during flow at high level performances and suggest important insights regarding the possibility of modulated temporal processing and spatial attention.

Keywords: flow, temporal processing, spatial attention, hierarchical linear modeling, perception, sport, music, high performance

1. INTRODUCTION

Several anecdotal claims regarding improved perception during flow states have been reported in various populations, yet it remains an open question as to whether there is a veridical change in perception, or if instead these reported improvements in perception are related to post-performance memory biases. As an example of this supposed improvement in performance, George Scott, a professional baseball player, stated in an interview: "When you're hitting the ball [well], it comes at you looking like a grapefruit. When you're not, it looks like a blackeyed pea" (Witt and Proffitt, 2005, p. 937). In an attempt to disentangle the question of whether perception is indeed modulated during optimum performance levels, Witt and Proffitt (2005) correlated performance when playing softball (i.e., batting averages) and the perceived size of a softball. According to Witt and Proffitt (2005), successful players perceived the ball to be bigger than less successful players, with this finding leading the authors to further claim that enhanced performance levels are indeed capable of modulating perception. Similar findings have been found with darts players, with throwing ability in darts influencing the perceived size of a target (Wesp et al., 2004; Cañal-Bruland et al., 2010), such that participants with better accuracy chose

bigger circles corresponding to the size of the target than participants with lower accuracy. In addition, people who perform better in archery report seeing larger targets when compared to their weaker counterparts (Lee et al., 2012), and high-performing golfers (compared to weaker golfers) perceive the size of the cup to be larger (Witt et al., 2008). Moreover, distances are perceived as longer by people who are overweight (Sugovic et al., 2016), by people carrying a heavy backpack (Proffitt et al., 2003), or when throwing a heavy object toward a certain destination (Witt et al., 2004). Witt (2019) provides a review on action-specific effects on modulated spatial perception for a review on action-specific effects on modulated spatial perception.

These types of performance-dependent modulations of perception extend to the temporal domain. For instance, Gray (2013) found that not only was perceived ball size larger amongst high performing baseball players, but also the perceived speed of the ball, which was rated to be slower by better players. Additionally, tennis players who played better than other players perceived the ball to move more slowly (Witt and Sugovic, 2010). This evidence dovetails with the phenomenological experience of time slowing during threatening events (Arstila, 2012) as well as how the prospect of reward can affect the subjective perception of time (Failing and Theeuwes, 2016). Different approaches might offer an explanation for these phenomenological circumstances, of how or why our subjective perception of time changes when performing well or when placed in fear provoking circumstances. One possible explanation could derive from memory biases. For example, it is possible that players who perform poorly could perceive the size of the target during the game as exactly the same size as better players, but recall the size of the targets to be smaller, perhaps as a means to justify their poor performance (Cooper et al., 2012). A memory bias could also lead to the perception that time seems to slow down, because richer than usual memories could later be improperly connected in such a way that they span a longer period than the experience on which they were actually based on (Arstila, 2012). Furthermore, neurophysiological correlates have been identified that are related to perceptual alterations (van der Kruijs et al., 2014) might play a crucial role in experiencing such phenomenological modulations. Ursano et al. (2007) discuss in their investigation about dissociative reactions during traumatic events the contribution of the cerebellum in perceptual alteration concerning time and space.

When conceptualizing the state of the literature that has explored the phenomenological experience of modulated perception (see for example Witt and Proffitt, 2005), it is important to note that mostly intra-individual differences in perception that may result from different levels of performance are not taken into account; instead, most investigations have simply focused on how better players compare to weaker players (e.g., players with better batting averages vs. worse batting averages). It is entirely possible that a better player might perceive the ball or a target to be larger than a weaker player. This essentially equates to a measure of good vs. bad players, and is therefore uninformative with respect to the question of whether or not performing at a high level modulates perception.

Furthermore, to the best of our knowledge most researchers have also failed to consider the subjective evaluation that an athlete might have regarding their own personal performance, which could be different from the objective evaluation of, for instance, averaging their hitting rate. With specific respect to Witt and Proffitt (2005), batting average was based on self-report and calculated on a relatively small number of attempts (1–2 games only, without the exact number of at-bat attempts reported). As can clearly be seen, a longitudinal approach to investigating such phenomena is needed.

In addition to investigating potential perceptual modulations that arise due to high performance, phenomenological experiences associated with flow experiences have also been robustly explored. The experience of flow (Jackson and Csikszentmihalyi, 1999; Engeser and Rheinberg, 2008) refers to high performance in a task (e.g., athletics, music, etc.) that often involves increased levels of focus until complete immersion occurs, attention that is not distracted by anything irrelevant, feelings of optimal challenges, and deep enjoyment (Csikszentmihalyi, 1975, 2000). Recent research suggests that flow can be characterized by nine different dimensions (Csikszentmihalyi, 2000, 2002): challenge-skill balance (demanding situations in which the individual is engaged but not overwhelmed to meet the challenge), clear goals that derive from the activity, unambiguous feedback that helps individuals to constantly adapt in order to achieve their goals, concentration on the task at hand (one's focus relies on the activity and is not distracted by irrelevant stimuli), action-awareness merging (total immersion in the activity), loss of self-consciousness (individual's self-awareness and concerns regarding external evaluations decreases), increased sense of control (knowledge about the ability to keep things under control, if necessary), and transformation of time (disordered perception of time). The first three (i.e., challenge-skill balance, clear goals, and unambiguous feedback) are required conditions for flow to occur, while the remaining items refer to the phenomenological characteristics frequently associated with flow. These dimensions have been studied in many different populations mainly using self-report approaches (Moneta, 2012; Swann et al., 2012; Chirico et al., 2015; Stamatelopoulou et al., 2018; Habe et al., 2019).

Of direct concern to the research conducted here, the notion that the perception of time can be modulated when in flow states has been frequently reported, although these reports are almost exclusively anecdotal. For instance, an elite track and field athlete claimed that "When I went to throw it [the javelin], it was like things were in slow motion, and I could feel the position I was in, and I held my position for a long time" (Jackson, 1995, p.82). This statement, and others regarding the altered perception of time, refers to the speed at which the passage of time is experienced (Thönes and Stocker, 2019).

A challenge for the claim that the perception of time slows down during flow states can be found in the difficultly of disentangling the subjectively perceived experience of time from objective perception. It is unlikely that flow states would lead to (or arise from) a speed up in neuronal communication, with the question being further muddied by the fact that attempts to measure perception during flow states would almost

surely take the individual out of that state. As such, questions related to time perception during flow states are limited to posteriori surveys, therefore, the underlying processes of any modulation in the perception of the passage of time still remain unknown (Wearden, 2015; Tanaka and Yotsumoto, 2017).

The aim of the present study is to investigate whether potential changes in temporal and spatial processing are modulated by increases in flow experiences. Possible modulations of temporal processing and spatial attention can be measured with a temporal order judgement (TOJ) paradigm, a task that has been widely used as a tool to measure temporal and spatial processing. By means of the TOJ task, two different values can be calculated: The just noticeable difference (JND) and the point of subjective simultaneity (PSS) (West et al., 2008; Lim and Sinnett, 2011). The former is a measure of temporal processing and refers to the smallest amount of time needed to accurately separate two stimuli 75% of the time, and thus be able to correctly identify the order of presentation. The latter is a measure of spatial attention and reflects the extent to which attention is distracted by a spatial cue, either peripheral (exogenous) or central (endogenous), such that the uncued side must be presented before the cued side in order for both stimuli to be perceived as having been presented simultaneously.

The cues in the TOJ task create a prior entry effect (Shore et al., 2001): Attended stimuli are perceived before unattended stimuli, showing that temporal processing is influenced by attention (Shore et al., 2001). By presenting such cues in the TOJ task prior to the onset of the first stimulus, attention should be, at least in theory, directed toward the cued side, resulting in the cued side being detected first, even when both items had been presented simultaneously. That is, if the left and right stimuli appear simultaneously, for example, the stimulus at the cued side will be perceived as having occurred first and the PSS would indicate a shift of attention toward the cued side (Shore et al., 2001). The shift might be greater for exogenous cues than for endogenous cues, due to increased volitional control over orienting effects for central cues (Shore et al., 2001). Notably, evidence has been observed that faster stimulus perception associated with the prior cue reliably results from the allocation of spatial attention and not from any potential response bias (Ulrich, 1987; Stelmach and Herdman, 1991; Shore et al., 2001; West et al., 2009).

The TOJ paradigm has been used in several situations as a viable approach for measuring perception. For example, Lim and Sinnett (2011) showed lower JND scores for musicians, suggesting better temporal discrimination in musicians than in controls. Similarly, modulations in visual attention were observed after extensive action video game play, with West et al. (2008) showing greater sensitivity to exogenous sensory stimuli and the potential that video game play modulates spatial attention. While neither of these studies considered whether temporal perception might be modulated when these groups of participants were in flow states, these modulations in information processing nonetheless are attributable to long-term experience, and on a neurological level, arguably due to increased neuro-plasticity. For instance, Granek et al. (2010) and Gong et al. (2016) used fMRI to show greater activity in the prefrontal cortex within video-game experts

during complex non-gaming tasks, and increased functional integration between two critical neural networks for visual attention, namely the salience network and the central executive network, when compared to novices. These different brain patterns can be potentially explained by the task demands on visual attention that are associated with video games. Using behavioral and electrophysiological measures, Qiu et al. (2018) investigated the effects of short- and long-term action video gaming on measures of visual attention. After a short session of playing an action video game, experts and novices showed performance improvements in visual attention, with experts outperforming novices before the session. Importantly, modulated electrophysiological measures in novices were found. These findings provide evidence for a correlation between plasticity of visual attention and action video gaming, even after a brief session of gaming.

In this study we explored whether such modulation of temporal processing and spatial attention can also be observed depending on flow experience, in the short-term. Given the continuum of performance levels within any performer, flow experiences provide an ideal place to investigate how phenomenological experience might possibly modulate perception. While this is clearly the case considering anecdotal evidence, it is unknown whether there are objective enhancements in perception when participants experience a higher feeling of flow compared with when they experience a lower state of flow. By measuring flow levels and temporal processing across multiple sport and music performances in practice or rehearsal sessions, we are able to address this question to an extent that has not been done previously, to the best of our knowledge. Precisely, if individuals in a flow state do experience a slow down in the perception of time, this should be correlated with improved temporal processing of stimuli (i.e., a smaller JND) when in a flow state compared to when they are not performing at that optimal level. Additionally, we extended this question to determine whether individuals in a flow state are less distracted by exogenous or endogenous cues, potentially suggesting enhanced spatial attention when performing in a higher flow state. The present study will help provide a more comprehensive explanation of modulated temporal processing and spatial attention during flow states with intra-individual differences in consideration.

2. MATERIALS AND METHODS

2.1. Participants

Eleven athletes (mean age = 23.6, SD = 3.53, 4 female and 7 male) and 16 musicians (mean age = 20.8, SD = 3, 10 female and 6 male) from various sports and musical disciplines were recruited. One additional subject (athlete) was used for piloting and excluded from the analyses due to irregularities in the testing procedure and refinement of the experiment (e.g., increasing the number of repetitions of the TOJ task). The athletes had 14 (SD = 5.87) years of experience and practiced 18 (SD = 6.83) hours per week on average. The musicians had 8.8 (SD = 2.94) years of experience and practiced 8.5 (SD = 6.33) hours per week on average. Due

to previous findings suggesting that skill level is correlated with the experience of flow (Catley and Duda, 1997; Engeser and Rheinberg, 2008), expertise was operationally defined as regular practice over several years in a particular discipline. Additionally, all athletes and musicians currently compete or perform at exceptional levels (e.g., NCAA Division II tennis players; performing musicians, etc.). The years of experience between athletes and musicians were significantly different, $t_{(13)} = 2.72$, $p = 0.017$, as well as the weekly practice hours, $t_{(20)} = 3.65$, $p = 0.002$. Altogether, 27 trained musicians and athletes (mean age = 21.9, $SD = 3.47$, 14 female and 13 male) participated. To cover a more general picture about flow across expertise types, a diverse sample of athletes (2 runners, 9 tennis players) and musician types (1 piano, 2 trumpets, 4 flutes, 2 clarinets, 2 saxophones, 1 bassoon, 1 oboe, 1 trombone, 1 tuba, 1 percussion) participated. The study was approved by the University of Hawaii at Manoa's committee on human subjects (CHS). All participants provided written informed consent before beginning the study. In order to compensate for their time, participants were offered the opportunity to participate in a mental preparation seminar held by one of the authors (RAP). Due to drop outs, altogether five experimental runs of two participants are missing.

2.2. Task

The experimental task consisted of a temporal order judgment task (TOJ), adapted from Lim and Sinnett (2011), designed to measure temporal processing and spatial attention. Two versions with different conditions of the TOJ task were presented in separate blocks, one with exogenous cues and the other with endogenous cues (**Figure 1**). The TOJ tasks were presented successively in separate sessions and counterbalanced blocks. Approximately half of the participants started with the endogenous condition. Each trial started with a fixation cross in the middle of the screen flanked by two placeholder squares. The length between the outer ends of the placeholder squares to the fixation cross was 5.4 cm. After 1,000 ms, either an exogenous or an endogenous cue was displayed for 45 ms. Exogenous cues were created by thickening placeholder squares to 4 pixels, whereas endogenous cues consisted of a central arrow (1.2 cm). Following the appearance of the cue with a delay of 45 ms, the first target (horizontal or vertical line) was displayed in either

the left or right placeholder square. The other target appeared in the other placeholder square after a specified stimulus onset asynchrony (SOA). Target orientation and appearance side were presented with the same probability of occurrence. The targets (1.2 cm) appeared within the placeholder squares (1.6 × 1.6 cm). Participants were then forced to make a choice on the keyboard to indicate on which side they perceived the target to appear first. To determine the SOAs for each trial, a 1-up-3-down adaptive staircase approach (Cornsweet, 1962) was used. Each block started with a SOA of 267 ms. Depending on whether 3 correct or one incorrect response was given, the SOA would respectively decrease or increase by 16.7 ms on the subsequent trial. Each block was finished after 14 reversals. Altogether, the TOJ task lasted ~7–10 min. The experiment was programmed and run using the software PsychoPy 3.0 (Peirce et al., 2019).

2.3. Questionnaire

To measure flow states, the Activity Flow State Scale (AFSS) (Payne et al., 2011a) was used. The AFSS captures the 9 dimensions of flow [Merging actions and awareness (MAA); Clear goals (CG); Concentration on task at hand (CO); Unambiguous feedback (UF); Challenge skill balance (CS); Transformation of time (TT); Sense of control (CN); Loss of self-consciousness (SC); Autotelic experience (AE)] according to Csikszentmihalyi (2000, 2002) with 26 statements. It has a high reliability with Cronbach's alpha coefficients for the 9 subscales ranging from 0.71 to 0.90 (Payne et al., 2011a). It has been shown to measure flow in different populations and was specifically constructed to measure flow across a wide range of activities (Payne et al., 2011b; Osin et al., 2016). The items are rated on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). A global flow score (the mean of all items) was computed for each participant.

2.4. Design

The experiment was divided into three sessions conducted ~2–5 days apart (**Figure 2**). In each session, all participants completed a pre-TOJ task, including both endogenous and exogenous conditions, a sports or music performance (practice or rehearsal sessions), a post-TOJ task, again including endogenous and

FIGURE 1 | (A) Endogenous cues are indicated by an arrow in the middle and **(B)** exogenous cues are indicated by a thick frame. Each trial started with a fixation cross followed by a cue, a cue-target interval followed by the appearance of the first target and a specific stimulus onset asynchrony (SOA) interval followed by the second target.

FIGURE 2 | The within-subjects design was divided into three repeated sessions with 2–5 days intervals. One session run is shown as an example: All participants completed a pre-performance TOJ task (including endogenous and exogenous trials), a sports or music practice or rehearsal session, a post-performance TOJ task (including endogenous and exogenous blocks, which were presented in a counterbalanced order between the participants and sessions), a control question and a self-report questionnaire (AFSS, displayed in randomized order).

exogenous conditions, a control question and the questionnaire (AFSS) at the end of each session.

2.5. Procedure

Participants were recruited according to their level of expertise. All sessions took place next to the participants' practice environment to ensure high ecological validity, and were managed by at least two researchers at a time. In the first session, participants were asked to provide informed consent, and then filled out a questionnaire collecting demographic information, including information about years of experience in their specific discipline and weekly practice amounts. The participants were then seated $\sim$ 60 cm from the monitor. The investigation was set up in a way that a maximum of six people could participate at the same time. Each session started with a written introduction for the upcoming task and a reminder to complete the task as quickly and accurately as possible. As soon as the participants were familiar with the task and felt ready, they began with the pre TOJ task by pressing the space bar on the keyboard. The task was divided into 2 blocks with a short break in between. Each block included either the exogenous or the endogenous condition, each lasting $\sim$ 3–5 min. After the TOJ task, the participants completed their sports or music performance (practice or rehearsal sessions), which lasted between 30 min to 2 h. The duration of the practice or rehearsal sessions depended on the usual practice schedule of the participants. Immediately after the practice or rehearsal session, the participants completed the post-TOJ task (same procedure as the pre-TOJ task) followed by the self-report questionnaire (AFSS) about their flow state during the sports or music session, altogether lasting around 10 min. Prior to beginning the AFSS, a control question ranging from 1 (strongly disagree) to 5 (strongly agree) was inserted in order to get an impression if participant's state, emotions, and body feelings were about the same during the practice or rehearsal session and the post-TOJ task, and intended to gauge whether the participants still experienced flow while doing the post-TOJ task. The task and the questionnaire were presented on 6 different 13.3" laptop computers each with a refresh rate of 60 Hz.

2.6. Analysis

To carry out the following reported analyses of this study, we used Microsoft Excel 2019 (Microsoft Corporation, 2019) and the statistical software R version 3.5.3 (R Core Team, 2019). The logistic regressions were performed using the R package "quickpsy" (Linares and López i Moliner, 2016). To fit the HLMs the R package "lme4" was used (Bates et al., 2015). To fit the bayesian HLMs the R package "brms" was used (Bürkner, 2016). The brms package implements bayesian HLMs in R using the probabilistic programming language "Stan" (Carpenter et al., 2017) under the hood, with an lme4-like syntax. Sorensen and Vasishth (2015) provide a detailed and accessible introduction to bayesian HLMs applied to cognitive science using Stan. The "Loo" package (Vehtari et al., 2017) was used to compare the bayesian models.

To estimate the JNDs and PSSs for each participant, we used a similar procedure as Lim and Sinnett (2011). At first, data from each participant was separated into endogenous cued trials, endogenous left or right cued trials, exogenous cued trials, and exogenous left or right cued trials. A logistic regression model was then fitted to each cue type for each participant's run, resulting in 36 models per participant, namely 12 JND scores (3 endo pre, 3 exo pre, 3 endo post and 3 exo post) and 24 PSS scores (3 endo pre right, 3 endo pre left, 3 exo pre right, 3 exo pre left, 3 endo post right, 3 endo post left, 3 exo post right and 3 exo post left). Two measures were then calculated for each participant's run. First, the JNDs were calculated independently for the endogenously and exogenously cued trials by taking the SOAs corresponding to 0.75 and 0.25 proportions, and then halving the distance between these SOAs (**Figure 3**). This halved distance is the specific JND for each condition (Endogenous/Exogenous) within participants for testing (pre/post), and session (1, 2, 3). Second, the PSSs were taken from the models of the right and left cued endogenous and exogenous trials for the SOAs corresponding to 0.50 proportion for right first responses (**Figure 4**). This is the point of maximal uncertainty, representing chance performance, and statistically, the PSS. To compare the PSS scores between the pre- and post-tasks, the distances between the left and right PSS separate

FIGURE 3 | Proportion responding right first as a logistic regression of stimulus onset asynchrony (SOA) for one participant as an example. JND measurements were taken for the endogenous (red) and exogenous (blue) cued items by halving the distance between the SOAs corresponding to 0.75 and 0.25 proportions (see gray segments for endogenous cued items and black segments for exogenous cued items).

FIGURE 4 | Proportion responding right first as a logistic regression of stimulus onset asynchrony (SOA) for one participant as an example. PSS measurements were taken for the endogenous right cued items (red), endogenous left cued items (dark red), exogenous right cued items (blue), and exogenous left cued items (dark blue) for the SOAs corresponding to 0.50 proportion (see gray segments for endogenous cued items and black segments for exogenous cued items). Notable is the larger gap between right cued items and left cued items for exogenous trials compared to endogenous (reflecting larger distraction by peripheral cues).

for endogenous and exogenous pre and post measurements were calculated.

All estimated JNDs outside a range of 2.5 of the absolute deviations around the median (Leys et al., 2013) were excluded, due to response error rates and non-convergence of the algorithm fit, resulting in 39 data points being excluded (~12.4%). With regards to the PSS, the distances between the left and right cued PSSs of the pre task were compared to the distances between the left and right cued PSSs of the post task, separated by session. All responses outside a range of 2.5 absolute deviations around the median (Leys et al., 2013) of the PSS pre-post difference were excluded, due to response error rates and non-convergence of the algorithm fit, resulting in the exclusion of 37 data points (~11.8%).

In order to test the pre-registered hypothesis of whether flow has an influence on the JND, hierarchical linear models (HLMs) were calculated. These models estimate the influence of flow on the change (moderation) of the JND from pre- to post-testing. HLMs account for data hierarchies as observed in repeated measurement designs. In our study, repeated points of measurement are nested within sessions and within participants, and this data dependency can be recognized by HLMs.

In order to test whether the mean values of the nested groups differ and whether the differences justify a three-level model structure, a fully unconditional model was created. To test if there is a general difference in the JND from the pre- to post-testing a predictor was added to the model to estimate this relationship. Additionally, we included condition (endogenous vs. exogenous)
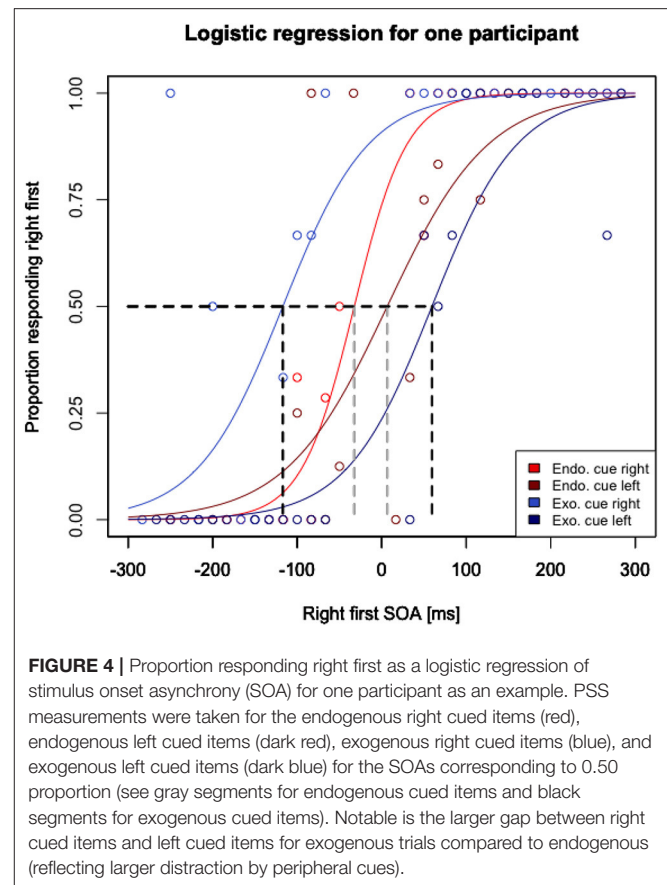
as a predictor in the model to test the general contrast between these two conditions. Lastly, the interaction of the difference from pre- to post-testing and flow as a predictor was included in the model. A significant fixed interaction effect between flow and the pre- to post-testing TOJ would indicate a moderation of the change in the JND by flow from the pre- to the post-testing, and therefore a relationship between flow and the JND.

Stegmueller (2013) have shown that frequentistic approaches to modeling HLMs are sometimes susceptible to relatively small sample sizes, whereas bayesian probabilistic models appear to be more robust. In particular, the bayesian probabilistic approach shows considerably better properties with regard to the estimation of confidence intervals. Therefore, due to our relatively small sample size, for the main analysis of the dependent variable JND, an additional probabilistic model, with a bayesian point of view, was estimated to support the results. An interaction term of pre- to post-testing and flow which does not include zero in its 95% credible interval would indicate an effect of flow on the JND.

2.6.1. Exploratory Analyses

In order to investigate whether the PSSs are influenced by the amount of experienced flow, HLMs for the dependent variable PSS were estimated in the same way as for the dependent variable JND. In order to test whether the JNDs and PSSs within the

TABLE 1 | Number of items, mean value, standard deviation, and Cronbach's α for the nine subscales of the AFSS.

Subscale	No. items	*M*	*S.D.*	Cronbach's α
MAA	3	3.19	0.96	0.67
CG	3	3.68	1.14	0.91
CO	4	3.41	1.04	0.83
UF	2	3.47	1.06	0.77
CS	3	3.57	0.87	0.69
TT	3	3.24	0.96	0.68
CN	2	3.46	1.21	0.86
SC	3	3.49	1.14	0.83
AE	3	3.42	0.97	0.80

n = 78; MAA, merging actions and awareness; CG, clear goals; CO, concentration on task at hand; UF, unambiguous feedback; CS, challenge skill balance; TT, transformation of time; CN, sense of control; SC, Loss of self-consciousness; AE, autotelic experience.

TABLE 2 | Estimated coefficients of the JND-model.

Predictors	Estimate	Std. Error	CI	*P*
(Intercept)	53.30	4.31	44.85 to 61.74	<0.001
Condition	10.42	2.26	−9.83 to −0.89	<0.001
Flow	2.84	3.44	−3.91 to 9.59	0.409
Prepost	−5.36	2.28	5.99 to 14.85	0.019
Prepost * Flow	-5.80	3.02	−11.71 to 0.12	0.055
Random effects				
Var: session:subject	152.93			
Var: subject	331.62			
σ^2	343.18			
Num. obs.	274			
Num. groups: session: subject	77			
Num. groups: subject	27			

CI, 95% Confidence Interval; Flow, Mean-centered Values.

musicians differ from those of the athletes, a predictor that estimates this distinction was calculated. To test whether the relationship between the JND or PSS and flow is dependent on the response of the control question, an additional interaction term that estimates this relationship was built into the models.

3. RESULTS

Across all three sessions, the mean of the global flow score as reported on the AFSS was 3.43 (out of 5), ranging between 1.58 and 4.85, suggesting that our participants did experience a broad variance of flow levels. The Cronbach's alpha coefficient for the AFSS was, $\alpha = 0.95$. The mean values, standard deviations and Cronbach's alpha for the nine subscales can be found in **Table 1**. The mean answer on the 5-point Likert scale to the control question [i.e., "My state (body feelings, emotions and thoughts) during the task on the computer was similar to the state I had during my music/sports performance."] was 3.33 ($SD = 1.09$), suggesting that participants tended to affirm the statement.

3.1. Analysis of the JND

A model that estimates the explained variance by the nested groups showed that the subject effect (level 3) was responsible for 15% of the explained variance, and 38% of the JNDs variance was explained by session effects (level 2). This suggests the necessity of using a three-level model structure (pre-/post-testing nested in sessions within subjects).

When including pre-/post-testing of the JND as a fixed coefficient in the model, representing the estimated general difference between the measured JNDs before the practice or rehearsal session and after the practice or rehearsal session, the model's fit was significantly improved compared with the fully unconditional model, $X^2(1) = 4.82, p = 0.03$. When including the factor pre-/post-testing as a random coefficient in the model (M1.1), there was no significant improvement in the model fit, $X^2(4) = 6.72, p = 0.15$.

Including the discrimination of the conditions exogenous and endogenous as an additional fixed coefficient to our model to test

for possible differences between these two conditions, the model's fit improved significantly, $X^2(1) = 19.66, p < 0.001$. There was no significant improvement in the model fit, $X^2(3) = 7.73, p = 0.052$, when including this predictor as a random coefficient.

To test whether flow is a moderator for the change of the JND between the pre- and the post-testing, an interaction term of the variable pre/post and flow was added as a fixed coefficient to the model (the coefficients of the model can be seen in **Table 2**). The interaction between the pre-/post-testing and flow amounts to −5.8 ms, $SE = 3.02$, 95% CI [−11.71 to 0.12], $t_{(206)} = −1.92, p = 0.055$. The model fit did not significantly improve as we added the interaction as a fixed coefficient, $X^2(2) = 3.64, p = 0.16$.

The estimated interaction of pre/post, flow and the condition (exogenous/endogenous) was not significant ($p = 0.27$), suggesting that there is no difference between the interaction of the pre-/post-testing and flow for the endogenous condition when compared with the exogenous task.

In regards to the JND, the interaction between pre/post and flow does not depend on the answer of the control question ($p = 0.85$). Thus, the connection between the values of the pre-/post-testing and flow is not stronger if the control question is answered higher.

3.1.1. Probabilistic Modeling of the JND
The estimated interaction coefficient of Flow and the difference from pre- to post-testing within the bayesian model is −5.75 ms, $SE = 3.04$, 95% Credible Interval [−11.8 to −0.02]. The 95% credible interval can be interpreted in a way that there is a 0.95 probability that the value of the intercept lies between −11.82 and −0.02 ms. This effect is visualized in **Figure 5**. Low values of flow indicate a negative difference between the pre- and the post-testing and a high value of flow shows a positive difference between the pre- and the post-testing. It should be noted that the probabilistic estimate differs from the frequentistic estimate in terms of whether zero should be included in the confidence or the credible interval, respectively. The results of the estimates therefore do not seem to agree entirely on the extent to which the effect can be perceived as significant. Furthermore, a model

FIGURE 5 | Estimated JND scores and their 95% credible intervals as a function of flow, divided into pre- and post-measured values.

TABLE 3 | Estimated coefficients of the PSS-model.

Predictors	Estimate	Std. Error	CI	P
(Intercept)	41.12	5.36	30.61 to 51.62	<0.001
Condition	64.12	9.18	46.13 to 82.111	<0.001
Flow	12.21	5.79	0.86 to 23.57	0.035
Prepost	6.84	5.52	−3.98 to 17.65	0.215
Prepost * Flow	−21.31	7.28	−35.58 to −7.05	0.003
Random Effects				
Var: subject (Intercept)	153.05			
Var: subject Condition	1427.08			
Cov: subject (Intercept) Condition	373.90			
σ^2	2053.72			
Num. obs.	276			
Num. groups: subject	27			

CI, 95% Confidence Interval; Flow, Mean-centered Values.

which includes the interaction term of pre- to post-testing and flow differs only slightly from a model which does not include this predictor in terms of its predictive accuracy (elpd_diff = −0.2), and not larger than the standard error of these estimations (se_diff = 0.3). This indicates that a model which includes the interaction term of pre- to post-testing and flow does not make a substantial contribution to the prediction of new data.

3.2. Analysis of the PSS

A model that estimates the explained variance by the nested groups shows that 13.1% of the variance of the PSS was explained by participants. This indicates the necessity of a two-level-structure (pre-/post-testing nested within subjects). A three-level-structure (nesting the values of the PSS in sessions within participants) leads to an over-fitting of the estimated model, and therefore was not considered.
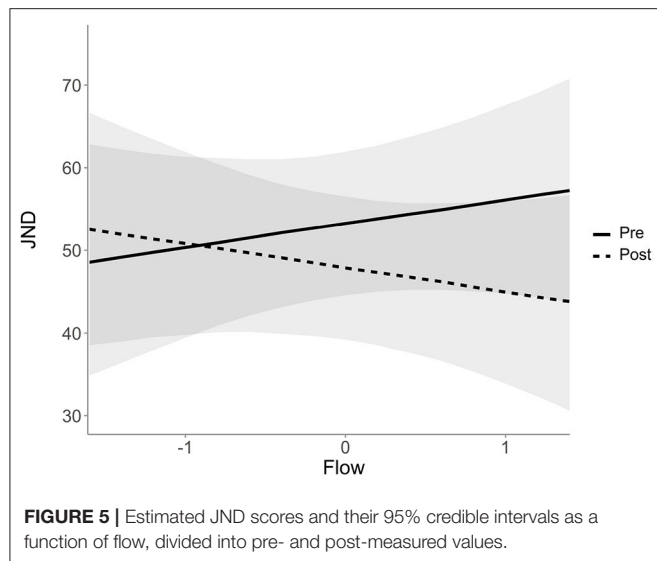
To test whether flow is a moderator for the change of the PSS between pre- and post-testing, an interaction term of the variable pre/post and flow was added to the model. As can be seen in **Table 3**, the estimation shows that a one unit change in the value of "flow" moderates the change from the pre- to the post-testing by −21.3 ms, $SE = 7.28$, 95% CI [−35.58 to −7.05], $t_{(227)} = −2.93$, $p < 0.01$. Essentially, this means that for every unit increase/decrease in experienced flow the PSS improved/diminished by 21 ms, respectively, indicating improved control of spatial attention with increased flow. This model also significantly accounted for the explained variance compared to a model without this interaction, $X^2(3) = 9.82$, $p = 0.02$. Adding the effect of the interaction as a random effect leads to an over-fitting of the data.

Probabilistic modeling of the data supports these results and estimates a coefficient of −20.9 ms, $SE = 7.34$, 95% CI [−35.06 to −6.41], for the interaction of flow and the change from pre- to post-testing. **Figure 6** shows the relationship from pre to post PSS changes depending on flow. Lower flow values are associated

FIGURE 6 | Estimated PSS scores and their 95% credible intervals as a function of flow, divided into pre- and post-measured values.

with an increase in PSS values and high flow values are associated with a decrease in PSS values.

Including the interaction of pre/post, flow, and condition (exogenous/endogenous) to test if the moderation depends on the condition, there is neither an improvement of the explained variance [$X^2(3) = 1.32$, $p = 0.72$] nor a significant three-way interaction [−13.93, $SE = 14.50$, 95% CI [−42.35 to 14.50], $t_{(226)} = −0.96$, $p = 0.34$]. This indicates that the moderation of the PSS by flow remains the same under both endogenous and exogenous conditions.

In regards to the PSS, the interaction between pre/post and flow does not depend on the answer of the control question ($p = 0.98$), suggesting that the connection between the values of the pre-/post-testing and flow is not stronger if the control question is answered higher.

3.3. Effects of the Domain on the JND and the PSS

When directly comparing the performance between musicians and athletes (Domain), there are no significant differences in the JND scores ($p = 0.53$). However, the rate of change from the pre- to the post-testing seems to differ significantly by 11.3 ms, $SE = 4.82$, 95% CI [1.89–20.79], $t_{(202)} = 2.35$, $p = 0.020$. Athletes have thus a 11.3 ms larger change from the pre- to the post-measurement compared to musicians. Regarding the PSS, no difference between musicians and athletes in general ($p = 0.83$) nor in their rate of change from pre- to post-testing ($p = 0.57$) was observed.

4. DISCUSSION

Phenomenological differences in perception while performing a demanding task have been reported on several occasions. To the best of our knowledge, this is the first study to address the question of whether increased flow experience can improve perception, with the aim to investigate the relationship between modulated temporal and spatial visual processing at different time points and flow states in a within-subjects design. Of critical importance, the participants in this experiment participated over several experimental sessions, thereby enabling the measurement of a range of self-perceived flow states from which we can assess whether increased self-reported flow correlated with improved spatial and temporal processing, as measured by the TOJ task. This controls for the possibility that previous research addressing similar questions simply measured perceptual performance between more skilled and lesser skilled participants: Wesp et al. (2004) demonstrated that accomplished darts players perceived the target to be larger than novices with lower abilities, Witt and Proffitt (2005) suggested that softball players who were successful at hitting recalled the ball to be bigger than players with less success, Lee et al. (2012) claimed that people who perform better in archery see the target as larger than weaker archers, and Witt et al. (2008) show that golfers with high performance perceived the size of the cup to be bigger compared to lower performing golfers. Furthermore, possible short-term modulations in temporal perception like the speed of the ball in baseball (Gray, 2013) or in tennis (Witt and Sugovic, 2010) was perceived to be slower among better players. Nevertheless, these studies failed to take into account the intra-individual differences of the participants, therefore it is difficult to claim that flow states (i.e., when a participant is performing optimally) does lead to improved perception. In the present study, intra-individual differences in flow state were repeatedly measured, with our findings suggesting a flow state dependent correlation instead of a person dependent correlation.

The results of the present pre-registered study indicate a relationship between the value of experienced flow and spatiotemporal information processing. The improved performance was manifested in improved temporal perception (i.e., reduced JND scores) and significantly improved spatial attentional control (i.e., reduced PSS scores) when flow states were highest. In a recent longitudinal experiment by Cowley et al. (2019), where flow was induced with the help of a video game-like high-speed steering task, possible trial-wise fluctuations of performance due to flow were found, suggesting that performance was enhanced when participants experienced increased flow. These results dovetail with our findings, and might likewise indicate a short-term modulation of information processing and performance by flow.

4.1. Temporal Processing and Flow

The results indicate a positive correlation between the value of experienced flow and temporal processing. Specifically, when participants reported higher flow states in the practice or rehearsal session, they also performed better in the post-performance TOJ task compared with the pre-performance TOJ task, approximately by 5.8 ms per one-unit change of flow. But the results cannot draw a clear picture of the significance of the effect of flow on the JND. Although probabilistic models demonstrated that a coefficient of zero is not in their credible interval, the effect does not seem to contribute significantly in terms of predicting new data. In addition, the frequentistic model, although clearly pointing in one direction of supporting the alternative hypothesis, did not show a significant effect and the interaction term did not improve the fit of the model. Due to the way the data was collected (in the field), it can be assumed that a high level of noise in the data contributed to wide confidence intervals (and slightly smaller credible intervals), which makes it difficult to provide a reliable statement about the effect of flow on the JND and its strength.

On average, across all participants and independent from reported flow levels, the JND from pre- to post-testing generally improved (decreases) by approximately 5 ms. Despite there being no significant difference found overall when comparing the JND between athletes and musicians, there was a difference in the general rate of change from pre- to post-testing between the two domains. Specifically, athletes' JND scores deteriorated by 11 ms compared to musicians. This means that the musicians have improved JND scores while the athletes' performance has weakened. The differences between the rate of change of the two domains may be explained by the fact that athletes experience increased fatigue (compared to musicians) due to their physical performance while musicians might benefit more from learning effects and increased concentration, or, in contrast to the athletes, are not fatigued. However, these assumptions are purely speculative and the underpinning reason for these differences should be explored in further research. These differences were not observed in relation to the PSS.

4.2. Spatial Attention and Flow

In regards to spatial attention, it appears that increased flow experiences resulted in significant improvements in the PSS. More precisely, PSS improved by approximately 21 ms for each one unit increase in reported flow, suggesting that spatial attention was indeed modulated during flow experiences. It is important to note that this improvement could in fact be underestimated given that performance specific demands on spatial attention are arguably less for musicians than

athletes (although, musicians are required to divide attention between their own performance and musical notes, and also the conductor and other fellow musicians). Future research could focus specifically on how flow experiences might affect attentional/perceptual mechanisms that are more tightly coupled with differing music/sport specific demands. Similar to Lim and Sinnett (2011) larger overall PSS scores (64ms difference) were observed for exogenous than for endogenous trials, which demonstrates that exogenous stimuli have a greater impact on attention. With direct respect to the perception of central objects (i.e., endogenous trials), it is likely that attentional focus is increased during flow experiences. This effect is reflected by participants' PSS scores, which indicate that a high level of flow reduces the distractibility from the central cues, resulting in a lower PSS.

4.3. Theoretical Implications

Attention plays a critical role in how time is experienced. Consider how a pleasant event (e.g., an excellent film or book) makes the subjective time seem to pass more quickly compared to something less pleasant (e.g., a boring lecture or perhaps this article for some). That is, the perception of time can be modulated depending on one's focus during the task (Phillips, 2012; Wearden, 2015). Arguably, the basis for time perception is rooted in the ability to process temporal order, with temporal resolution allowing for the successful recognition of stimulus order or simultaneity (Thönes and Stocker, 2019). Yet, the question remains as to how exactly the phenomenological perception of the passage of time during flow states and the processing of temporal information are related.

As a theoretical framework for the subjectively perceived slow down in time perception, Tse et al. (2004) might offer an alternative. These authors found that during unexpected events attention is arguably highly engaged, potentially leading to an increase in the amount of information processed during that time and to the subjectively experienced expansion of time. The authors conducted an oddball paradigm, in which participants were required to respond to a low-probability target that occurred within a range of high-probability stimuli. Participants were required to judge the duration of the presented low-probability target and decide whether it lasted longer or shorter than the standard items. The authors found that unexpected stimuli that were in fact presented for a shorter amount of time than the standard stimuli were, in fact, judged to be presented for the same length, suggesting that the engagement of attention for unexpected stimuli potentially leads to an increase in the amount of processed information, and subjectively perceived time. Essentially, more information is extracted from an unexpected signal. Arstila (2012) further posit why such faster rates of information processing lead to the experience of time as moving slower than usual. Specifically, these authors suggest that one component of the perception of time is determined by the speed of things in the external world. This re-afferent system plays a crucial role in determining the time that we perceive. The experience of faster acting than usual also implies that external objects might be slower than usual. Applying this logic to our study would suggest that participants who experienced a high

flow state (in which attention plays a crucial role), processed more temporal information during the post TOJ task than in the pre task, and therefore, their JND scores improved (i.e., reduced). Moreover, we might realize that we are able to shift our attention from one stimulus to another more quickly than usual and therefore, our re-afferent system provides us with information that our internal processes are faster than normal.

One possible alternative explanation for our results could simply be that participants were subject to learning effects because of the repeated measures design. This learning effect could manifest itself within a session, between the pre- and the post-testing or across subsequent sessions. The change between the sessions is embedded in the hierarchical structure of the model, which accounts for the variance of this random effect. While a general modulation of JND does occur over sessions, these same differences were not observed for PSS. Moreover, it should be noted that the main interest of this study was not in the general effect of a single performance on the JND, which would be reflected in the change from pre- to post-testing, but whether this change is moderated by flow, for which the data suggest to be the case.

5. LIMITATIONS AND FUTURE RESEARCH

At this point it is impossible to make any statements regarding the direction of the causality of whether flow leads to improved perception, or if instead improved perception leads to flow. Thus, the question remains unanswered as to whether participants who experienced flow during the practice or rehearsal session have therefore an enhanced information processing, or if participants who show an enhanced information processing therefore experience higher flow states. While this concern should be partially alleviated given that our findings do imply that perception improves as flow increases *within* our participants, this is a question that should be elucidated by further research. Randomized controlled trials and experiments in which flow can be actively manipulated could provide insights into the causality relationships between flow states and spatiotemporal processing. Shehata et al. (2018) present a possibility to actively manipulate flow states and to induce them with the help of a "Music Rhythm Game," which arguably generates flow over the difficulty of the game (from "boredom" to "overload"). The difficulty in this is, of course, whether flow states caused by simple but controllable environments are similar or not to those experienced in self-chosen multifaceted sports and music environments. Another limitation that could be tackled with such a task, might be that the flow inducing sports or music performance was between the TOJ measurements and not at the same time.

An additional aspect of the current study that could be expanded in future research has to do with the inclusion of both athletes and musicians who play different sports and instruments, respectively. While this did result in differences between groups in terms of years of experience and training regimens, it should be noted that when including group as a factor no major differences were observed in the findings. We included different activities (i.e., athletes and musicians) in an attempt to more broadly

explore how flow experiences from varying performance types might affect perception. Future research could consider exploring specific activities (e.g., only tennis players or only pianists) in order to see if perception is differently modulated based on the flow that is experienced from different activities.

Despite the inclusion of a control question to determine whether the state was approximately the same during the practice/rehearsal session and post task, and although the participants tended to agree with the control question, it is difficult to conclude that the flow-state experience was truly transferred to the post TOJ task. It was expected that there would be a stronger correlation between flow and pre- and post-measurement performance when the control question was answered higher. However, this was not the case. This could possibly be explained by the fact that the control question might not adequately assess whether the flow state of the participant extended to the TOJ task. In addition, although we attempted to keep the amount of time between music/sport session and testing as short as possible, memory biases could have influenced our results in that we presented the control question and the flow questionnaire at the end of the experimental trial and the items were answered retrospectively. In future investigations it would certainly be interesting to develop a task where flow and perception can be simultaneously measured, so as to completely avoid any potential memory confound. This could potentially be accomplished by using physiological measures, such as skin conductance, pulse rate (Tozman et al., 2015), brain activity or oculomotor indicators (Shehata et al., 2018). Knierim et al. (2018) provide an overview of peripheral nervous system indicators of flow. Furthermore, aspects that, in addition to the flow state, could also lead to an improvement in spatiotemporal processing, such as increased alertness after enhanced performance, should be controlled in further studies.

Due to the testing that took place in the field, the data are subject to high noise exposure, therefore the calculated models have difficulties in estimating accurate confidence/credible intervals. This limitation is notable as this potentially impedes an accurate determination of the effect size. For instance, and with particular regard to the pre-registered hypothesis about the moderation of the JND, no statistically significant and unambiguous answer could be found.

Lastly, even if the results regarding the modeling of the PSS may appear promising with regard to the understanding of the possibility that perception is modulated during flow states, the results were obtained in form of an exploratory analysis and should be confirmed in the sense of a pre-registered confirmatory analysis.

6. CONCLUSION

To the best of our knowledge the present study is the first to provide evidence that subjectively experienced improvements in

perception during flow states are related to improved temporal and spatial visual processing. In this study, self-reported flow states and perceptual processing, as measured by a TOJ task, were obtained over several time points. The correlation between flow and the JND scores indicate an improved temporal processing during flow, with the correlation between flow and the PSS scores indicating significantly enhanced spatial attention. Combined, these findings provide evidence to suggest that anecdotal accounts of improved perception during flow states might actually reflect objective reality.

DATA AVAILABILITY STATEMENT

This study was pre-registered, in which the general procedure, the hypothesis for the confirmatory analysis about the relationship between flow and the JND, and the exploratory analysis about the relationship between flow and the PSS were documented. The pre-registration, "Transparent Changes" document, all generated data sets, experimental code, analyses and additional analyses are publicly available via the Open Science Framework: https://osf.io/xfrmu/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Committee on Human Subjects, University of Hawaii at Manoa. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

REFERENCES

Arstila, V. (2012). Time slows down during accidents. *Front. Psychol.* 3:196. doi: 10.3389/fpsyg.2012.00196

Bates, D., Mächler, 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

Bürkner, P.-C. (2016). brms: an r package for bayesian generalized linear mixed models using stan. *J Stat Softw*.

Cañal-Bruland, R., Pijpers, J. R., and Oudejans, R. R. D. (2010). The influence of anxiety on action-specific perception. *Anxiety Stress Coping* 23, 353–361. doi: 10.1080/10615800903447588

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: a probabilistic programming language. *J. Stat. Softw.* 76, 1–32. doi: 10.18637/jss.v076.i01

Catley, D., and Duda, J. L. (1997). Psychological antecedents of the frequency and intensity of flow in golfers. *Int J Sport Psychol.* 28, 309–322.

Chirico, A., Serino, S., Cipresso, P., Gaggioli, A., and Riva, G. (2015). When music "flows". state and trait in musical performance, composition and listening: a systematic review. *Front. Psychol.* 6:906. doi: 10.3389/fpsyg.2015.00906

Cooper, A. D., Sterling, C. P., Bacon, M. P., and Bridgeman, B. (2012). Does action affect perception or memory? *Vision Res.* 62, 235–240. doi: 10.1016/j.visres.2012.04.009

Cornsweet, T. N. (1962). The staircrase-method in psychophysics. *Am J. Psychol.* 75, 485–491.

Cowley, B. U., Palomäki, J., Tammi, T., Frantsi, R., Inkilä, V.-P., Lehtonen, N., et al. (2019). Flow experiences during visuomotor skill acquisition reflect deviation from a power-law learning curve, but not overall level of skill. *Front. Psychol.* 10:1126. doi: 10.3389/fpsyg.2019.01126

Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *J. Hum. Psychol.* 15, 41–63.

Csikszentmihalyi, M. (2000). *The Contribution of Flow to Positive Psychology*. Philadelphia, PA: Templeton Foundation Press.

Csikszentmihalyi, M. (2002). *Flow: The Psychology of Optimal Experience, 2nd Edn*. New York, NY: Harper and Row.

Engeser, S., and Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motiv. Emot.* 32, 158–172. doi: 10.1007/s11031-008-9102-4

Failing, M., and Theeuwes, J. (2016). Reward alters the perception of time. *Cognition* 148, 19–26. doi: 10.1016/j.cognition.2015.12.005

Gong, D., He, H., Ma, W., Liu, D., Huang, M., Dong, L., et al. (2016). Functional integration between salience and central executive networks: a role for action video game experience. *Neural Plasticity* 2016:9803165. doi: 10.1155/2016/9803165

Granek, J. A., Gorbet, D. J., and Sergio, L. E. (2010). Extensive video-game experience alters cortical networks for complex visuomotor transformations. *Cortex* 46, 1165–1177. doi: 10.1016/j.cortex.2009.10.009

Gray, R. (2013). Being selective at the plate: processing dependence between perceptual variables relates to hitting goals and performance. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 1124–1142. doi: 10.1037/a0030729

Habe, K., Biasutti, M., and Kajtna, T. (2019). Flow and satisfaction with life in élite musicians and top athletes. *Front. Psychol.* 10:698. doi: 10.3389/fpsyg.2019.00698

Jackson, S. A. (1995). Factors influencing the occurrence of flow state in elite athletes. *J. Appl. Sport Psychol.* 7, 138–166.

Jackson, S. A., and Csikszentmihalyi, M. (1999). *Flow in Sports*. Champaign, IL: Human Kinetics.

Knierim, M. T., Rissler, R., Dorner, V., Maedche, A., and Weinhardt, C. (2018). "The psychophysiology of flow: a systematic review of peripheral nervous system features," in *Information Systems and Neuroscience*, eds F. Davis, R. Riedl, J. vom Brocke, P. M. Léger, and A. Randolph (Springer), 109–120. doi: 10.1007/978-3-319-67431-_13

Lee, Y., Lee, S., Carello, C., and Turvey, M. (2012). An archer's perceived form scales the "hitableness" of archery targets. *J. Exp. Psychol. Hum. Percept. Perform.* 38:1125. doi: 10.1037/a0029036

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 764–766. doi: 10.1016/j.jesp.2013.03.013

Lim, A., and Sinnett, S. (2011). "Exploring visual attention in musicians: temporal, spatial and capacity considerations," in *Proceedings of the Annual Meeting of the Cognitive Science Society* (Merced, CA), 33.

Linares, D., and López i Moliner, J. (2016). quickpsy: an r package to fit psychometric functions for multiple groups. *R J.* 8, 122–131. doi: 10.32614/rj-2016-008

Microsoft Corporation (2019). *Microsoft Excel*.

Moneta, G. B. (2012). "On the measurement and conceptualization of flow," in *Advances in Flow Research*, ed S. Engeser (New York, NY: Springer), 23–50.

Osin, E. N., Malyutina, A. V., and Kosheleva, N. V. (2016). Self-transcendence facilitates meaning-making and flow: evidence from a pilot experimental study. *Psychology Russia State Art* 9, 80–96. doi: 10.11621/pir.2016.0207

Payne, B. R., Jackson, J. J., Noh, S. R., and Stine-Morrow, E. A. (2011a). Activity flow state scale. *PsycTESTS* 10.

Payne, B. R., Jackson, J. J., Noh, S. R., and Stine-Morrow, E. A. (2011b). In the zone: flow state and cognition in older adults. *Psychol. Aging* 26, 738. doi: 10.1037/a0022359

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* 51, 195–203. doi: 10.3758/s13428-018-01193-y

Phillips, I. (2012). Attention to the passage of time. *Philos. Perspect.* 26, 277–308. doi: 10.1111/phpe.12007

Proffitt, D. R., Stefanucci, J., Banton, T., and Epstein, W. (2003). The role of effort in perceiving distance. *Psychol. Sci.* 14, 106–112. doi: 10.1111/1467-9280.t01-1-01427

Qiu, N., Ma, W., Fan, X., Zhang, Y., Li, Y., Yan, Y., et al. (2018). Rapid improvement in visual selective attention related to action video gaming experience. *Front. Hum. Neurosci.* 12:47. doi: 10.3389/fnhum.2018.00047

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Shehata, M., Elnagar, S., Yasunaga, S., Nakauchi, S., and Shimojo, S. (2018). Flow of the eye: gaze direction as an objective measure of flow experience. *J. Vision* 18, 1205–1205. doi: 10.1167/18.10.1205

Shore, D. I., Spence, C., and Klein, R. M. (2001). Visual prior entry. *Psychol. Sci.* 12, 205–212. doi: 10.1111/1467-9280.00337

Sorensen, T., and Vasishth, S. (2015). Bayesian linear mixed models using stan: a tutorial for psychologists, linguists, and cognitive scientists. *arXiv preprint* arXiv:1506.06201.

Stamatelopoulou, F., Pezirkianidis, C., Karakasidou, E., Lakioti, A., and Stalikas, A. (2018). "being in the zone": a systematic review on the relationship of psychological correlates and the occurrence of flow experiences in sports' performance. *Psychology* 9:2011. doi: 10.4236/psych.2018.98115

Stegmueller, D. (2013). How many countries for multilevel modeling? A comparison of frequentist and bayesian approaches. *Am. J. Polit. Sci.* 57, 748–761. doi: 10.1111/ajps.12001

Stelmach, L. B., and Herdman, C. M. (1991). Directed attention and perception of temporal order. *J. Exp. Psychol. Hum. Percept. Perform.* 17:539.

Sugovic, M., Turk, P., and Witt, J. K. (2016). Perceived distance and obesity: It's what you weigh, not what you think. *Acta Psychol.* 165, 1–8. doi: 10.1016/j.actpsy.2016.01.012

Swann, C., Keegan, R. J., Piggott, D., and Crust, L. (2012). A systematic review of the experience, occurrence, and controllability of flow states in elite sport. *Psychol. Sport Exerc.* 13, 807–819. doi: 10.1016/j.psychsport.2012.05.006

Tanaka, R., and Yotsumoto, Y. (2017). Passage of time judgments is relative to temporal expectation. *Front. Psychol.* 8:187. doi: 10.3389/fpsyg.2017.00187

Thönes, S., and Stocker, K. (2019). A standard conceptual framework for the study of subjective time. *Conscious. Cogn.* 71, 114–122. doi: 10.1016/j.concog.2019.04.004

Tozman, T., Magdas, E. S., MacDougall, H. G., and Vollmeyer, R. (2015). Understanding the psychophysiology of flow: a driving simulator experiment to investigate the relationship between flow and heart rate variability. *Comput. Hum. Behav.* 52, 408–418. doi: 10.1016/j.chb.2015.06.023

Tse, P. U., Intriligator, J., Rivest, J., and Cavanagh, P. (2004). Attention and the subjective expansion of time. *Percept. Psychophys.* 66, 1171–1189. doi: 10.3758/bf03196844

Ulrich, R. (1987). Threshold models of temporal-order judgments evaluated by a ternary response task. *Percept. Psychophys.* 42, 224–239.

Ursano, R. J., Fullerton, C. S., and Benedek, D. M. (2007). "Peritraumatic dissociation: time perception and cerebellar regulation of psychological, interpersonal, and biological processes," in *Traumatic Dissociation: Neurobiology and Treatment*, eds E. Vermetten, M. Dorahy, and D. Spiegel (American Psychiatric Publishing, Inc.), 181–190. Available online at: https://psycnet.apa.org/record/2007-05420-009

van der Kruijs, S. J., Bodde, N. M., Carrette, E., Lazeron, R. H., Vonck, K. E., Boon, P. A., et al. (2014). Neurophysiological correlates of dissociative symptoms. *J. Neurol. Neurosurg. Psychiatry* 85, 174–179. doi: 10.1136/jnnp-2012-302905

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Stat. Comput.* 27, 1413–1432. doi: 10.1007/s11222-016-9696-4

Wearden, J. (2015). Passage of time judgements. *Conscious. Cogn.* 38, 165–171. doi: 10.1016/j.concog.2015.06.005

Wesp, R., Cichello, P., Gracia, E. B., and Davis, K. (2004). Observing and engaging in purposeful actions with objects influences estimates of their size. *Percept. Psychophys.* 66, 1261–1267. doi: 10.3758/bf03194996

West, G. L., Anderson, A. A., and Pratt, J. (2009). Motivationally significant stimuli show visual prior entry: evidence for attentional capture. *J. Exp. Psychol. Hum. Percept. Perform.* 35:1032. doi: 10.1037/a0014493

West, G. L., Stevens, S. A., Pun, C., and Pratt, J. (2008). Visuospatial experience modulates attentional capture: evidence from action video game players. *J. Vision* 8, 13–13. doi: 10.1167/8.16.13

Witt, J. K. (2019). Action's influence on spatial perception: resolution and a mystery. *Curr. Opin. Psychol.* 32, 153–157. doi: 10.1016/j.copsyc.2019.07.045

Witt, J. K., Linkenauger, S. A., Bakdash, J. Z., and Proffitt, D. R. (2008). Putting to a bigger hole: golf performance relates to perceived size. *Psychon. Bull. Rev.* 15, 581–585. doi: 10.3758/pbr.15.3.581

Witt, J. K., and Proffitt, D. R. (2005). See the ball, hit the ball. *Psychol. Sci.* 16:937. doi: 10.1111/j.1467-9280.2005.01640.x

Witt, J. K., Proffitt, D. R., and Epstein, W. (2004). Perceiving distance: a role of effort and intent. *Perception* 33, 577–590. doi: 10.1068/p5090

Witt, J. K. and Sugovic, M. (2010). Performance and ease influence perceived speed. *Perception* 39, 1341–1353. doi: 10.1068/p6699

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.

Efficacy, Trainability, and Neuroplasticity of SMR vs. Alpha Rhythm Shooting Performance Neurofeedback Training

Anmin Gong[1], Wenya Nan[2], Erwei Yin[3], Changhao Jiang[4]* and Yunfa Fu[5]*

[1]School of Information Engineering, Engineering University of Armed Police Force, Xi'an, China, [2]Department of Psychology, College of Education, Shanghai Normal University, Shanghai, China, [3]Tianjin Artificial Intelligence Innovation Center (TAIIC), National Institute of Defense Technology Innovation, Academy of Military Sciences China, Beijing, China, [4]Key Laboratory of Sports Performance Evaluation and Technical Analysis, Capital University of Physical Education and Sports, Beijing, China, [5]School of Automation and Information Engineering, Kunming University of Science and Technology, Kunming, China

Previous literature on shooting performance neurofeedback training (SP-NFT) to enhance performance usually focused on changes in behavioral indicators, but research on the physiological features of SP-NFT is lacking. To explore the effects of SP-NFT on trainability and neuroplasticity, we conducted a study in which 45 healthy participants were randomly divided into three groups: based on sensory-motor rhythm of C3, Cz and C4 (SMR group), based on alpha rhythm of T3 and T4 (Alpha group), and no NFT (control group). The training was performed for six sessions for 3 weeks. Before and after the SP-NFT, we evaluated changes in shooting performance and resting electroencephalography (EEG) frequency power, participant's subjective task appraisal, neurofeedback trainability score, and EEG feature. Statistical analysis showed that the shooting performance of the participants in the SMR group improved significantly, the participants in the Alpha group decreased, and that of participants in the control group have no change. Meanwhile, the resting EEG power features of the two NFT groups changed specifically after training. The training process data showed that the training difficulty was significantly lower in the SMR group than in the Alpha group. Both NFT groups could improve the neurofeedback trainability scores and change the feedback features by means of their mind strategy. These results may provide evidence of trainability and neuroplasticity for SP-NFT, suggesting that the SP-NFT is effective in brain regulation and thus provide a potential method to improve shooting performance.

Keywords: neurofeedback, shooting performance, motor sensory rhythm, resting EEG, trainability

HIGHLIGHTS

- The neurofeedback based on SMR may be more effective than that based on alpha rhythm in improving the non-expert shooter's pistol shooting performance.
- SMR and Alpha NFT have trainability. Participants can increase their feedback features and neurofeedback trainability scores during NFT stage.
- SMR and Alpha neurofeedback exert some effects on the effects on trainability and neuroplasticity, and feedback training can make specific directional changes in resting EEG.

INTRODUCTION

Shooting is a simple motor behavior that can be easily affected by mental states such as attention and emotion. The relationship between shooting performance and central nervous system features has raised interest in many scholars, and this relationship has been widely investigated for shooting athletes through electroencephalography (EEG; Hatfield et al., 1984; Del Percio et al., 2009; Bertollo et al., 2016). For instance, it has been reported that during the preparation process in shooting, shooting experts show continuously increased alpha rhythm (8–12 Hz) at the T3 electrode (Hatfield et al., 1984). Furthermore, during the preparation process, a steady increase of theta rhythm (4–7 Hz) power at the frontal midline is shown in shooting experts rather than novice shooters (Doppelmayr et al., 2008), and the event-related desynchronization (ERD) and event-related coherence (ERCoh) of alpha rhythm in shooting experts is significantly less than that of novice shooters (Del Percio et al., 2009, 2011). In addition, a decrease in alpha power of the occipital region is reported during successful trials compared to fail shooting trials (Loze et al., 2001). These results suggest that shooting performance is closely related to the shooter's brain activity during shooting preparation.

It is unknown if it is naturally possible to enhance shooting performance by interventions that influence brain activity. Therefore, in the movement neuroscience, brain regulation for enhancing shooting performance using neurofeedback training (NFT) has become a new research focus. NFT converts EEG signals into sound or animation, which is easily understood by the participants, to help people understand their own physical status. Participants can selectively enhance or suppress neurophysiological signals of a specified frequency band through repeated training and effectively regulate their brain function.

Many studies have used NFT to improve the sports performance of athletes (Raymond et al., 2005; Faridnia et al., 2012; Strizhkova et al., 2012; Cheng et al., 2015a; Mikicin et al., 2015; Ring et al., 2015). For instance, Raymond et al. (2005) improved the dance performance of the college dance sports team by increasing the alpha/theta ratio of the Pz electrode. Strizhkova et al. (2012) improved the complex coordinated activities performance of gymnasts by increasing alpha of F1, F2, P3 and P4 electrodes. Faridnia et al. (2012) reduced the sport competition anxiety of swimmers by increasing SMR and low beta and decrease theta and a high beta of C3 and C4 electrodes. Ring et al. (2015) improved the golfers' putting performance by reducing theta and high alpha power of the Fz electrode. Mikicin et al. (2015) reduced student-athletes' attention-reaction by increasing beta1 and SMR and decrease theta and beta2 of C3 and C4 electrodes. Cheng et al. (2015a) improved golfers' putting performance by increasing SMR of the Cz electrode.

In the literature, only two studies attempted to improve shooting performance by NFT (SP-NFT). One SP-NFT study utilized SMR (Sensorimotor Rhythm, 12–15 Hz EEG rhythm, usually collected from C3, Cz, and C4), beta and alpha rhythm mixed NFT protocol to improve shooting performance (Rostami et al., 2012). This approach was taken because

a lot of previous studies have found that the increase in SMR is often accompanied by attention increases. The NFT based on SMR has been widely used in the treatment of attention deficit hyperactivity disorder (ADHD), and its activities are also found to be closely related to the optimization of the skilled action execution motor performance such as golf putting and dart throwing (Vernon et al., 2004; Cheng et al., 2015a,b). Therefore, the researchers expected to increase the participants' attention by increasing SMR, so as to improve their shooting performance.

On the other hand, it has been demonstrated that there is an asymmetry between the left and right hemispheres and an increase of alpha rhythm in the left temporal region during rifle and archery shooting preparation (Hatfield et al., 1984; Salazar et al., 1990). Based on this phenomenon, the other SP-NFT study utilized NFT to enhance EEG low-frequency activity over the left temporal region (T3) and successfully improved archers shooting performance (Landers et al., 1991). The authors explained this NFT could improve the shooting performance by simulating brain activity during actual rifle and archery shooting preparation.

Although both studies used NFT to improve shooting performance, it is unknown which of the two methods is more effective. In addition, according to the current standards of neurofeedback experimental research, there are still some problems in these two studies. Mirifar et al. (2017) suggested that neurofeedback combining visual and auditory feedback may be more effective than visual or auditory feedback alone. However, the two kinds of SP-NFT mentioned above only use visual feedback. Additionally, in Landers's study, participants were tested for feedback effects after only one feedback training session. Some scholars have suggested that successful neurofeedback regulation may require a minimum of three to four sessions, and initial improvements can only be seen within the first five to ten sessions (Konareva, 2005; Hammond, 2011). More importantly, both studies only analyzed the efficacy of NFT from a kinematic perspective and did not provide a detailed analysis of the effect of NFT on trainability and neuroplasticity (Enriquez-Geppert et al., 2014). That is, there was no indication of whether the EEG features of feedback can be actively modulated by the participants during the training process, i.e., trainability, or if this training can produce some lasting changes to the EEG rhythm of the brain, i.e., neuroplasticity.

In previous NFT studies, the NFT trainability was firstly examined before confirming that the NFT can affect cognition/behavior. For instance, Cho et al. (2007) confirmed the NFT trainability of alpha activity in the midline parietal region and found that NFT could enhance participants' ability to maintain alpha activity. Zoefel et al. (2011) first studied the trainability of upper alpha NFT in the parietal and occipital regions and found the effect of this NFT on cognitive improvement. Enriquez-Geppert et al. (2014) demonstrated the trainability of frontal midline theta NFT and revealed the role of training in cognitive ability improvement. These literatures suggest that proving NFT trainability is the basis of studying the effects of NFT on cognition/behavior.

Studies regarding neuroplasticity are mainly based on the correlations between neurological characteristics and cognitive/behavioral indicators. Many studies found a significant correlation between neurological characteristics of resting states and behavioral/cognitive indicators (Babiloni et al., 2010; Zhou et al., 2012; Wan et al., 2014; Zhang et al., 2015). Therefore, scholars believe that if NFT changes not only the cognitive/behavioral indicators but also the relevant neurological characteristics, it will be an important evidence supporting its influence in cognition/behavior by altering the neurological characteristics. Many scholars have studied the neuroplasticity generated by NFT from different perspectives. Besides finding that NFT can produce neuroplasticity changes in resting EEG, NFT can lead to changes in white and gray matter in the brain and the resting brain network features (Cho et al., 2007; Zoefel et al., 2011; Ghaziri et al., 2013; Kluetsch et al., 2014).

Here we sought to improve SP-NFT by using both visual and auditory feedback and increasing the number of sessions, while also examining trainability and neuroplasticity. To examine trainability, we recorded the extent to which each participant could actively modulate EEG features during the training process. To examine neuroplasticity, we measured resting-state EEG before and after NFT. Current studies have found significant correlation between resting EEG features such as individual alpha frequency (IAF) and frequency band power and the performance of participants completing cognitive tasks, motor imagery tasks, and even shooting tasks (Koch et al., 2008; Zhou et al., 2012; Zhang et al., 2015; Gong et al., 2017). Other studies also found that there were significant differences in resting EEG features among participants of different sport levels (Gong et al., 2017). Therefore, we believe that resting EEG is a good neurological indicator reflecting the brain "baseline" state and can be used to analyze the brain neuroplasticity induced by training (Khanna et al., 2015; Zhang et al., 2015).

Previous studies on SP-NFT have not specifically focused on the changes in resting EEG caused by NFT. Therefore, they could not fully demonstrate whether NFT really changed brain activity or whether the observed effect was simply a placebo effect of neurofeedback (Schönenberg et al., 2017; Xiang et al., 2018).

These research limitations are one of the main reasons why NFT technology has been controversial and has not gained as much popularity in the kinematics field (Gruzelier, 2014; Mirifar et al., 2017).

To better explore the effects of SP-NFT on trainability and neuroplasticity, we hypothesized the following: (1) SP-NFT has an effect on trainability, participants can master the SP-NFT, and improve neurofeedback trainability scores; and (2) SP-NFT has an effect on neuroplasticity, such that resting-state EEG activity is significantly changed.

MATERIALS AND METHODS

Participants

The 45 healthy college students (male, age: 19.5 ± 2 years) from the Armed Police Engineering University voluntarily took part in this research. All participants completed pistol course learning, had qualifying grades, and mastered fixed-target pistol shooting skills. Before training, all participants were divided into three groups according to their age, height, weight, and somatotype. The SMR group ($N = 15$) aimed to enhance SMR (12–15 Hz) power of C3, Cz, and C4 channel, similar to the study of Rostami et al. (2012). The Alpha group ($N = 15$) training aimed to enhance the alpha rhythm (8–12 Hz) power of T3 channel and decrease that of T4 channel, similar to the study of Landers et al. (1991). The control group ($N = 15$) did not receive NFT, just underwent a shooting test. This study protocol was in accordance with the Declaration of Helsinki and approved by the university ethics committee. Before the experiment, the experimental processes and purpose were explained, and the written informed consent was obtained from all participants. Participants were free to withdraw from the experiment at any time.

Experimental Design

Figure 1 shows the experimental design and flow of our study. First, all participants underwent a shooting performance pre-test and a resting EEG pre-test within 3 days before NFT. Next, during the NFT period, the SMR group and Alpha group underwent six NFT sessions within 3 weeks at convenience. Finally, within 3 days after all NFTs, they completed
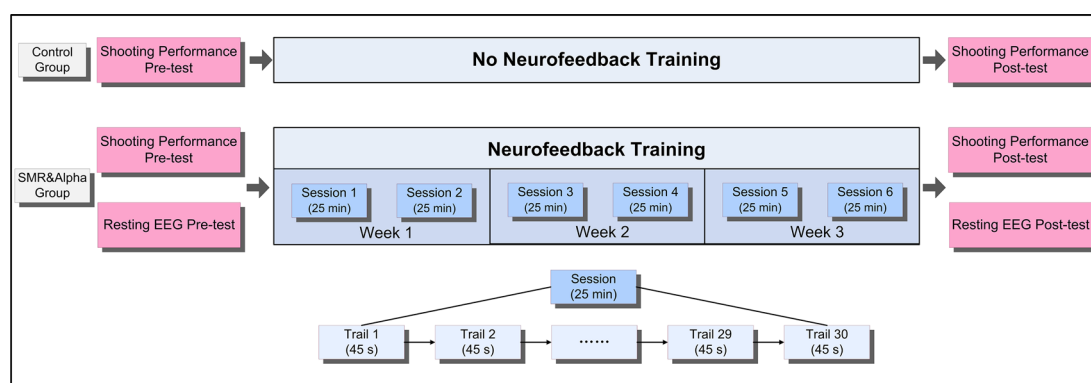

FIGURE 1 | The experimental design and flow.

a shooting performance post-test and a resting EEG post-test. It is worth noting that the control group only underwent shooting performance pre-test and post-test, and all participants did not perform any excess shooting tasks during the 3 weeks.

Shooting Performance Pre-test and Post-test

All participants underwent two pistol shooting tasks to evaluate the effect of NFT on shooting performance. The shooting test was organized by the College of Basic Military Education. Participants used a type 92 pistol, aiming at a target 25 m away. The dimensions of the target were 52 × 52 cm. It included 10 rings, with a diameter of 10 cm and 10 ring edges, each extending 5 cm followed by 9, 8, 7, and 6 rings. The corresponding shooting score was obtained by hitting the position on the target, and a miss was recorded as 0. For example, if the shooter hit the center of the target, a score of 10 was recorded. Every shooter was asked to take a standing position in a single firing mode. Prior to each shot, the participants were informed of their previous shooting score. All participants performed 25 shots at their own pace.

Acquisition and Preprocessing

The resting EEG acquisition device was a Beijing SymTom 32-D EEG amplifier. The EEG signals were recorded from 32 electrodes according to the international 10–20 system, including Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, Pz, FC3, FC4, CP3, CP4, FT7, FT8, TP7, TP8, FCz, CPz, Oz, PO3, and PO4. The ground electrode was placed over the forehead and the reference was the left and right mastoids. The impedance of each electrode was kept below 5 kΩ and the sampling frequency was 1,000 Hz. The participants were asked to sit on soft and comfortable seats, while asking them to remain relaxed but not to fall asleep, and do not try to recall anything. EEG data of participants were collected with eyes-closed for 5 min and eyes-open for 5 min. To control the alertness level, the participants' behavior and the quality of the EEG signals were monitored online in real-time. If the EEG had abnormal changes due to coughing, manual activity, sleepiness, etc., the participants were verbally reminded to cooperate.

The EEG signals at all recording channels were analyzed offline in MATLAB (2014a) and EEGLAB toolbox (Delorme and Makeig, 2004). The signals were firstly segmented into 2 s epochs and the data epochs corrupted by artifact were rejected by visual inspection (Delorme and Makeig, 2004). Independent component analysis (ICA; fast ICA algorithm) was used to remove blinking artifacts during the eyes-open resting state (Jung et al., 2001). Blinking artifacts were recognized by calculating the correlation coefficient with the EEG signals of Fp1 and Fp2 channels. Any component with a correlation coefficient greater than 0.8 was considered as the blinking artifact and was removed (the 0.8 was empirical data). Then a 6-order Butterworth band-pass filter of 0.5–40 Hz was used to filter the signal. Finally, the clean EEG data were used for subsequent frequency band power analyses. The frequency band power of the EEG signal was calculated using the Welch method. The calculated time window spanned 5,000 sampling points (5 s) and the overlap rate was 50%.

To control for differences in individual frequency bands between participants, the frequency bands of interest were defined relative to the IAF (Klimesch, 1999). The EEG power spectrum is calculated in the occipital region during the resting eyes-closed state, and the peak frequency position is found in the range of 7–13 Hz frequency bands. Finally, this peak frequency position was recorded as the IAF of the participants, and other frequency bands were defined relative to the IAF as follows: theta as IAF −6 Hz to IAF −3 Hz, alpha as IAF −2 Hz to IAF +2 Hz, beta as IAF +3 Hz to IAF +20 Hz.

Implementation of Shooting Performance NFT

According to previous studies, an effective neurofeedback experiment should preferably include at least five training sessions, and the interval between two training sessions should be at least 1 day (Mirifar et al., 2017). Therefore, our study design included six sessions of SP-NFT in 3 weeks, in which each participant would undergo two sessions a week with an interval of at least 1 day in-between (**Figure 1**).

To ensure double-blinded NFT, experimenters were only required to enter the unique identity number of the participants into the NFT system (Ros et al., 2019). The system automatically identified the feedback type of the participants and ran the corresponding NFT mode. Therefore, the experimenter was not aware of the feedback mode in which the participant was training.

The EEG signals during SP-NFT were recorded from Cz, C3, C4, T3, and T4 for all participants in both NFT groups (**Figure 2**). The EEG signals were transmitted to the computer by the USB interface, and the SP-NFT program written in MATLAB 2014 was used to process and calculate the EEG features for feedback to participants in real-time. The feedback feature of the SMR group was the average SMR power of Cz, C3, and C4, while the feedback feature of the Alpha group was the alpha power of the T3 electrode minus alpha power of T4 electrode.

Each training session lasted about 25 min and included 30 trials. Each trial included a 15-s relaxation break and 30-s SP-NFT. During the relaxation break, the participants sat in silence and did not deliberately recall anything. At the end of the relaxation stage, the median value of the feedback feature in this period was shown as a red line on the feedback interface and served as the baseline of the SP-NFT trial. In the training stage, the feedback feature of the participants was calculated in real-time and presented to the participants in the form an image of a shooting target at different clarity levels, and a dynamic blue curve, both indicating the magnitude of the feedback feature. The participants were requested to focus on the shooting target image and to try to improve the feedback feature by means of their own mental strategy. Participants are advised not to be overly nervous during the NFT period, to remain relaxed but focused, to try to imagine the movements of the shooting preparation stage or to focus on the target image on the computer screen. When the feedback feature increases, the curve rises and the image becomes clearer; if the feedback

FIGURE 2 | Schematic figure of NFT.
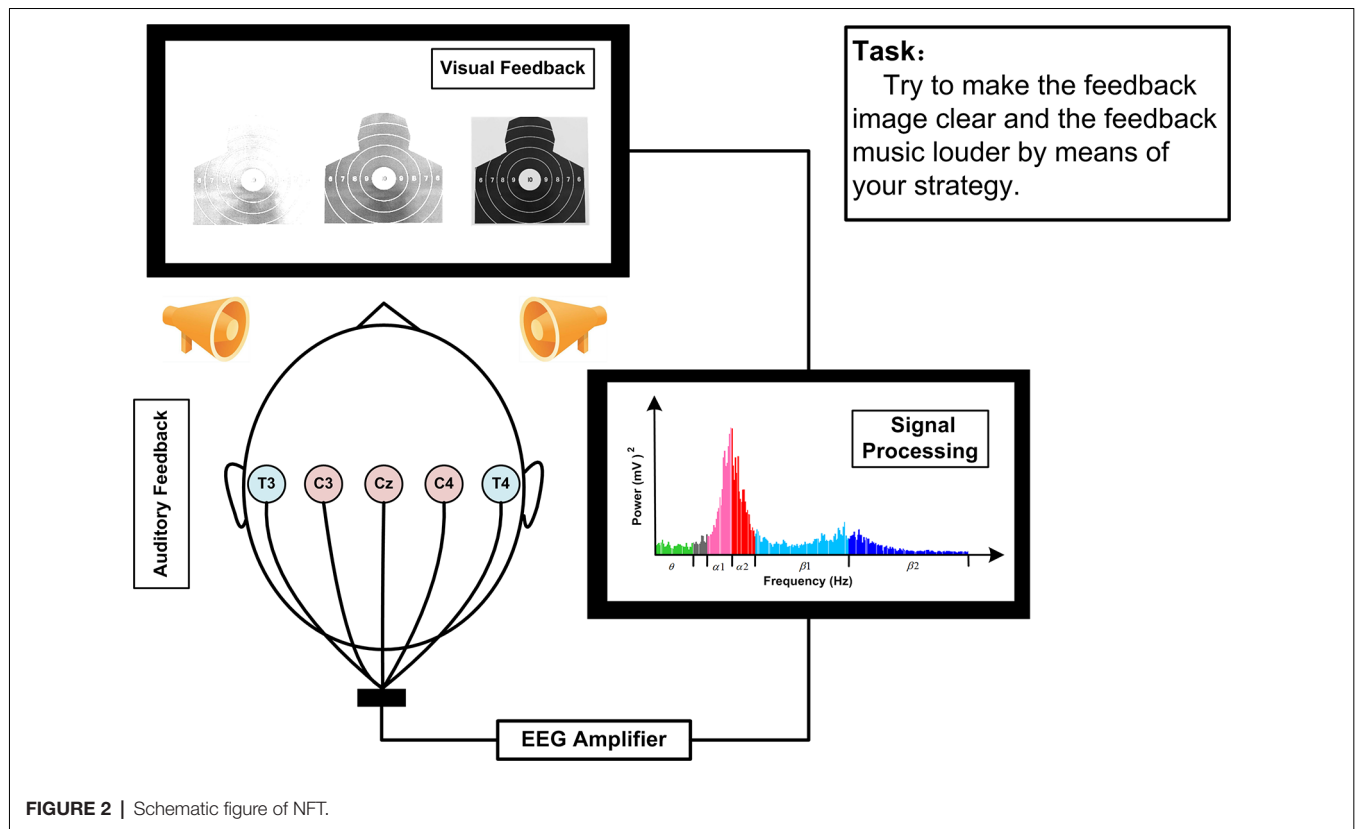
feature decreases, the curve falls and the image becomes less clear. To implement the important auditory component of our NFT paradigm, the system included a digital sound equalizer to control the play of feedback music. The parameters of the equalizer were controlled according to the real-time feedback EEG feature: at the beginning of the training stage, the music was lowest, and when the magnitude of the feedback feature was higher than the baseline, the feedback music would gradually increase in volume and clarity.

To evaluate the success of SP-NFT, the system automatically calculated the neurofeedback trainability scores after each trial. The trainability score was determined as the ratio of time the feedback features were higher than the baseline over the total time of the training period. If the feedback features were always higher than the baseline during the feedback period, the neurofeedback trainability scores were recorded as 100 points. On the contrary, if the feedback features were always lower than the baseline, the neurofeedback trainability scores were recorded as 0. At the end of each feedback trial, the feedback score displayed and shown to the participant.

Subjective Task Appraisal of Participants

To compare the subjective outcomes of SP-NFT between SMR groups and the Alpha group, a participants' subjective task appraisal scale was used to evaluate the two kinds of training. After each SP-NFT session, participants reported the degree of fatigue, commitment, and difficulty in relation to the training. The reported scale used a five-level Likert scale: one indicated the

lowest degree and five indicated the highest degree. For example: for fatigue degree, one means no fatigue at all, and five means very tired.

Statistical Analysis

It is noteworthy that we tested the probability distribution of the samples using the Kolmogorov–Smirnov test. Regarding shooting performance, resting EEG IAF, and resting EEG frequency band power, it was found that these samples were not all subjected to Gaussian distribution. Thus, we used the Wilcoxon signed-rank test for these samples to examine the difference between pre-test and post-test in each group. For subjective task appraisal index and NFT trainability score, feedback feature and repeated-measures ANOVA was applied because all the samples were subjected to Gaussian distribution.

Shooting Performance Index

For each participant, the shooting performance index was the mean value of the 25 shot scores. To examine the differences in shooting performance between pre-test and post-test, Wilcoxon signed-rank test was used to test the median difference of shooting performance index between pre-test and post-test in the SMR group, Alpha group, and control group, respectively. Then, to compare the difference in the effect of the SMR group and the Alpha group on the shooting performance, we also carried out the Wilcoxon rank-sum test on the shooting scores of the post-test minus pre-test for the SMR group and the Alpha group.

Subjective Task Appraisal Index

For the subjective task appraisal index of the two SP-NFT groups, a statistical analysis was applied to test the difference between the two groups. To test whether the participants had experienced any changes during the six training sessions based on the subjective task appraisal, a repeated-measure ANOVA with the factors group (SMR vs. Alpha) and the factors session was performed (1–6).

Neurofeedback Trainability Scores and Feedback Feature

To examine the neurofeedback trainability of two SP-NFT groups, two repeated-measures ANOVAs were performed. The first ANOVA with factors group (SMR vs. Alpha) and factors session (1–6) compared the real-time calculated neurofeedback trainability scores for each trial, averaged within each session and compared across sessions and groups. This score was calculated as the ratio of the time that the feedback feature remained above baseline for each NFT trial. The second ANOVA with within-subject factor state (Train vs. Relax) and within-subject factor session (1–6) was conducted for the averaged values of the individualized feedback feature across all trials within each session.

Resting EEG Rhythm Power Indexes

Finally, we analyzed the changes in resting EEG features before and after training from three perspectives: (1) the resting EEG IAF (calculated by the eyes-closed resting EEG in the occipital region); (2) changes in the EEG power spectrum and the specific frequency band power in channels of interest (COI) during training; and (3) calculation of the average whole-brain EEG frequency band power topographic map of all participants and visually comparing the changes of the whole brain EEG before and after training. For the resting EEG IAF and band power of COI, the difference between pre-test and post-test was tested using the Wilcoxon signed-rank test. The visual comparison of the whole brain map only reflects the change of the average power spectrum before and after the feedback training; it does not reflect statistical testing. All statistical tests are performed in MATLAB 2014.

RESULTS

Shooting Performance Before and After the SP-NFT

Figure 3 indicates the comparisons of the shooting scores of the three groups in pre-test and post-test. The left is the SMR group,

FIGURE 3 | The shooting scores box plot of the pre-test and post-test of three groups.
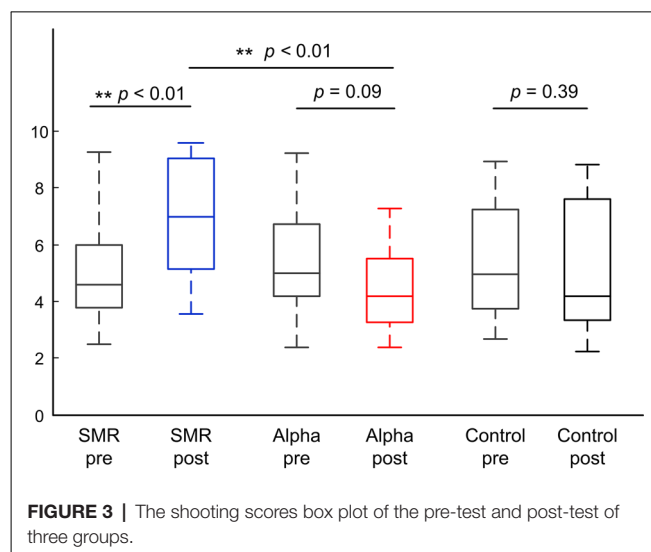
the middle is the Alpha group, and the right is the Control group. Blue, red, and black box indicates post-test shooting performance and the gray box indicates pre-test shooting performance. The statistical results show that the median shooting score after training is significantly higher than that before training in the SMR group ($z = -3.55$, $p < 0.01$). The median shooting score after training is marginal significantly lower than that before training in the Alpha group ($z = 1.80$, $p = 0.09$). The median shooting score have no significant change in control group ($z = 0.85$, $p = 0.39$). In terms of shooting score difference between two SP-NFT groups, the SMR group is significantly higher than the Alpha group ($z = -3.06$, $p < 0.01$).

Subjective Task Appraisal of Participants

Table 1 shows the mean and SD of the three subjective task appraisal indexes for the six feedback sessions. As can be seen from the table, the fatigue degree of the two groups was approximately 2, while the degrees of effort and difficulty were approximately 3 in both groups. Results of repeated-measure ANVOA statistical analyses show that for fatigue index, the group factor effect is not significant $F_{(1,14)} = 0.44$, $p > 0.05$, and the session factor effect is not significant $F_{(5,70)} = 0.47$, $p > 0.05$. For commitment index, the group factor effect is significant $F_{(1,14)} = 10.54$, $p < 0.01$, such that the Alpha group is significant higher than SMR group; the session factor effect is not significant $F_{(5,70)} = 0.68$, $p > 0.05$. For difficulty index, the group factor effect is significant $F_{(1,14)} = 17.02$, $p < 0.01$, with higher

TABLE 1 | Mean and SD of each feedback subjective task appraisal of the two neurofeedback training (NFT) groups.

	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6
Fatigue (SMR group)	2.27 (1.03)	1.93 (0.70)	2.20 (0.86)	2.20 (0.94)	1.93 (0.96)	2.00 (1.07)
Fatigue (Alpha group)	2.13 (0.74)	2.13 (0.92)	2.00 (0.85)	1.93 (0.80)	2.00 (0.85)	1.80 (0.68)
Commitment (SMR group)	2.80 (0.56)	2.93 (0.70)	2.80 (0.86)	2.53 (0.92)	2.87 (0.99)	2.60 (0.74)
Commitment (Alpha group)	3.40 (0.99)	3.13 (0.92)	3.27 (0.80)	3.00 (0.76)	3.00 (0.93)	3.13 (0.74)
Difficulty (SMR group)	2.27 (1.03)	2.33 (0.90)	2.33 (1.11)	2.33 (1.11)	2.40 (1.12)	2.40 (1.06)
Difficulty (Alpha group)	2.93 (0.59)	2.67 (0.82)	3.00 (0.65)	3.07 (0.80)	2.87 (0.83)	2.93 (0.96)

From top to bottom, there are the results of fatigue, commitment, and difficulty of SMR group and Alpha group, respectively, and from left to right, session 1 to session 6.

difficulty in the Alpha group compared to the SMR group; the session factor effect is not significant $F_{(5,70)} = 0.18$, $p > 0.05$. The interactions between session and group of the three indicators are not significant.

Dynamic Changes in Parameters of SP-NFT

The Dynamics of Trainability Scores

Figure 4 shows the mean and 1.96× standard errors of neurofeedback trainability scores for each session. The horizontal axis is the feedback training session, the vertical axis is the neurofeedback trainability scores, the blue line indicates the SMR group, and the red indicates the Alpha group. The neurofeedback trainability scores of both NFT groups increased as the number of training sessions increased. The results of repeated—measure ANOVA showed the group factor effect is significant ($F_{(1,14)} = 5.43$, $p < 0.01$, *post hoc*: SMR group > Alpha
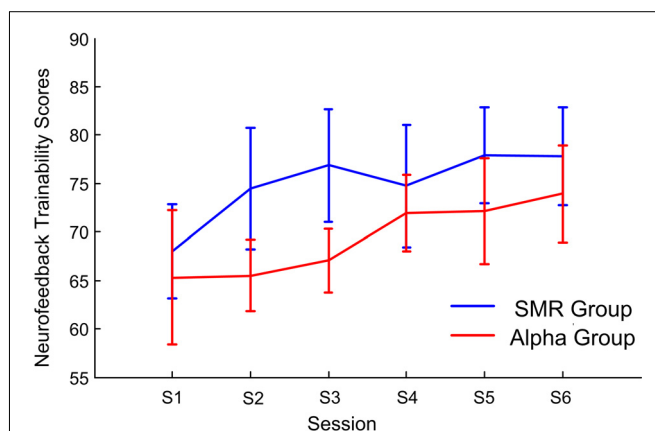
FIGURE 4 | The average and 1.96× standard errors of the neurofeedback trainability scores vary with the feedback session.

group). For the session factor effect is significant ($F_{(5,70)} = 3.13$, $p < 0.01$, *post hoc*: session 6 > session 1). This finding indicated that participants could master the neurofeedback well after six sessions of SP-NFT, and the neurofeedback trainability scores were significantly improved.

The Dynamic of Feedback Features

Figure 5 shows the dynamics of feedback features with training sessions. The left side is the SMR group and the right side is the Alpha group. The gray line is the feedback feature in a resting state; the blue and red lines indicate the feedback feature of SMR and Alpha groups, respectively. For the SMR group, the results of repeated—measure ANOVA showed that the state factor effect was significant ($F_{(1,14)} = 78.25$, $p < 0.001$, *post hoc*: Train > Relax) while the session factor effect was not significant ($F_{(5,70)} = 0.68$, $p > 0.05$). For the Alpha group, the results of repeated- measure ANOVA showed that the state factor effect was significant ($F_{(1,14)} = 7.32$, $p < 0.01$, *post hoc*: Train > Relax) whereas the session factor effect was not significant ($F_{(5,70)} = 0.28$, $p > 0.05$). Comparing the two groups, it was found that the feedback features of both NFT groups in the training state were significantly higher than that of the relaxation state, while the *F* value of the Alpha group was lower than that of the SMR group. This may also be one of the reasons why participants in the Alpha group considered the training was more difficult than those in the SMR group.

The Effect of SP-NFT on the Resting EEG

Comparison of Pre-test and Post-test IAF

The median value (SD) of IAF calculated according to the resting eyes-closed EEG is as follows. In the SMR group, the pre-test is 10.64 ± 0.69 Hz and the post-test is 10.45 ± 0.56 Hz, and there was no significant difference between pre-test and post-test ($z = 0.99$, $p = 0.32$). In the Alpha group, the pre-test is 10.30 ± 0.83 Hz and the post-test is 10.23 ± 0.92 Hz, and
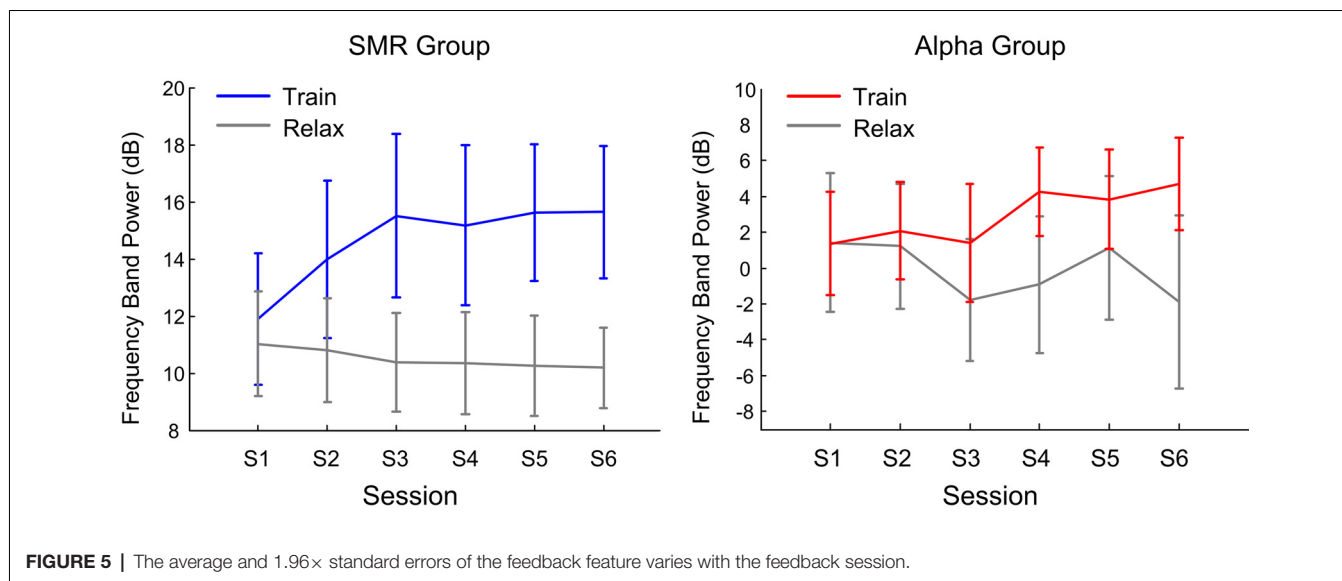
FIGURE 5 | The average and 1.96× standard errors of the feedback feature varies with the feedback session.

there was no significant difference between pre-test and post-test ($z = 1.42$, $p = 0.16$).

Comparison of Pre-test and Post-test Resting EEG Power Spectrum on COI

Figure 6 is the power spectrum of the resting eyes-closed EEG on the feedback channel before and after training for the SMR group and Alpha group, respectively. The left side corresponds to the SMR group and the right side, to the Alpha group. The asterisk (*) indicates that there was a significant change in the frequency band power between before

and after training ($p < 0.05$). **Figure 7** shows the power spectrum of the resting eyes-open EEG, with similar details as in **Figure 6**. The power spectrum of the two groups changed after the training compared with before the training, and the change was in the same direction as the direction of SP-NFT: the training for the SMR group consisted in increasing the SMR power of Cz, C3, and C4, causing the resting eyes-closed alpha frequency power of the Cz channel ($z = -2.96$, $p < 0.05$) and the resting eyes-open beta frequency power of the C3 and C4 channels to significantly increase (C3: $z = -4.39$, $p < 0.05$; C4: $z = -3.22$, $p < 0.05$). The

FIGURE 6 | The contrast of pre-test and post-test electroencephalography (EEG) power spectrum of the eyes-closed resting state. *Indicates there was significant difference in the frequency band power between the pre-test and post-test ($p < 0.05$).

Resting Eyes-Open EEG

FIGURE 7 | The contrast of pre-test and post-test EEG power spectrum of the eyes-open resting state. *Indicates there was significant difference in the frequency band power between the pre-test and post-test (*p* < 0.05).

training of the Alpha group consisted of increasing the alpha power of the T3 electrode and decreasing the alpha power of the T4 electrode, leading to a significant increase of the resting eyes-closed alpha frequency power at the T3 electrode ($z = -3.01$, $p < 0.05$), significant decreases of the resting eyes-closed and eyes-open alpha frequency power (eyes-closed: $z = 2.86$, $p < 0.05$; eyes-open: $z = 2.88$, $p < 0.05$) as well as the resting eyes-open beta frequency power at the T4 electrode ($z = 2.95$, $p < 0.05$).

Comparison of the Whole Brain Resting EEG Frequency Band Power

Figure 8 is a whole-brain topographic map of the resting eyes-closed EEG frequency power difference between the NFT groups before and after training. The upper is the SMR group and the lower is the Alpha group. From left to right are theta, alpha and beta frequency band. Red indicates that the EEG power of the post-test is higher than that of the pre-test, and the blue indicates that the EEG power of the post-test

is lower than that of a pre-test. **Figure 9** is a whole-brain topographic map of the resting eyes-open EEG frequency power difference between the groups before and after training. In the SMR group, three frequency bands in the prefrontal, frontal channels, and central channels were increased after feedback. For the Alpha group, the frequency band power of the left hemisphere increased after feedback, while that in the right hemisphere was slightly weakened. The resting EEG changes in both NFT groups were consistent with the enhanced direction of SP-NFT.

DISCUSSION

In this article, we explored and compared the efficacy, trainability, and neuroplasticity of SMR vs. alpha rhythm SP-NFT. We improved traditional neurofeedback paradigms by including an auditory feedback component, in addition to a visual feedback component. Furthermore, we increased the number of NFT sessions and carried out six NFT sessions for each participant, rather than the typical one or two sessions. In addition, we have also taken into account trainability and neuroplasticity, which have not been fully explored by previous researchers.

The Effect Analysis of SP-NFT on Shooting Performance

We evaluated the shooting performance of the three groups after SP-NFT and found that the groups achieved different results. The median shooting score reflects the overall shooting level. There was a significant improvement in the shooting performance of the SMR group, whereas there was a decrease in the performance of the Alpha group. Significant improvement in the shooting performance of the participants who participated in SMR training is consistent with the results of Rostami et al. (2012). In previous findings, an increase in SMR was often accompanied by an increase in attention (Vernon et al., 2004; Cheng et al., 2015b). NFT based on SMR has been widely used in the treatment of ADHD, and its activities have been found to be closely related to the optimization of the skilled action execution motor performance, such as golf putting and dart throwing (Vernon et al., 2004; Cheng et al., 2015a,b). These results also extend the potential facilitation effects of SMR training to athletes and healthy people in need.

Nonetheless, participants who underwent alpha training did not achieve improvement and even displayed a decline in shooting performance. The feedback feature of the Alpha group was characterized by the alpha power difference between T3 and T4, and the participants were given positive feedback when the feedback feature increased. That is, the participants got positive feedback irrespective of the increase in alpha power of T3 or a decrease in the alpha power of T4. Collura (2013) pointed out that in neurofeedback experiments, the training aimed at reducing activity may be a kind of "squeeze" enhancement training. Enhancing or reducing EEG activity in a brain region may lead to increased activity in that region (Plotkin and Rice, 1981). Although researchers try to suppress the activity of a certain brain area, the results of the training may instead

lead to the enhancement of the brain area. According to this view, Alpha group enhanced the alpha rhythm of left temporal region and decreased the alpha rhythm of the right temporal region; however, the results may strengthen the activity of both temporal hemispheres of the participants because the participants did not acquire their shooting skills through training and the shooting performance did not improve. Therefore, we speculate that if the feedback feature is changed to the alpha power of the Hemi-temporal region, better training results may be obtained.

On the other hand, in terms of EEG features, Landers et al. (1991) used the Slow Cortical Potential (SCP) signal, whereas we used the alpha rhythm signal in the present study. Although the experimental principle is the same, it may lead to different experimental results. In terms of participant selection, the participants of Landers et al. (1991) were pre-professional athletes, while our participants were military students with amateur shooting levels. These differences may also lead to the inconsistency between the results of our research and previous studies.

The Effect of SP-NFT on Trainability

We defined trainability as the ability of participants to control their own NFT features in training. The experimental results in "Dynamic Changes in Parameters of SP-NFT" section show that the trainability in both groups increased gradually throughout the six SP-NFT sessions. Neurofeedback trainability scores for the sixth session were significantly higher than those for the first session. The feedback features increased gradually up to the first four sessions, and the change range between the fifth and sixth sessions was stable. These results are consistent with previous studies of alpha training and frontal middle line theta training (Cho et al., 2007; Zoefel et al., 2011; Enriquez-Geppert et al., 2014). In addition, we also inquired about the strategy used by the participants who had high neurofeedback trainability scores during NFT. Most of them reported that they were "focusing on one point in the target image" or "to concentrate on the motor imagination of the shooting preparation stage." These results indicate that both feedback training modes are effective and the participants can actively modulate their EEG rhythm power features. In particular, the subjective task appraisal showed that the degree of difficulty for the Alpha group was significantly higher than for the SMR group, which indicates that SMR feedback might be easier and more convenient for participants.

Effects of SP-NFT on Neuroplasticity on Resting EEG

The third focus of our research was to study the effects of SP-NFT on neuroplasticity, that is, whether SP-NFT can change the brain neural activity and whether the brain activity of the participant has changed throughout a period of training (Ghaziri et al., 2013; Megumi et al., 2015; Faller et al., 2019). In this study, we tested this by examining the resting EEG of the participants before and after NFT.

The results of our study showed that after training, the resting eyes-closed and eyes-open EEGs significantly changed
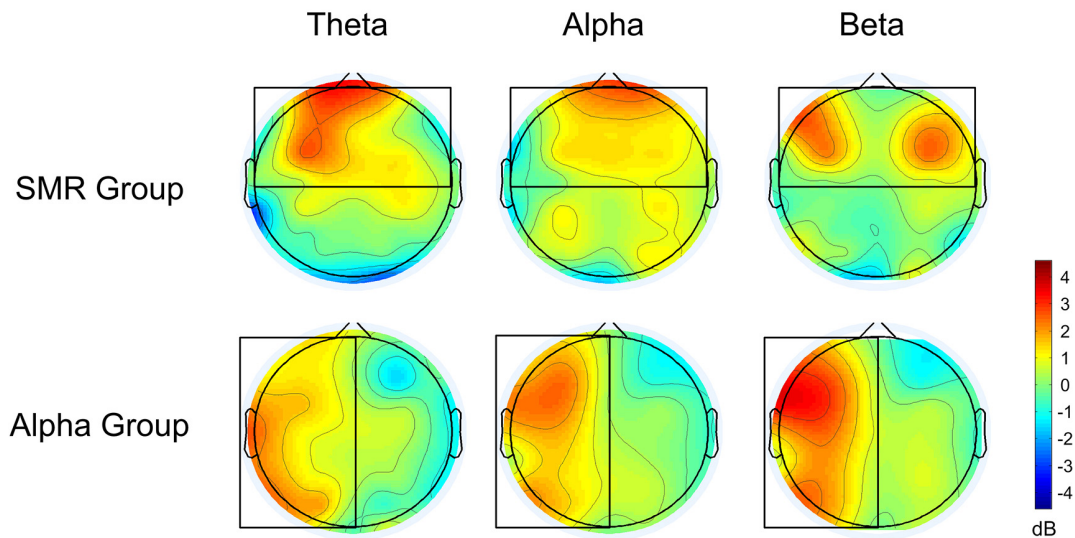
FIGURE 8 | The difference of EEG power between pre-test and post-test of the eyes-closed resting state.
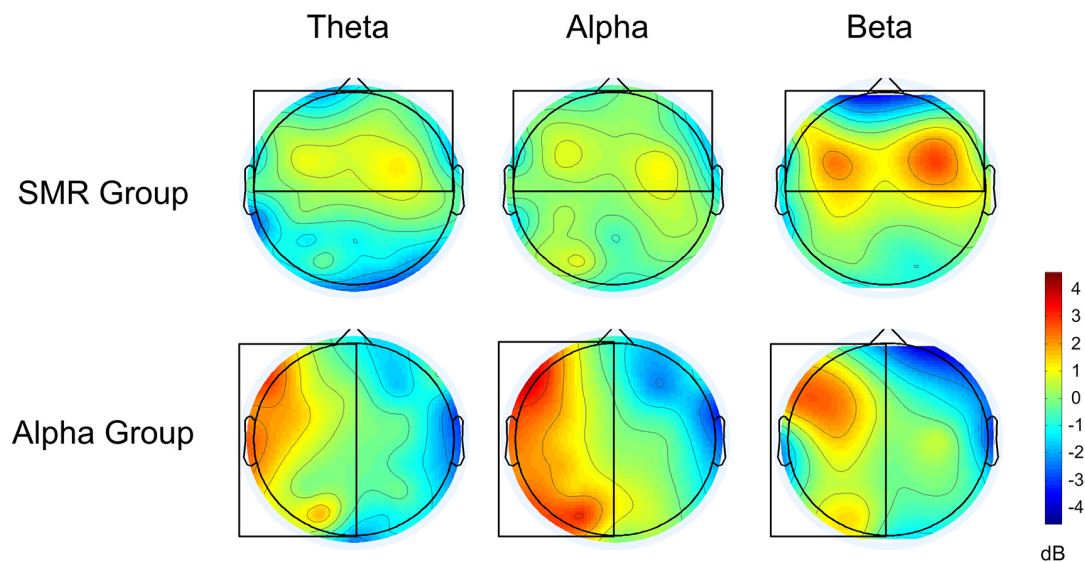
FIGURE 9 | The difference of EEG power between pre-test and post-test of eyes-open resting.

in both NFT groups. From the power spectrum of COI and the power topographic map of the whole brain, the position at which the resting state EEG changed was the region where the feedback electrode was placed. This suggests that NFT can cause specific changes in the channels and frequency bands involved in feedback, and it provides evidence that NFT

exhibits neuroplasticity at the EEG level. However, from the brain topographic map, we find that in addition to the trained target channels and frequency bands, the adjacent channels and frequency bands have also undergone trend changes, which are a non-specific change. Collura (2013) suggest this could be a kind of influence of the "entrainment" effect: in addition

to the frequency bands and channels involved in feedback, the EEG power of other nearby frequency bands and the cerebral cortex also showed a certain degree of change. Other neurofeedback also has reported similar phenomena (Cheng et al., 2015a).

The resting-state EEG of the participants changed significantly after SP-NFT, which is not only strong evidence that neurofeedback promotes neuroplasticity, but it also may provide a reason why SP-NFT can improve behavior indices (shooting performance). This shows that in future training processes of skill learning, we may be able to incorporate appropriate NFT throughout the entire training process and improve the participants' attention and mental abilities, thereby improving the participants' physical ability and skill level.

Research Limitations and Improvements

Through comparative experiments, we studied the effects of two kinds of SP-NFT on training ability and brain regulation. However, based on the experimental conditions, the study had several limitations and could have benefited from some improvements:

(1) No active control group was set up. In this study, we divided the participants into two training groups and one passive control group. The control group only underwent shooting performance pre-test and post-test. Although such experimental settings have little effect on the research of trainability and neuroplasticity, there may be a placebo effect in the study of behavioral indicators. The control group did not undergo NFT, thus the observed effects may not be due to specificity of training in the given EEG frequency bands. In later studies, an active control group should be set up to avoid this problem.

(2) We used a single feedback interface. Although we have adopted the feedback mode of combined stimuli of sound and image, and further optimized the traditional feedback system, the same feedback images and music were applied in all six SP-NFT sessions, which may cause the participants to lose interest in the later stages of training, and thus, lead to a decline in the effect of SP-NFT.

(3) In the alpha rhythm NFT, we chose to increase the alpha power of the T3 channel and reduce that of the T4 channel. Compared with the SMR group, which increased the SMR power on the three channels at the same time, the signal-to-noise ratio (SNR) obtained by alpha training may be lower. Therefore, the experimental results that alpha training is more difficult than SMR training may also be affected by the SNR of the training paradigm. Subsequent research should take the SNR of the feedback signal and the difficulty of feedback training into consideration.

CONCLUSION

To compare the efficacy, trainability and neuroplasticity effects of SMR and alpha SP-NFT, 45 participants were recruited into the experiment, and 30 participants trained with SP-NFT during six sessions in 3 weeks, respectively. Through the analysis of the experimental results, the following main conclusions were obtained:

(1) By comparing the results before and after SP-NFT, we showed that the shooting performance of the SMR group after SP-NFT was significantly higher than that before SP-NFT, while the shooting performance of the Alpha group after SP-NFT was lower than that before SP-NFT. Thus, for non-professional shooters, enhanced SMR may be more effective to improve shooting performance than enhancing the alpha power difference in the left and right temporal regions.

(2) The results regarding neurofeedback trainability scores and feedback features show that the participants can master the SP-NFT technology correctly to improve the neurofeedback trainability scores and feedback features through training. By comparing the subjective task appraisal results, we found that SMR NFT was subjectively assessed as less difficult and was easier to master than alpha NFT.

(3) A comparison of the results of EEG before and after SP-NFT showed that the power spectrum of COI and that of the whole brain of the participants presented direct and specific changes, indicating that SP-NFT may promote neuroplasticity and has a continuous effect on resting EEG activity.

Overall, the results in this article provide some evidence of the effects of SP-NFT on trainability and neuroplasticity and further contribute to the application of SP-NFT to improve shooting performance stage.

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 Engineering University of the Chinese People's Armed Police Force ethics committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AG collected experimental data and wrote the original manuscript. WN analyzed experiment results. CJ and YF designed experiments and revised manuscript. EY provided important advice and help on key content of the manuscript.

FUNDING

REFERENCES

Babiloni, C., Marzano, N., Iacoboni, M., Infarinato, F., Aschieri, P., Buffo, P., et al. (2010). Resting state cortical rhythms in athletes: a high-resolution EEG study. *Brain Res. Bull.* 81, 149–156. doi: 10.1016/j.brainresbull.2009.10.014

Bertollo, M., Fronso, S. D., Filho, E., Conforto, S., Schmid, M., Bortoli, L., et al. (2016). Proficient brain for optimal performance: the MAP model perspective. *PeerJ* 4:e2082. doi: 10.7717/peerj.2082

Cheng, M. Y., Huang, C. J., Chang, Y. K., Koester, D., Schack, T., and Hung, T. M. (2015a). Sensorimotor rhythm neurofeedback enhances golf putting performance. *J. Sport Exerc. Psychol.* 37, 626–636. doi: 10.1123/jsep. 2015-0166

Cheng, M. Y., Hung, C. L., Huang, C. J., Chang, Y. K., Lo, L. C., Shen, C., et al. (2015b). Expert-novice differences in SMR activity during dart throwing. *Biol. Psychol.* 110, 212–218. doi: 10.1016/j.biopsycho.2015.08.003

Cho, M. K., Jang, H. S., Jeong, S. H., Jang, I. S., Choi, B. J., and Lee, M. G. (2007). α neurofeedback improves the maintaining ability of α activity. *Neuroreport* 19, 315–317. doi: 10.1097/wnr.0b013e3282f4f022

Collura, T. F. (2013). *Technical Foundations of Neurofeedback*. London: Routledge.

Del Percio, C., Babiloni, C., Bertollo, M., Marzano, N., Iacoboni, M., Infarinato, F., et al. (2009). Visuo-attentional and sensorimotor α rhythms are related to visuo-motor performance in athletes. *Hum. Brain Mapp.* 30, 3527–3540. doi: 10.1002/hbm.20776

Del Percio, C., Iacoboni, M., Lizio, R., Marzano, N., Infarinato, F., Vecchio, F., et al. (2011). Functional coupling of parietal α rhythms is enhanced in athletes before visuomotor performance: a coherence electroencephalographic study. *Neuroscience* 175, 198–211. doi: 10.1016/j.neuroscience.2010. 11.031

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003. 10.009

Doppelmayr, M., Finkenzeller, T., and Sauseng, P. (2008). Frontal midline theta in the pre-shot phase of rifle shooting: differences between experts and novices. *Neuropsychologia* 46, 1463–1467. doi: 10.1016/j.neuropsychologia.2007. 12.026

Enriquez-Geppert, S., Huster, R., Scharfenort, R., Mokom, Z. N., Zimmermann, J., and Herrmann, C. S. (2014). Modulation of frontal-midline theta by neurofeedback. *Biol. Psychol.* 95, 59–69. doi: 10.1016/j.biopsycho.2013. 02.019

Faller, J., Cummings, J., Saproo, S., and Sajda, P. (2019). Regulation of arousal *via* online neurofeedback improves human performance in a demanding sensory-motor task. *Proc. Natl. Acad. Sci. U S A* 116, 6482–6490. doi: 10.1073/pnas. 1817207116

Faridnia, M., Shojaei, M., and Rahimi, A. (2012). The effect of neurofeedback training on the anxiety of elite female swimmers. *Ann. Biol. Res.* 3, 1020–1028.

Ghaziri, J., Tucholka, A., Larue, V., Blanchette-Sylvestre, M., Reyburn, G., Gilbert, G., et al. (2013). Neurofeedback training induces changes in white and gray matter. *Clin. EEG Neurosci.* 44, 265–272. doi: 10.1177/1550059413 476031

Gong, A. M., Liu, J. P., Li, F., Liu, F., Jiang, C., and Fu, Y. (2017). Correlation between resting-state electroencephalographic features and shooting performance. *Neuroscience* 366, 172–183. doi: 10.1016/j.neuroscience. 2017.10.016

Gruzelier, J. H. (2014). EEG-neurofeedback for optimizing performance I: a review of cognitive and affective outcome in healthy participants. *Neurosci. Biobehav. Rev.* 44, 124–141. doi: 10.1016/j.neubiorev.2013.09.015

Hammond, D. C. (2011). What is neurofeedback: an update. *J. Neurother.* 15, 305–336. doi: 10.1080/10874208.2011.623090

Hatfield, B. D., Landers, D. M., and Ray, W. J. (1984). Cognitive processes during self-paced motor performance: an electroencephalographic profile of skilled marksmen. *J. Sport Psychol.* 6, 42–59. doi: 10.1123/jsp.6.1.42

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Hum. Brain Mapp.* 14, 166–185. doi: 10.1002/hbm.1050

Khanna, A., Pascualleone, A., Michel, C. M., and Farzan, F. (2015). Microstates in resting-state EEG: current status and future directions. *Neurosci. Biobehav. Rev.* 49, 105–113. doi: 10.1016/j.neubiorev.2014.12.010

Klimesch, W. (1999). EEG α and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/s0165-0173(98)00056-3

Kluetsch, R. C., Ros, T., Théberge, J., Frewen, P. A., Calhoun, V. D., Schmahl, C., et al. (2014). Plastic modulation of ptsd resting-state networks by EEG neurofeedback. *Acta Psychiatr. Scand.* 130, 123–136. doi: 10.1111/acps. 12229

Koch, S. P., Koendgen, S., Bourayou, R., Steinbrink, J., and Obrig, H. (2008). Individual α-frequency correlates with amplitude of visual evoked potential and hemodynamic response. *NeuroImage* 41, 233–242. doi: 10.1016/j.neuroimage. 2008.02.018

Konareva, I. N. (2005). Modifications of the EEG frequency pattern in humans related to a single neurofeedback session. *Neurophysiology* 37, 388–395. doi: 10.1007/s11062-006-0015-0

Landers, D. M., Petruzzello, S. J., Salazar, W., Crews, D. J., Kubitz, K. A., Gannon, T. L., et al. (1991). The influence of electrocortical biofeedback on performance in pre-elite archers. *Med. Sci. Sports Exerc.* 23, 123–129. doi: 10.1249/00005768-199101000-00018

Loze, G. M., Collins, D., and Holmes, P. S. (2001). Pre-shot EEG α-power reactivity during expert air-pistol shooting: a comparison of best and worst shots. *J. Sports Sci.* 19, 727–733. doi: 10.1080/02640410152475856

Megumi, F., Yamashita, A., Kawato, M., and Imamizu, H. (2015). Functional MRI neurofeedback training on connectivity between two regions induces long-lasting changes in intrinsic functional network. *Front. Hum. Neurosci.* 9:160. doi: 10.3389/fnhum.2015.00160

Mikicin, M., Orzechowski, G., Jurewicz, K., Paluch, K., Kowalczyk, M., and Wróbel, A. (2015). Brain-training for physical performance: a study of EEG-neurofeedback and α relaxation training in athletes. *Acta Neurobiol. Exp.* 75, 434–445.

Mirifar, A., Beckmann, J., and Ehrlenspiel, F. (2017). Neurofeedback as supplementary training for optimizing athletes' performance: a systematic review with implications for future research. *Neurosci. Biobehav. Rev.* 75, 419–432. doi: 10.1016/j.neubiorev.2017.02.005

Plotkin, W. B., and Rice, K. M. (1981). Biofeedback as a placebo: anxiety reduction facilitated by training in either suppression or enhancement of α brainwaves. *Exp. Brain Res.* 49, 590–596. doi: 10.1037/0022-006x.49. 4.590

Raymond, J., Sajid, I., Parkinson, L. A., and Gruzelier, J. H. (2005). Biofeedback and dance performance: a preliminary investigation. *Appl. Psychophysiol. Biofeedback* 30, 65–73. doi: 10.1007/s10484-005-2175-x

Ring, C., Cooke, A., Kavussanu, M., McIntyre, D., and Masters, R. (2015). Investigating the efficacy of neurofeedback training for expediting expertise and excellence in sport. *Psychol. Sport Exerc.* 16, 118–127. doi: 10.1016/j. psychsport.2014.08.005

Ros, T., Enriquez-Geppert, S., Zotev, V., Young, K., Wood, G., Whitfield-Gabrieli, S., et al. (2019). Consensus on the reporting and experimental design of clinical and cognitive-behavioural neurofeedback studies (CRED-nf checklist). *PsyArXiv*. doi: 10.31234/osf.io/nyx84

Rostami, R., Heidar, S., Kobra, K., Abadi, M. N., and Salamati, P. (2012). The effects of neurofeedback on the improvement of rifle shooters' performance. *J. Neurotherapy* 16, 264–269. doi: 10.1080/10874208.2012. 730388

Salazar, W., Landers, D. M., Petruzzello, S. J., Han, M., Crews, D. J., and Kubitz, K. A. (1990). Hemispheric asymmetry, cardiac response, and performance in elite archers. *Res. Q. Exerc Sport* 61, 351–359. doi: 10.1080/ 02701367.1990.10607499

Schönenberg, M., Wiedemann, E., Schneidt, A., Scheeff, J., Logemann, A., Keune, P. M., et al. (2017). Neurofeedback, sham neurofeedback, and cognitive-behavioural group therapy in adults with attention-deficit hyperactivity disorder: a triple-blind, randomized, controlled trial. *Lancet Psychiatry* 4, 673–684. doi: 10.1016/s2215-0366(17)30291-2

Strizhkova, O., Cherapkina, L., and Strizhkova, T. (2012). Neurofeedback course applying of high skilled gymnasts in competitive period. *J. Hum. Sport Exerc.* 7, S185–S193. doi: 10.4100/jhse.2012.7.proc1.21

Vernon, D., Frick, A., and Gruzelier, J. (2004). Neurofeedback as a treatment for ADHD: a methodological review with implications for future research. *J. Neurother.* 8, 53–82. doi: 10.1300/j184v 08n02_04

Wan, F., Nan, W. Y., Vai, M. I., and Rosa, A. (2014). Resting α activity predicts learning ability in α neurofeedback. *Front. Hum. Neurosci.* 8:500. doi: 10.3389/fnhum.2014.00500

Xiang, M. Q., Hou, X. H., Liao, B. G., Liao, J. W., and Hu, M. (2018). The effect of neurofeedback training for sport performance in athletes: a meta-analysis. *Psychol. Sport Exerc.* 36, 114–122. doi: 10.1016/j.psychsport.2018.02.004

Zhang, R., Xu, P., Chen, R., Li, F., Guo, L., Li, P., et al. (2015). Predicting inter-session performance of SMR-based brain-computer interface using the spectral entropy of resting-state EEG. *Brain Topogr.* 28, 680–690. doi: 10.1007/s10548-015-0429-3

Zhou, G., Liu, P., He, J., Dong, M., Yang, X., Hou, B., et al. (2012). Interindividual reaction time variability is related to resting-state network topology: an electroencephalogram study. *Neuroscience* 202, 276–282. doi: 10.1016/j.neuroscience.2011.11.048

Zoefel, B., Huster, R. J., and Herrmann, C. S. (2011). Neurofeedback training of the upper α frequency band in EEG improves cognitive performance. *NeuroImage* 54, 1427–1431. doi: 10.1016/j.neuroimage.2010.08.078

A Psychophysiological Model of Firearms Training in Police Officers: A Virtual Reality Experiment for Biocybernetic Adaptation

John E. Muñoz[1], Luis Quintero[2]*, Chad L. Stephens[3] and Alan T. Pope[3,4]

[1] *Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada,* [2] *Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden,* [3] *Langley Research Center, National Aeronautics and Space Administration, Hampton, VA, United States,* [4] *Learning Engagement Technologies, Poquoson, VA, United States*

Crucial elements for police firearms training include mastering very specific psychophysiological responses associated with controlled breathing while shooting. Under high-stress situations, the shooter is affected by responses of the sympathetic nervous system that can impact respiration. This research focuses on how frontal oscillatory brainwaves and cardiovascular responses of trained police officers ($N = 10$) are affected during a virtual reality (VR) firearms training routine. We present data from an experimental study wherein shooters were interacting in a VR-based training simulator designed to elicit psychophysiological changes under easy, moderate and frustrating difficulties. Outcome measures in this experiment include electroencephalographic and heart rate variability (HRV) parameters, as well as performance metrics from the VR simulator. Results revealed that specific frontal areas of the brain elicited different responses during resting states when compared with active shooting in the VR simulator. Moreover, sympathetic signatures were found in the HRV parameters (both time and frequency) reflecting similar differences. Based on the experimental findings, we propose a psychophysiological model to aid the design of a biocybernetic adaptation layer that creates real-time modulations in simulation difficulty based on targeted physiological responses.

Keywords: biocybernetic adaptation, virtual reality, psychophysiological model, electroencephalography, heart rate variability, simulation, firearms training

INTRODUCTION

Burdea and Coiffet (2003) identified "traditional" areas of application of virtual reality (VR) as medicine, education, arts, entertainment, and the military as well as "emerging" areas of manufacturing, robotics and data visualization, training being a particular focus area in many of these applications. For use in training, VR environments provide advantages over physical training environments. VR training in medical surgery contexts has shown performance improvements beyond traditional training approaches (Lehmann et al., 2005). The use of immersive training systems taps into gross and fine motor skill acquisition, maintenance, and expert-level performance (Faria et al., 2018). Firearms training is considered an especially appropriate setting for VR technology deployment. Due to the safety concerns associated with live fire weapons training, the United States Department of Defense places a high value on the potential use of VR environments for firearms training for service members who are impaired by polytrauma (Oliver et al., 2019).

Under high-stress situations, the shooter is affected by responses of the sympathetic nervous system that can impact respiration. Relaxing the body and keeping a natural breathing pattern have been identified as major components of firearms training (Johnson, 2007). A training guide well known to the professional police officer trainer community asserts that particular sympathetic responses are desired during military fighting situations [e.g., 100–115 BPMs for heart rate (HR) levels] (Mason, 1998). According to marksmanship training guidelines (Johnson, 2007), an important factor that needs to be trained in overall marksmanship scenarios is shooting during the natural respiratory pause. Since the lack of oxygen might disturb the performance of cognitive skills and visual acuity, training autogenic breathing (autonomic self-regulation training) is considered as a main component of firearms training (Sajnog, 2013), which will calm the body and keep a natural breathing pattern. In order to maintain both body and mind collected, proper oxygenation could be provided by controlling the optimal respiration pace and peripheral responses such as HR and HR variability (HRV).

To support investigation of physiological self-regulation in police officers while training in a target-shooting scenario, a fully immersive, VR-based training simulator called Biocyber Physical Simulator (BioPhyS) for firearms training has been created (Muñoz et al., 2016b). BioPhyS is an example of a system that is capable of employing a form of physiological computing known as biocybernetic adaptation wherein real-time data from the brain and body is used by a control loop to adapt the user interface (Ewing et al., 2016). The design of the BioPhyS system relies on investigations of physiological responses associated with the psychophysiological responses of firearm trainees during a targeted state associated with concentration and calmness.

The research reported here focuses on how frontal oscillatory brainwaves and cardiovascular responses of experienced police officers are affected during a head-mounted display-VR (HMD-VR) training routine. This article uses insights from a previously reported research study (see Muñoz et al., 2019, for more details) to investigate how brain and heart of police officers react to a simulated training for firearms use. Particularly, this research is focused on:

- *Quantifying the neurophysiological and cardiovascular responses* of a group of trained police officers while interacting with a controlled and immersive virtual simulation of shooting practice.
- Investigating *perceived motion sickness and user experience* of the police officers after being immersed in a VR simulator for firearms training while wearing physiological sensors.
- Establishing a *psychophysiological model of firearm training for real police officers* while using immersive HMD-VR and non-intrusive physiological sensors. The model considers specific metrics from electroencephalographic (EEG) and cardiovascular sensors (e.g., HR and HRV) and how those metrics vary when the difficulty in the simulator was modulated.

- Illustrating a *streamlined pipeline for the integration of physiological adaptation into VR* simulators using novel tools such as biocybernetic software technologies and biofeedback design elements.

RELATED WORK

Lele (2013) points to the quality of immersion as the basis for a VR system's usefulness. Psychophysiological measurement has increasingly been incorporated into research using HMD-VR setups (Pugnetti et al., 2001), particularly in studies examining presence and immersion. Cortical, as well as cardiovascular measures, have relevance for characterizing the psychophysiological response of individuals experiencing VR environments.

In one study, service members with polytrauma demonstrated improved accuracy and precision following VR-based firearm training (Oliver et al., 2019), due to its capability to measure metrics that are not measured in a traditional qualification course, and allowing instructors to focus on other aspects to enhance shooter precision such as posture, sight alignment, or the elimination of bad habits. The United States Navy has developed the tactically reconfigurable artificial combat enhanced reality (TRACER) system to train sailors for combat. TRACER is extolled as a dynamic, engaging and less predictable virtual training environment (Iatsyshyn et al., 2019). In an analysis of the development status and characteristics of VR technology in China (Zhang et al., 2019), it is asserted that VR be beneficial in improving the psychological wellbeing of military personnel and help soldiers to adapt to various war environments. In this regard, exposure therapy based on VR has been extensively showed as an efficacious treatment for active-duty army soldiers with post-traumatic stress disorder (Rizzo and Shilling, 2017). EEG has been employed in investigations of various cognitive variables in 3D virtual learning environments. Frontal alpha EEG was recorded in a study of its role in attentional control; the study also demonstrated the importance of the immersion and engagement afforded by a 3D virtual learning environment (Berger and Davelaar, 2018). In a study of the effect of competition, while shooters were immersed in a virtual environment representing a shooting range, changes in alpha oscillatory activity were found during aiming that were associated with better performance (Pereira et al., 2018). In a previous study by the authors (Muñoz et al., 2016b), frontal alpha activity (as well as delta) was found to be a brainwave pattern that allows differentiation between baseline and active shooting states in a VR simulator. Increases in alpha EEG activity have also been associated with subjects who performed well in spatial navigation tasks in a VR environment (Pugnetti et al., 1996).

Virtual reality experiences can also affect the cardiovascular responses of users. A significant difference in HR values was found during 5-min long interactions with a VR experience that required users to perform simple manual tasks involving the arrangement of virtual elements with multiple shapes (Malińska et al., 2015). A recent study (Marín-Morales et al., 2019) employed analyses of EEG and electrocardiographic (ECG)

signals to explore differences in the emotional reactions (arousal and valence dimensions) between an exploration in a real-museum and its VR representation, and demonstrated high accuracy of a machine learning model in classifying the nature of a stimuli as real or virtual.

Machine learning approaches have also been used in psychophysiological studies. Specifically, the use of supervised classifiers [e.g., linear discriminant analysis (LDA) and support vector machines (SVM)] to categorize emotional states and create a model for audio-visual or game difficulty adaptation (Novak et al., 2012). Other studies have utilized multilayer perceptron to classify anxiety, boredom, and flow states to compare the effects of mental-state adaptation and performance-based adaptation in a shooting desktop-based game (Alves et al., 2018). A custom-made version of the conventional Tetris was adapted to create real-time adjustments of the game difficulty in three levels by using a SVM classifier that processed signals from skin response, blood volume and EEG (Chanel et al., 2011). Furthermore, learning-based classifiers have also been used to detect high and low anxiety in drivers from ECG and accelerometer data (Dobbins and Fairclough, 2018), or to create a virtual driving platform to maximize engagement in people with autism spectrum disorder (Bian et al., 2019). Applications for airplane pilots used classifiers to identify features from EEG and skin response signals that can model the users in scenarios of attention-related human performance limiting states (Harrivel et al., 2016) or to find relationships of cardiovascular features with psychophysiological stress while performing piloting maneuvers (Hanakova et al., 2017). To the best of our knowledge, this is the first project that aims at characterizing psychophysiological responses of police officers on duty for designing biocybernetic loops in VR firearms training.

VIRTUAL REALITY FOR FIREARMS TRAINING

In this section, we describe the design and development of a fully immersive, HMD-VR based simulator for firearms shooting training as well as briefly introduce a software tool, the Biocybernetic Loop (BL) Engine, used to integrate physiological intelligence to the VR simulator.

Biocyber Physical Simulator System for Firearms Training

The BioPhyS contains an outdoor military target-shooting range with representative props such as cable reels, wooden tables, barricades, weapons, and water towers (see **Figure 1**, left). Twenty targets were laid out at different distances from the shooting point; each target moves along a horizontal track. Three weapons were used for training: The Pistols M1911 and SIG Sauer P250, and the Reichsrevolver M1879 (see **Figure 1**, right), each with different impact force on the targets.

A setup panel called Wizard of Oz (WoZ) was designed to allow trainers to modify in real-time the conditions of the simulation; it also provided a controlled environment to study the participants' responses and behaviors when exposed to specific

stressors. The set of variables was defined to create physiological modulation and were grouped according to its effect on the scenario, as shown in **Table 1**.

BioPhyS can be used with wireless HTC Vive controllers and an air pistol gun adapted with the HTC Vive Tracker to handle and shoot the weapon. The touchpad of the controllers offers teleportation to navigate around the shooting range, the lateral buttons are used to grab the virtual gun, and the trigger is used for shooting.

Physiological Intelligence and the Biocybernetic Loop Engine

Conventional methods to integrate physiological sensing technologies into games and VR applications entail the development of specific software clients that stream data collected from the sensors and capturers able to read the data packages directly in game engines such as Unity3D (Muñoz et al., 2016a). Additionally, scripts for signal processing and data analysis are required to create truly intelligent algorithms able to finally integrate biocybernetic adaptation. To streamline the process, the BL Engine is a software tool that acts as a middle layer facilitating the integration of physiological intelligence to games and VR applications developed in Unity3D (Muñoz et al., 2017b). By allowing (i) communication with multiple physiological sensors, (ii) a drag-and-drop console to create adaptive rules, and (iii) specialized scripts to communicate applications and modify variables in real-time; the BL Engine is a software tool to simplify the integration of biocybernetic adaptation in VR projects. This article outlines our initial stage of designing the biocybernetic adaptation layer after carrying out a physiological characterization study in police officers while interacting with the BioPhyS with controlled difficulty levels.

PSYCHOPHYSIOLOGICAL CHARACTERIZATION STUDY WITH POLICE OFFICERS

The BioPhyS was used to conduct a controlled study with police officers using a repeated measures design, to understand the cardiovascular and neurophysiological responses under different simulation difficulties.

System Setup

A room-scale tracking system of the VR headset HTC Vive Pro was used to provide a fully immersive experience. The users were able to walk in an area of up to 12 square meters (maximum 5 m between both tracking lighthouses) and interact with the virtual environment by using one wireless controller. A VR One MSI backpack (VR ready computer) computer to run BioPhyS, and an additional screen was used to configure the scenarios and to mirror the VR simulation (see **Figure 2**).

Participants

Ten police officers from the police division of Hampton city (VA, United States) were recruited for the study. Participants were
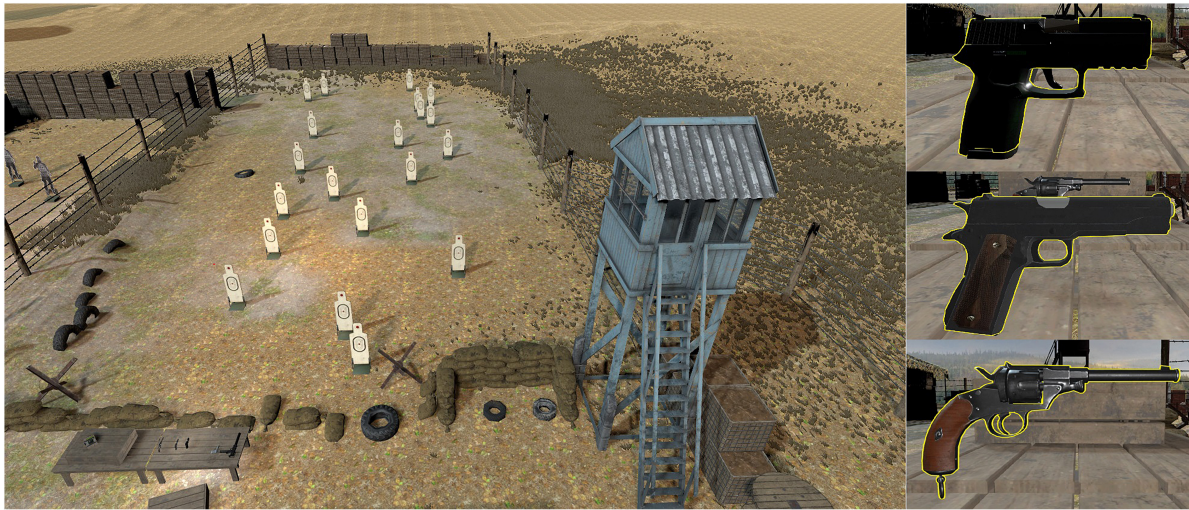
FIGURE 1 | Screenshots of the BioPhyS virtual simulator created in VR. **Left:** top-view of the shooting range with the targets placed randomly and a table with the guns. **Right:** view of the grabbable weapons that can be shot during the simulation.

mostly males (nine males and one female) with ages from 21 to 43 years old. The participation in the study was advertised by the police department as voluntary and the inclusion criteria was having past experience with real target shooting. The experiment was described as a playtest and the details were given to the participants before starting; every police officer also signed informed consent. **Table 2** summarizes the characteristics of the police officers who participated in the study.

Physiological Metrics

Police officers were wearing both wearable EEG and HR monitors during the training session using the HMD-VR system. The signals were synchronously recorded on a secondary computer.

Electroencephalography

Brainwave activity was recorded using the wearable headband Muse BCI. The sensor includes four EEG electrodes in the TP9, Fp1, Fp2, and TP10 channel positions following the 10–20 standards. The device generates samples at a frequency of 500 Hz and contains proprietary algorithms that compute relevant parameters to quantify brain activity patterns such as the oscillatory rhythms (Sanei and Chambers, 2013) delta (δ, 1–4 Hz), alpha (α, 8–12 Hz), beta (β, 12–30 Hz), theta (θ, 4–8 Hz),

and gamma (γ, 30–100 Hz). The proprietary software Muse Lab[1] was used to compute the power spectral density (PSD) of the EEG raw data for each channel for a frequency range from 0 to 110 Hz, using a Hamming window with a window-length of 256 samples and 90% overlap. The EEG metrics that were more relevant for exploration were:

- Absolute Bandpowers: which use a logarithmic function with the sum of the PSD of the EEG data in a specific frequency range. The absolute α, β, ∂, θ, and γ bandpowers were computed for the frontal electrodes of the Muse BCI sensor.
- EEG indexes: three different indexes were explored considering previous studies (Kamzanova et al., 2011; Fairclough and Gilleade, 2012; McMahan et al., 2015). First, the Engagement index was computed by using the ratio of beta to alpha + theta. The frontal asymmetry index was computed by subtracting the Alpha in Fp2 (right hemisphere) with the Alpha in Fp1 (left hemisphere). Finally, the Theta/Beta ratio was also computed to see the neurophysiological responses in those two bandpowers.

Cardiovascular

The chest strap sensor Polar H10 was used to record the cardiac responses of police officers. It includes built-in algorithms that were used to calculate ECG parameters needed for the HRV analysis. The main variable computed is the R-to-R intervals (RRI) with units of 1/1024 seconds.[2] HR and RRIs are broadcast by the sensor and saved locally using a custom-made client running on Windows and based on a Bluetooth Low Energy Windows API. The PhysioLab toolbox (Muñoz et al., 2017a) was used to compute both time and frequency domain HRV parameters. Extracted features include the standard deviation

TABLE 1 | List of simulation variables carefully defined to create the physiological modulation in the BioPhyS.

Simulation effect	Simulation variable	Range
Targets	Number of targets	3–20 (targets)
	Target size	0.3–3 (units)
	Target speed (horizontal)	0–5 (m/s)
	Target hardness	1–20 (units)
Simulation environment	Day light	0–10 (units)
	Rain intensity	0–1 (units)

---

[1]https://sites.google.com/a/interaxon.ca/muse-developer-site/muselab

[2]https://support.polar.com/en/support/Selection_Info_Analysis_for_R_R_data

FIGURE 2 | System setup for the study describing the equipment used.
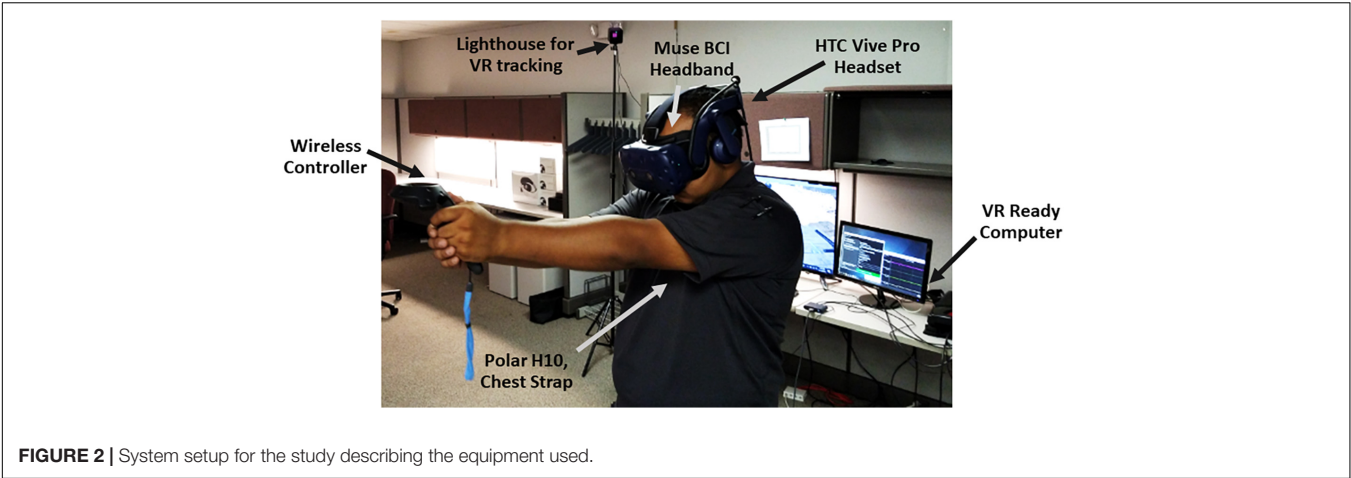
TABLE 2 | Police officer's demographics and firearm training.

User ID	Age	Gender	Target shooting experience	Military veteran	Time as LEO	Firearms qualifications requirement	Specialized weapons training
01	37	Male	>3 years	No	17 years	Bi-Annual	SWAT, Firearms Instructor
02	37	Male	>3 years	Yes	3 years	Annual, Bi-Annual	None
03	43	Male	>3 years	Yes	In Academy	Annual, Bi-Annual	None
04	21	Male	<1 year	No	In Academy	Bi-Annual	None
05	32	Male	>3 years	No	11 years	Annual, Bi-Annual	Firearms instructor, SWAT, Patrol rifle
06	26	Female	>3 years	No	3 years	Bi-Annual	Firearms instructor
07	31	Male	>3 years	Yes	In Academy	Bi-Annual	None
08	33	Male	>3 years	No	15 years	Bi-Annual	Firearms instructor, SWAT
09	27	Male	<1 year	No	In Academy	Bi-Annual	None
10	23	Male	<1 year	No	In Academy	Bi-Annual	None

LEO, Law Enforcement Officer; SWAT, special weapons and tactics.

of the RRIs (SDNN) and the root mean square of successive differences (RMSSD) values from the time domain. Similarly, frequency domain parameters included high frequency (HF, 0.15–0.40 Hz), low frequency (LF, 0.04–0.15 Hz), and very low frequency (VLF, 0.0033–0.04 Hz), which were extracted from the PSD of the RRI signal. The PSD is computed by using a Welch estimator with a Hanning window, and spectrum components are averaged by an area-under-the-curve approach. The polar chest strap has shown acceptable performance in calculating HR and RRI measurements under different scenarios including non-resting situations (Plews et al., 2017).

Other Measurements
Simulation Performance
The BioPhyS computed the participants' performance during firearm shooting training. It recorded specific simulation variables such as the total amount of shot bullets, number of destroyed targets, and headshots. The shooting performance metric was defined as the ratio between the number of destroyed targets and the shot bullets.

Simulation Sickness
The level of motion sickness produced during interaction with the VR system was evaluated using the Simulator Sickness Questionnaire (SSQ; Kennedy et al., 1993). It assesses

total severity of simulation sickness supported by three main symptom clusters called oculomotor (eyestrain, difficulty focusing, blurred vision, and headache), disorientation (dizziness and vertigo), and nausea (nausea, stomach awareness, increased salivation, and burping).

Post-session Subjective Interview
Police officers were briefly interviewed at the end of the session to gather their reactions in terms of (i) overall user experience, (ii) pros and (iii) cons of training with the BioPhyS, and (iv) envisioned improvements.

Experimental Protocol
Training Scenarios
The training scenarios were jointly designed together with a military veteran of the research team. The training protocol included three difficulty modes: easy, medium, and hard; each of them lasting 3 min. The easy configuration laid out 10 static targets randomly distributed in the 10 first horizontal tracks of the training scenario. The medium setup used 10 moving targets at a speed of 0.5 m/s. The hard scenario used 20 targets moving at 1 m/s and it was specifically designed to provoke frustration. Both the target size and hardness were maintained constant across the difficult levels. The same weapon Reichsrevolver was used during the experiment setup with a shooting power equals to

the hardness of the target, thus targets were instantly destroyed with one shot. Although unrealistic, the revolver was preferred considering its slow response, thus aiding a more careful aiming and shooting instead of less mentally and physically prepared training. A baseline condition was used to record physiological signals from the police officers during a passive stand-up situation, wearing the VR headset and physiological sensors, and holding the virtual weapon without shooting or interacting with the virtual environment.

Procedure

The study protocol was reviewed for ethical treatment of human subjects and approved by the Office of the Chief of Police of the City of Hampton, VA, serving as the research ethics committee. All police officer subjects volunteered to participate and gave written informed consent in accordance with the Declaration of Helsinki. The experiment lasted around 40 min per participant, including questionnaires, the connection of sensors and interaction. The informed consent was signed at the beginning of the experiment together with a short demographics form. Then, the sensors were connected to the police officer's forehead and chest; the VR headset was also worn and participants were given the instruction to shoot some targets before starting to verify that the protocol was understood properly. Baseline physiological measurements were taken for 3 min, emphasizing that users needed to avoid any facial expression, such as speaking or visual navigation while moving the head. During the active shooting moments, police officers were instructed to also avoid squinting the eyes for aiming in order to minimize signal artifacts in the EEG sensor caused by facial movements. The easy-medium-hard scenarios were manually set up by the researcher, the participants interacted with them for a 2-min resting period between sessions. Finally, users filled out the SSQ and the short post-session interview.

Data Analysis

Collected physiological data was processed offline using MatLab (v2013b). Individual cardiovascular and EEG parameters were computed and averaged for statistical analysis. EEG bandpowers and index analysis initially explored frontal Fp1 and Fp2 electrodes separately; however, frontal lobe activity was ultimately weighted by averaging the contribution of both electrodes. Data normality was checked using Kolmogorov-Smirnoff tests. Data with normal distribution were statistically analyzed using parametric tests, whereas non-parametric tests were used for non-normal distributions to determine the influence of the simulation difficulty as the main effect. *Posthoc* analysis using Bonferroni correction was used to follow up the findings.

RESULTS

Physiological Responses
Electroencephalography

Electroencephalography data recorded from the frontal lobe using the Fp1 and Fp2 electrodes of the Muse BCI system revealed significant brain activity patterns associated with the

different difficulty levels used in the BioPhyS. Two specific EEG patterns showed significant results. Firstly, the frontal theta levels (θ), $\chi^2(3) = 10.21$, $p < 0.05$, showed significant changes across simulation difficulty levels. Wilcoxon tests were used to follow up initial findings while Bonferroni adjustments were applied, so all effects are reported at a 0.0125 level of significance. Results revealed that frontal theta (θ) activity differed significantly from *easy* to *hard* difficulty levels ($Z = -1.58$, $p = 0.009$) for the police officers (see **Figure 3**). Secondly, although the theta/beta ratio was significantly influenced by the simulation difficulty factor, $\chi^2(3) = 10.69$, $p < 0.05$, non-significant differences were found with the Wilcoxon test.

Cardiovascular Responses

Cardiovascular responses were quantified via HR and HRV parameters. HRs of police officers were significantly affected by the simulation difficulty, $\chi^2(3) = 18.84$, $p < 0.05$. *Post hoc* tests (Wilcoxon with Bonferroni) showed significant differences between the baseline measurements of HR compared with the *easy* ($Z = -2.80$, $p = 0.005$), *medium* ($Z = -2.80$, $p = 0.005$), and *hard* ($Z = -2.80$, $p = 0.005$), difficulty levels (see **Figure 4**).

The HRV analysis showed similar results for both time and frequency domains. RMSSD, $\chi^2(3) = 18.48$, $p < 0.05$, and VLF, $\chi^2(3) = 18.84$, $p < 0.05$; values were significantly affected by the simulation difficulty. Particularly the differences between baseline and *easy* ($Z = -2.80$, $p = 0.005$), *medium* ($Z = -2.80$, $p = 0.005$), and *hard* ($Z = -2.80$, $p = 0.005$) were significant in the *post hoc* analysis.

Simulation Performance

Simulation (or shooting) performance was computed as the ratio between the number of destroyed targets and the shot bullets. Simulation difficulty significantly influenced the shooting

FIGURE 3 | Theta brainwave patterns analyzed for the baseline, easy, medium, and hard difficulties in the BioPhyS. Asterisk (*) denotes significant results following Wilcoxon tests.

FIGURE 4 | Heart rate (HR) levels affected during the training with the BioPhyS. Asterisk (*) denotes significant differences following Wilcoxon tests.
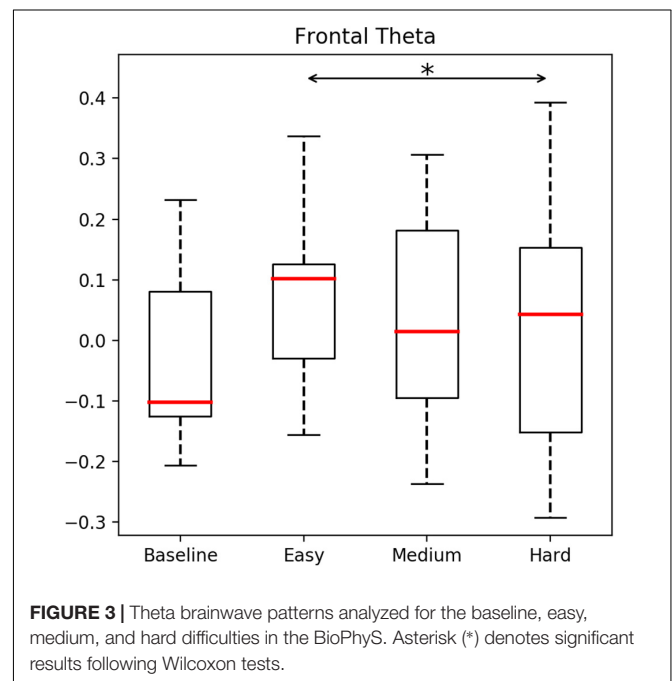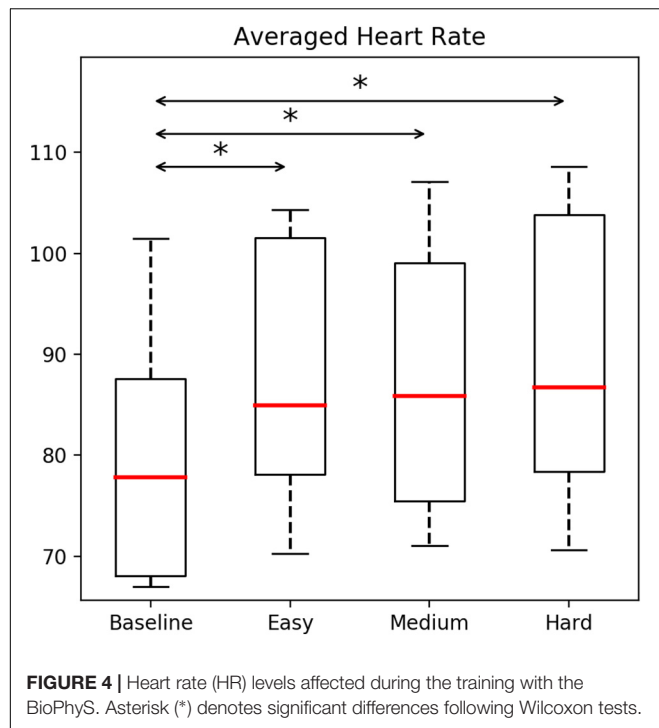
performance of police officers, $\chi^2(3) = 18.84$, $p < 0.05$, revealing 64, 46, and 24% performance for the *easy*, *medium*, and *hard* difficulties, respectively.

Simulation Sickness

Simulation sickness was measured immediately following the experience by asking police officers about their physical and cognitive status using the SSQ. Data from one of the users was discarded due to recording errors. Test subjects reported eight or fewer SSQ symptoms which is categorized as minimal symptoms, indicating the VR experience did not impact users' operation of the system (Kennedy et al., 2001; Saredakis et al., 2020). Further analysis of the SSQ was conducted and cut-off scores were defined based on the 75th percentile of the calibration sample, to represent the majority of users in the test sample, to detect whether users were adversely affected by exposure to the virtual reality scenario (Kennedy et al., 1993). The reference scores are 15 for total severity, the thresholds for the subscales are 9.5 for nausea, 15.2 for oculomotor, and 0 for disorientation. Results revealed that three police officers were having sickness score above the threshold in all four aspects, and a total of six participants had high scores in the subscales of nausea and disorientation. Eye strain and difficulty focusing were reported by half of the police officers at the end of the training, although they never reached severe intensities.

Additionally, the mean and standard deviation were computed for the three distinct symptom clusters and the total score for the SSQ, and the relative severity was calculated by comparison with the nine calibration simulators utilized in the questionnaire. Results, depicted in boxplot format in **Figure 5**, indicate SSQ total severity (M = 15.4, SD = 11.1), nausea (M = 8.5,

SD = 7.4), oculomotor (M = 16.8, SD = 12.4) and disorientation (M = 13.9, SD = 13.9) that fall between the first and second simulators with higher sickness. Therefore examination of the data and further development of the simulator is necessary to reduce effects of the VR simulation used in this study on users.

Subjective Interview

Overall user experience was described positively by all police officers participating in the study. Two participants emphasized that the simulation made them focus on breathing and concentrate on remaining calm while shooting. Five participants described the experience as "good" and "cool" while highlighting their enthusiasm for trying VR for the first time. One subject mentioned that "graphics were really good" while others mentioned being "stressful as things became more difficult." As advantages (pros) of the simulation system, three participants liked the simplicity of the setup, mentioning that the sensors were minimally invasive and comfortable. Three participants highlighted the realism of the simulation and the accuracy of the motion tracking. The possibility to personalize the training scenarios was mentioned by one of the users as the most remarkable advantage of the system. As disadvantages (cons) of the simulation system, five police officers mentioned being concerned about the sights and the aiming issues resulting from not allowing them to squint their eyes. Additionally, three officers reported some blurriness in the HMD. Three participants also mentioned mapping problems between the controllers and the virtual weapon since "the controller not as much like a weapon as would like." Finally, police officers proposed several features

FIGURE 5 | Simulator Sickness Questionnaire (SSQ) scores reported by test subjects for nausea (M = 8.5, SD = 7.4), oculomotor (M = 16.8, SD = 12.4), disorientation (M = 13.9, SD = 13.9), and total severity (M = 15.4, SD = 11.1).
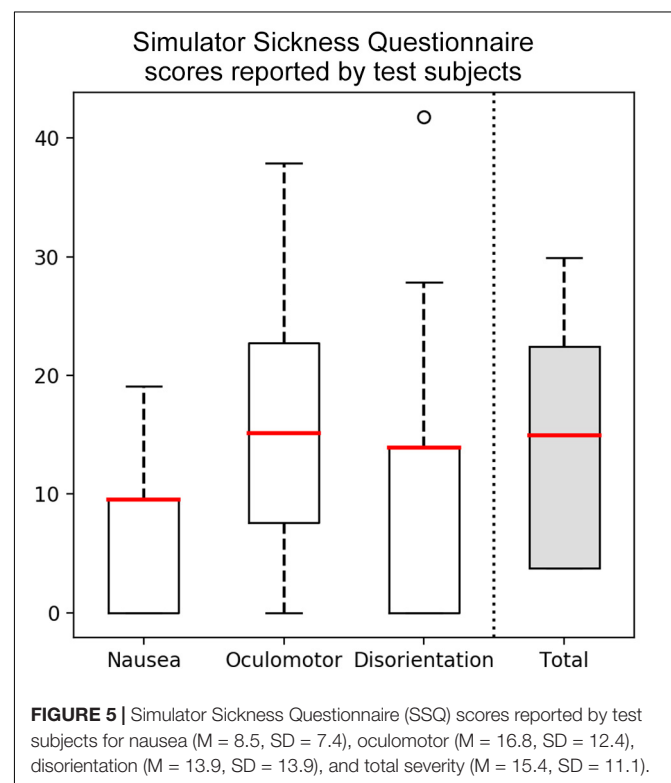
that can be implemented to improve the system, from which we highlight the following ones: (i) change controllers for air pistols (or more realistic devices), (ii) allow users to see sights of the weapon to improve shooting accuracy, and (iii) integrate scenario-driven interactions.

Finally, the four police officers trained as firearms instructors endorsed the potential of the system for firearm training, the simplicity of the setup including the HMD-VR headset and physiological sensors and the training scenario personalization.

THEORETICAL MODEL TO INTEGRATE BIOCYBERNETIC ADAPTATION

Psychophysiological Model

A psychophysiological model based on empirical findings under various levels of challenge is an integral part of the design of a biocybernetic adaptation layer that creates real-time modulations in simulation difficulty based on targeted physiological responses. The resulting psychophysiological model derived for the trained police officer population in this study is used to inform the biocybernetic adaptation approach that can be integrated into the simulation system by using the BL Engine. With the BioPhyS system, this model is developed by characterizing the difference between resting states when compared with active shooting states in a VR simulator in both brain responses and sympathetic signatures. This characterization helps specify the values of physiological variables to be targeted in the biocybernetic adaptation system as well as serve as a control measure for assessing the effects of adding a biocybernetic adaptation layer to a VR training system. Considering the results from the characterization study, we created a psychophysiological model (see **Figure 6**) that reveals opportunities to integrate biocybernetic adaptation in the BioPhyS using the police officers' data.

Preliminary Computational Model

The simulation difficulty in the BioPhyS covered easy, medium and hard scenarios. The design of those scenarios intended to evoke responses from police officers during specific human states such as minimal engagement (simple challenge considering their skills), engagement (balanced challenge/skill), and frustration (tough challenge considering their skills). The captured neurophysiological and cardiovascular responses from police officers across the simulation difficulty levels reflect specific physiological signatures that must be used to define the automatic adaption intended. Firstly, two specific EEG metrics showed statistical significance across simulation difficulties, frontal theta and theta/beta ratio. However, for real-time adaptation, frontal theta is preferred since it can be captured directly from the sensor without any further computation effort (e.g., dividing one bandpower by other). Secondly, cardiovascular responses also revealed significant differences in HR and RMMSD and VLW for HRV. HR is also preferred considering computational efficiency issues for real-time adaptation. Additionally, one of the firearms instructors of the research team pointed out literature suggesting specific sympathetic responses that are desired for trainees

to elicit during the training. Literature research revealed the 100–115 BPMs range as the targeted HR zone defined by firearm trainers to stress the heart enough to facilitate timely cognitive responses without hampering fine motor skill performance (Mason, 1998). Therefore, we used frontal theta as concentration (mental focus) and HR as calmness (cardiovascular regulation) metrics to create an adaptive rule in the BL Engine. The rule is shown schematically in **Figure 7**. The implementation allows the rules to be easily modified by dragging-and-dropping specific blocks that allow connecting logical instructions.

The rule involves capturing physiological data from data receiver blocks and averaging it using the array buffer and math array blocks. Averaged physiological data is then compared against the threshold values that are inferred from the psychophysiological model to finally create the changes in the simulation based on the detected values. For instance, this rule creates modulations toward increasing simulation difficulty (e.g., increasing targets speed and reducing target size) if police officers are not reaching the intended targeted values for getting them engaged (e.g., −0.1 dB frontal alpha and 100 BPMs for HR). Additionally, data from the BL Engine can be sent to the BioPhyS to provide a biofeedback display (panel at the right side of **Figure 8**) for the police officers and investigate the effects of implicit and explicit feedback on training performance (Kuikkaniemi et al., 2010).

Although controlled experiments to validate the real-time model have not been carried out yet, the proposed model and computational solution serve as an initial asset for the future integration of biocybernetic adaptation.

DISCUSSION

We present a repeated-measures experiment of firearms training using an HMD-VR setup and neurophysiological and cardiovascular measurements to model the psychophysiological responses of 10 ($n = 10$) police officers to different simulation difficulty levels. To induce specific graded stressors in the simulation, the system was designed to manually modulate specific parameters in the VR simulator (e.g., target speed). Three different scenarios (*easy*, *medium*, and *hard*) were used to investigate the police officer's physiological and subjective responses while actively shooting in the simulation and results were compared including measurements from resting states. Results show how police officers' brainwaves and cardiovascular responses were significantly affected by modulations in the simulation difficulty. Particularly, frontal theta values were significantly different between *easy* and *hard* difficulty levels whereas HR and HRV data (RMSSD and VLF) were also significantly different during resting once compared with the *easy*, *medium*, and *hard* difficulties.

Police Officers and Cognitive Readiness

Crucial elements for police firearms training include mastering very specific psychophysiological responses associated with controlled breathing while shooting. Novel immersive VR
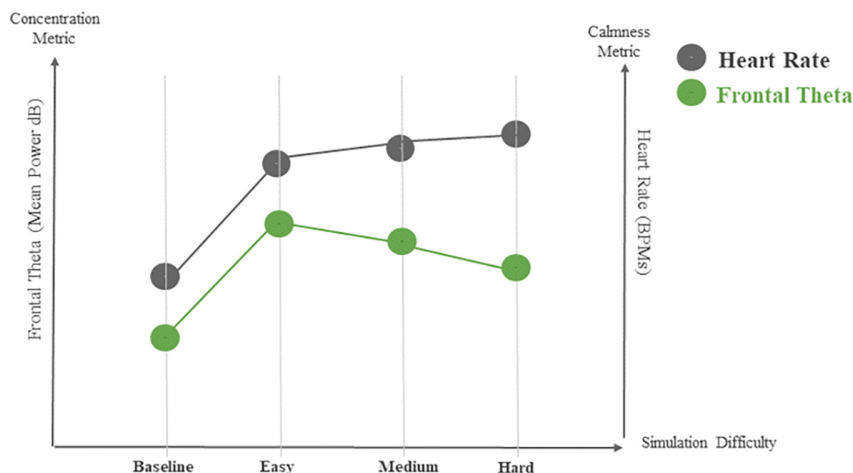
FIGURE 6 | Psychophysiological model for the BioPhyS considering EEG and HR data.
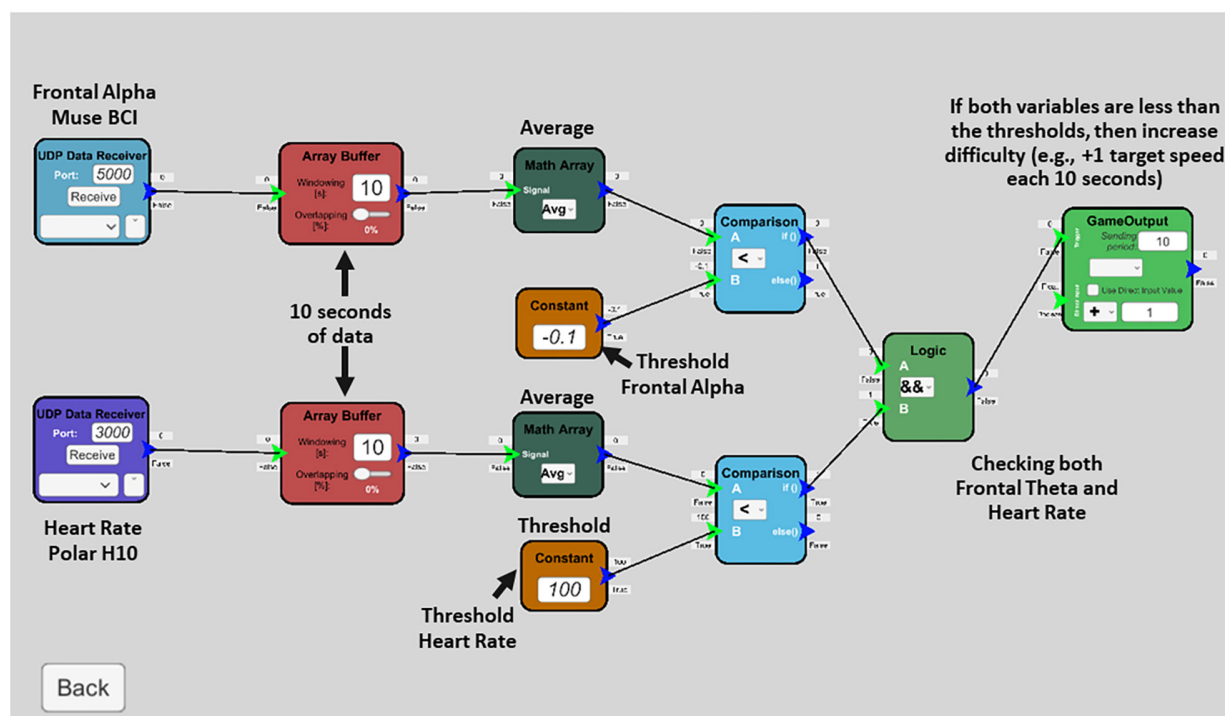
FIGURE 7 | Adaptive rule created in the BL Engine based on the psychophysiological model.

systems as the BioPhyS, allow the recreation of realistic scenarios where trainees can be exposed to different stressors while physiological responses are recorded. We showed how frontal theta values were significantly higher during the simulation difficulties that were not created to frustrate the police officers. Decrements in theta (θ) values have been commonly associated with non-attentional states such as drowsiness, severe sleepiness (Liu et al., 2009), or lower levels of arousal (Sanei and Chambers, 2013). Thus, controlled modulations toward increasing frontal theta (θ) levels can help

police officers to train cognitive readiness during shooting scenarios. Moreover, values of the theta/beta ratio collected during the experiment showed that this metric could be used to differentiate between difficulty levels in simulations for firearm training.

On the other hand, police shooters are affected not only by responses of the central nervous system, but also the peripheral system as well. A consistent and mastered skill of controlling cardiovascular responses able to produce desired patterns of respiration has been identified as a major component of firearms

FIGURE 8 | Integration of the BioPhyS and the BL Engine showing how it can be used to provide biofeedback based on the calmness and focus metrics.

training (Johnson, 2007). HR levels significantly increased from the resting or *baseline* recording where police officers were not engaged with shooting activities. A trend in **Figure 4** shows how HR levels were incrementally impacted by the simulation difficulty, demonstrating how via systematic modulations in the simulation parameters, cardiovascular responses of shooters can be affected. Similar responses have been found in police officers after training practices under pressure (Oudejans, 2008). Although only few parameters were used to modulate the simulation difficulties, further research using more metaphorical simulation variables (e.g., daylight and rain intensity) could reveal the ideal configuration to elicit the desired cardiovascular responses (Marchiori et al., 2018). Finally, we believe the overall user experience was positively perceived by police officers due to (i) the wearability of the physiological sensors which avoided extra discomfort produced by more invasive setups (e.g., multi-channel ECG signals or EEG caps), (ii) the realism of the simulation provided by the VR equipment used as well as the accuracy of the motion tracking achieved through the room-scale play area using two lighthouses, and (iii) the usefulness of the system perceived by the instructors and other police officers who were enthusiastic about using this technology in real training.

Characterization, Psychophysiological Model, and the Physiological Intelligence Layer

The characterization study with 10 police officers serves as a starting point for the design of physiological computing system that would be able to create the biocybernetic adaptation in the BioPhyS. From the model (see **Figure 6**) two trends can be identified, while HR levels increased proportionally with the simulation difficulty, the frontal theta values decreased. Theta waves have been found as a sensitive neurophysiological marker to describe participant's discomfort (Heo and Yoon,

2020) and motion sickness (Park et al., 2008) in gaming/VR related studies. Moreover, cardiovascular responses such as HR levels have been also studied in firearms training showing heightened states of arousal after active shooting (Heim et al., 2006). Thus, differences between the active shooting and baseline revealed that just by engaging people in the virtual shooting activity without considering the difficulty level, specific psychophysiological patterns can be identified. This is important since it allows differentiating between the physiological signature of being wearing the headset and sensors as well as being standing holding the controllers and the cognitive and cardiovascular cost of shooting. In other words, this provides a certain level of context awareness, so the adaptive system would be able to create more personalize modulations (Novak, 2014). Finally, the adaptive rule created using the BL Engine tool (see **Figure 7**) shows how to move from the theoretical model to a real software implementation by: (i) combining both cardiovascular and neurophysiological features into a clearly defined and transparent adaptive rule, (ii) allowing the modifications of the rule and values in real-time, so speeding up threshold's adjustment for real-time adaptation, and (iii) fully integrating the VR simulator with the BL Engine, so a bidirectional communication will take place enabling research on biofeedback visualization strategies (Kuikkaniemi et al., 2010). Relatedly, Kuikkaniemi et al. (2010) hypothesized that the increase in immersion due to adaptive biofeedback that they demonstrated in the context of first-shooter games could also be achieved in other contexts such as 3D virtual environments (Malińska et al., 2015). There is also evidence that neurofeedback, which BioPhyS employs, may boost the effects of cognitive training (Dessy et al., 2018). Biocybernetic adaptation integrated during exercise-based videogames called Exergames was used effectively to encourage older users to exert in targeted cardiac zones (Muñoz et al., 2018). A similar concept can be explored here, where the difference between the trainee's HR

response and the target or setpoint HR can be used to drive attributes of the simulation task, e.g., target speed or hindering rain intensity, which, in turn, would be expected to drive the trainee's HR.

Future Work With Virtual Reality and Biocybernetic Adaptation

In the BioPhyS system programmed for self-regulation training, attributes of the VR simulation task, e.g., target speed or rain intensity, are adjusted in ways designed to encourage certain psychophysiological response changes and to discourage others. Physiological self-regulation training with the BioPhyS leverages two memory process principles to promote the transfer of skill learning. Stimuli in the VR environment are designed to match stimuli in the real-world shooting situation, thereby promoting the transfer of firearm skill learning via a memory process known as encoding specificity (Pugnetti et al., 2001; Jaiswal et al., 2010). Physiological states that are determined to be effective in the real-world shooting situation are rewarded in the training setting by biocybernetic adaptation, thereby promoting transfer of learning via a memory process known as state-dependent learning (Pugnetti et al., 1996). In order to investigate how to integrate biocybernetic adaptation strategies into novel immersive systems for firearm training, the BioPhyS can leverage physiological computing technologies to be empowered with intelligent capabilities. Specifically, in this project, the BL Engine offers strategies suitable for enhancing shooting performance by controlling the elements in the virtual environment that affect the simulation difficulty and the officer's concentration and relaxation levels (concentrated and so-called "calm, cool, and collected" state). To improve their scores in the shooting scenarios, police officers would have to use self-regulation strategies (e.g., respiration and attentional control) that will keep them calm and focused, hence reducing the probability of making mistakes in real-life situations (Blacker et al., 2018). Additionally, the physiological challenges can be modified (in real-time, if needed) by trainers allowing a very dynamic and personalized training.

Including physiological sensing technologies in novel immersive and interactive technologies such as augmented or VR has been posed as the new generation of gaming interfaces (Pope et al., 2014). This integration is far from being simple and straightforward since including the human in the loop carries several complexities widely discussed (Novak, 2014). Nevertheless, current advances in sensing technologies, game engine software, and hardware accessibility have allowed the development of very sophisticated *biocybernetically* adaptive systems empowered with the *immersivity* and realism of 52 (Kosunen, 2018; Muñoz et al., 2019). In order to get a more solid and integrated VR-biocybernetic confluence, advances in three aspects are required:

- Improvements in the hardware integration: to aid the design and development process of *biocybernetically* empowered VR applications, the integration of

physiological sensing devices, and VR headsets should be minimally intrusive and less cumbersome. Although in our experience we used minimally intrusive sensors, the use of VR headsets that include sensors to record body signals (e.g., Neurable,[3] Looxidlabs,[4] Interaxon; Aimone et al., 2018) will aid the widespread of such technological symbiosis.

- Advances in the software intelligence: even though several good examples of using biocybernetic adaptation in games implemented simplistic adaptive techniques to maximize the desired target psychophysiological state (Pope et al., 2014; Ewing et al., 2016), more sophisticated adaptive algorithms can be used to provide a more robust and personalized experience. For instance, techniques based on control feedback principles such as PID controllers can be used to create more accurate modulations in the simulation variables, thus facilitating a faster and more sustained "desired" physiological response (Parnandi and Gutierrez-Osuna, 2014; Muñoz et al., 2019). Machine learning techniques can also be used to detect complex psychophysiological states such as stress or workload and adapt the content accordingly (Panicker and Gayathri, 2019).

- Spreading out successful use cases outside academia: progress in *biocybernetically* adaptive technologies has been mostly presented in academic environments with only few exceptions (Pope et al., 2014). Although this construct has shown clear evidence of how it can be used in scenarios like gaming, healthcare, and training (Fairclough, 2017), examples of successful and *commercializable* use cases of *biocybernetically* empowered products are still scarce.

- Considerations for simulator and VR sickness are vital for meaningful results to be gleaned from this and similar research embedding physiological computing into simulated (VR, AR, MR, etc.) experiences/environments. Metrics for the effects that simulated stimuli have on test subjects have been developed and examined extensively (Kennedy et al., 1993, 2001; Saredakis et al., 2020) and should be applied in future studies involving similar technologies.

STUDY LIMITATIONS

Our study encloses numerous limitations that are listed as follows:

- Sample size: the size of the sample, $n = 10$, was small. Therefore, caution should be exercised in generalizing the results of this VR characterization study. The results should be confirmed in future studies with a larger sample size. Additionally, participants were diverse in terms of training skills and demographics.

---

[3]http://www.neurable.com/
[4]https://looxidlabs.com/

- Validation of the physiologically adaptive model: although both the psychophysiological model (**Figure 6**) and the adaptive rule created in the BL Engine (**Figure 7**) serve as valuable inputs for the real-time biocybernetic loop implementation, a controlled experiment validating the proposed methods is needed to better understand the effects of physiological adaptation to train self-regulation skills in police officers. Data variability, for instance, is a major challenge for implementing the real-time model since differences in physiological responses between resting states and active shooting differ considerably. For instance, HR differences are largely the same in all simulation difficulties whereas frontal theta only showed differences between easy and hard difficulties.

- Limitations of the EEG sensor: the EEG wearable had few electrode sites, limiting EEG data density. Furthermore, motion artifact in the EEG and ECG signals can be an issue with an active VR task such as target shooting. Nevertheless, informative studies employing these three technologies have been successfully conducted (Marín-Morales et al., 2019).

- Exposure time for each training scenario: a short duration, 3 min, was the time used to record the physiological responses during the pre-established training scenarios (*easy*, *medium*, and *hard*), which should be analyzed carefully since firearms training sessions could last more time (Johnson, 2007). However, it is worth mentioning that critical incidents can last seconds and training in physiological regulation could play an important role in preparing police officers for such scenarios. Besides, it is also important to consider that extracting cardiovascular and neurophysiological features in short time intervals (e.g., 1–5 min) could bring more accurate information of the specific psychophysiological responses rather than averaging an entire, long session (e.g., more than 5 min).

- Limitations in the weapon used: due to technical issues in integrating more realistic air-pistols in our study, we used the standard controllers of the VR equipment which can be seen as less accurate in terms of creating a vivid simulation. Also, the use of the virtual revolver differs from the pistols used in real scenarios, thus can be seen as an extra limitation.

Further research is needed to provide a more conclusive understanding of how biocybernetic adaptation can impact police officers' performance in firearms training using VR simulations.

CONCLUSION

This study characterized the dynamical neural and cardiac responses of police officers in a VR based target shooting scenario and provides the foundation for a biocybernetic system with an intelligent adaptation layer. The simulation difficulty factor was shown to influence shooting performance, frontal theta brainwave activity and HR, providing a specification of values of the physiological variables to be targeted in a biocybernetic

adaptation system as well as to serve as a control measure for assessing the effects of adding a biocybernetic adaptation layer to a VR training system. Assessment of the results from the Simulator Sickness Questionnaire showed that the experience was generally well tolerated. Further development and refinement of the VR simulation capability (including ergonomic improvements to the headset) are suggested to mitigate unintended effects of simulation stimuli on users of such systems. An explanation was given for employing the neural and cardiac findings to integrate physiological intelligence into the VR simulator using a specialized software tool, the BL Engine. Next steps on this research must employ the provided model to quantify the effects of using biocybernetic adaptation in VR-based firearms training for police officers.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are not publicly available as participants did not provide their informed consent for the public sharing of their data. The datasets are available on request to john.munoz.hci@uwaterloo.ca.

ETHICS STATEMENT

Ethical review and approval was completed for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this manuscript.

AUTHOR CONTRIBUTIONS

JM and AP designed and defined the BioPhyS approach. JM conducted the experiment and collected the data. LQ developed the VR simulation and carried out the integration with the BL Engine. CS contributed and reviewed the manuscript. All authors revised and approved the current version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

process, (ii) the National Institute of Aerospace (NIA) that supported the convergence of the stakeholders for this project, (iii) the PRISM team from NASA Langley for its very supportive feedback throughout the project development, (iv) Zeltech employees at Hampton facilities. Special thanks to personnel from the Hampton, Virginia Police Department who participated actively in our studies; and Jeremy Sklute and Mike Priddy from J&F Alliance who helped in improving the system's realism, and (v) Stockholm University for providing open access publication fees. This

collaboration with NASA was established through the Langley Research Center Commercialization Office with the J&F Alliance Group, including Dr. Munoz and Dr. Quintero, who licensed several NASA patents related to simulation and game psychophysiological modulation. The use case that the J&F Alliance Group determined to be high potential was military/police training. The support and cooperation of the Langley Research Center Commercialization Office (Jesse Midgett) and System Wide Safety Project (John Koelling) are gratefully appreciated.

REFERENCES

Aimone, C. A., Mackenzie, S. T., Moffat, G. D., Banville, H., and Proulx, N. H. (2018). *Wearable Computing Device with Electrophysiological Sensors*. Toronto, CA: INTERAXON INC.

Alves, T., Gama, S., and Melo, F. S. (2018). "Flow adaptation in serious games for health," in *Proceedings of the 2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH)* (Piscataway, NJ: IEEE), 1–8.

Berger, A. M., and Davelaar, E. J. (2018). Frontal alpha oscillations and attentional control: a virtual reality neurofeedback study. *Neuroscience* 378, 189–197. doi: 10.1016/j.neuroscience.2017.06.007

Bian, D., Wade, J., Swanson, A., Weitlauf, A., Warren, Z., and Sarkar, N. (2019). Design of a physiology-based adaptive virtual reality driving platform for individuals with ASD. *ACM Trans. Accessible Comput.* 12, 1–24. doi: 10.1145/3301498

Blacker, K. J., Hamilton, J., Roush, G., Pettijohn, K. A., and Biggs, A. T. (2018). Cognitive training for military application: a review of the literature and practical guide. *J. Cogn. Enhanc.* 3, 30–51. doi: 10.1007/s41465-018-0076-1

Burdea, G. C., and Coiffet, P. (2003). *Virtual Reality Technology*. Hoboken, NJ: John Wiley & Sons.

Chanel, G., Rebetez, C., Bétrancourt, M., and Pun, T. (2011). Emotion assessment from physiological signals for adaptation of game difficulty. IEEE Transactions on Systems, Man, and Cybernetics-Part A. *Syst. Hum.* 41, 1052–1063. doi: 10.1109/tsmca.2011.2116000

Dessy, E., Van Puyvelde, M., Mairesse, O., Neyt, X., and Pattyn, N. (2018). Cognitive performance enhancement: do biofeedback and neurofeedback work? *J. Cogn. Enhanc.* 2, 12–42. doi: 10.1007/s41465-017-0039-y

Dobbins, C., and Fairclough, S. (2018). "Detecting negative emotions during real-life driving via dynamically labelled physiological data," in *Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)* (Piscataway, NJ: IEEE), 830–835.

Ewing, K. C., Fairclough, S. H., and Gilleade, K. (2016). Evaluation of an adaptive game that uses eeg measures validated during the design process as inputs to a biocybernetic loop. *Front. Hum. Neurosci.* 10:223. doi: 10.3389/fnhum.2016.00223

Fairclough, S., and Gilleade, K. (2012). "Construction of the biocybernetic loop: a case study," in *Proceedings of the 14th ACM International Conference on Multimodal Interaction ICMI '12* (New York, NY: ACM), 571–578. doi: 10.1145/2388676.2388797

Fairclough, S. H. (2017). "Physiological computing and intelligent adaptation," in *Emotions and Affect in Human Factors and Human-Computer Interaction*, Ed. M. Jeon (Amsterdam: Elsevier), 539–556. doi: 10.1016/b978-0-12-801851-4.00020-3

Faria, A. L., Cameirão, M. S., Couras, J. F., Aguiar, J. R. O., Costa, G. M., and Bermúdez i Badia, S. (2018). Combined cognitive-motor rehabilitation in virtual reality improves motor outcomes in chronic stroke – a pilot study. *Front. Psychol.* 9:854. doi: 10.3389/fpsyg.2018.00854

Hanakova, L., Socha, V., Socha, L., Szabo, S., Kozuba, J., Lalis, A., et al. (2017). "Determining importance of physiological parameters and methods of their evaluation for classification of pilots psychophysiological condition," in *Proceedings of the 2017 International Conference on Military Technologies (ICMT)* (Piscataway, NJ: IEEE), 500–506.

Harrivel, A. R., Liles, C., Stephens, C. L., Ellis, K. K., Prinzel, L. J., and Pope, A. T. (2016). *Psychophysiological Sensing and State Classification for Attention Management in Commercial Aviation*. Kissimmee, FL: AIAA Infotech@ Aerospace, 1490.

Heim, C., Schmidtbleicher, D., and Niebergall, E. (2006). The risk of involuntary firearms discharge. *Hum. Fact.* 48, 413–421. doi: 10.1518/001872006778606813

Heo, J., and Yoon, G. (2020). EEG studies on physical discomforts induced by virtual reality gaming. *J. Electr. Eng. Technol.* 1–7. doi: 10.1007/s42835-020-00373-1

Iatsyshyn, A., Valeriia, K., Romanenko, Y. O., Deinega, I. I., Andrii, I., Popov, O., et al. (2019). "Application of augmented reality technologies for preparation of specialists of new technological era," in *Proceedings of the 2nd International Workshop on Augmented Reality in Education (AREdu 2019)* (Ukraine: Kryvyi Rih). Available online at: https://www.researchgate.net/publication/339944436_Application_of_augmented_reality_technologies_for_preparation_of_specialists_of_new_technological_era (accessed March 31, 2020).

Jaiswal, N., Ray, W., and Slobounov, S. (2010). Encoding of visual–spatial information in working memory requires more cerebral efforts than retrieval: evidence from an EEG and virtual reality study. *Brain Res.* 1347, 80–89. doi: 10.1016/j.brainres.2010.05.086

Johnson, B. R. (2007). *Crucial Elements of Police Firearms Training*. Flushing, NY: Looseleaf Law Publications.

Kamzanova, A. T., Matthews, G., Kustubayeva, A. M., and Jakupov, S. M. (2011). EEG indices to time-on-task effects and to a workload manipulation (cueing). *World Acad. Sci. Eng. Technol.* 80, 19–22.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* 3, 203–220. doi: 10.1207/s15327108ijap0303_3

Kennedy, R. S., Stanney, K. M., Drexler, J. M., Compton, D. E., and Harm, D. L. (2001). "Configural scoring of simulator sickness, cybersickness, and space adaptation syndrome: similarities and differences?," in *Psychological Issues in the Design and Use of Virtual and Adaptive Environments*, eds L. Hettinger and M. Haas (Mahwah, NJ: Lawerence Erlbaum Assosciates).

Kosunen, I. (2018). *Exploring the Dynamics of the Biocybernetic Loop in Physiological Computing*. Finland: University of Helsinki.

Kuikkaniemi, K., Laitinen, T., Turpeinen, M., Saari, T., Kosunen, I., and Ravaja, N. (2010). "The influence of implicit and explicit biofeedback in first-person shooter games," in *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, Atlanta, 859. doi: 10.1145/1753326.1753453

Lehmann, K. S., Ritz, J. P., Maass, H., Hakmak, H. K., Kuehnapfel, U. G., Germer, C. T., et al. (2005). A prospective randomized study to test the transfer of basic psychomotor skills from virtual reality to physical reality in a comparable training setting. *Ann. Surg.* 241, 442–449. doi: 10.1097/01.sla.0000154552.89886.91

Lele, A. (2013). Virtual reality and its military utility. *J Ambient Intell Human Comput* 4, 17–26. doi: 10.1007/s12652-011-0052-4

Liu, C. C., Hosking, S. G., and Lenné, M. G. (2009). Predicting driver drowsiness using vehicle measures: recent insights and future challenges. *J. Safety Res.* 40, 239–245. doi: 10.1016/j.jsr.2009.04.005

Malińska, M., Zużewicz, K., Bugajska, J., and Grabowski, A. (2015). Heart rate variability (HRV) during virtual reality immersion. *Int. J. Occup. Saf. Ergon.* 21, 47–54. doi: 10.1080/10803548.2015.1017964

Marchiori, E., Niforatos, E., and Preto, L. (2018). Analysis of users' heart rate data and self-reported perceptions to understand effective virtual reality characteristics. *Inform. Technol. Tour.* 18, 133–155. doi: 10.1007/s40558-018-0104-0

Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Gentili, C., et al. (2019). Real vs. immersive-virtual emotional experience: analysis of psycho-physiological patterns in a free exploration of an art museum. *PLoS One* 14:e0223881. doi: 10.1371/journal.pone.0223881

Mason, P. H. (1998). *Recovering From the War: A Guide for all Veterans, Family Members, Friends and Therapists.* New York, NY: Patience Press.

McMahan, T., Parberry, I., and Parsons, T. D. (2015). "Evaluating electroencephalography engagement indices during video game play," in *Proceedings of the Foundations of Digital Games Conference*, California.

Muñoz, J. E., Gouveia, E. R., Cameirão, M. S., and Badia, S. B. (2017a). PhysioLab-a multivariate physiological computing toolbox for ECG, EMG and EDA signals: a case of study of cardiorespiratory fitness assessment in the elderly population. *Multimedia Tools Appl.* 77, 11521–11546. doi: 10.1007/s11042-017-5069-z

Muñoz, J. E., Rubio, E., Cameirao, M., and Bermúdez, S. (2017b). "The biocybernetic loop engine: an integrated tool for creating physiologically adaptive videogames," in *Proceeding of the 4th International Conference in Physiological Computing Systems*, Madrid.

Muñoz, J. E., Paulino, T., Vasanth, H., and Baras, K. (2016a). "PhysioVR: a novel mobile virtual reality framework for physiological computing," in *Proceedings of the e-Health Networking, Applications and Services (Healthcom), 2016 IEEE 18th International Conference on (IEEE)*, Piscataway, NJ. 1–6.

Muñoz, J. E., Pope, A. T., and Velez, L. E. (2016b). "Integrating biocybernetic adaptation in virtual reality training concentration and calmness in target shooting," in *Physiological Computing Systems*, eds H. P. Silva, D. Majoe, and S. Fairclough (Berlin: Springer), 218–237.

Muñoz, J. E., Pope, A. T., and Velez, L. E. (2019). "Integrating biocybernetic adaptation in virtual reality training concentration and calmness in target shooting," in *Proceeding of the International Conference, Physiological Computing Systems (PhyCS 2015-2018) Revised Selected Papers Series: Lecture Notes in Computer Science* (Berlin: Springer). doi: 10.1007/978-3-030-27950-9_12

Muñoz, J. E., Rubio, E., Cameirao, M. S., and Bermúdez i Badia, S. (2018). "Closing the loop in exergaming - health benefits of biocybernetic adaptation in senior adults," in *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, Melbourne, VIC.

Muñoz, J. E., Stephens, C. L., Velez, L. E., and Pope, A. T. (2019). "Modeling biocybernetic adaptation through feedback control systems principles: implications in human-computer interaction and virtual reality applications," in *Proceedings of the The Second Neuroadaptive Technology Conference*, Liverpool, 39. doi: 10.1007/978-3-642-46847-6_3

Novak, D. (2014). "Engineering issues in physiological computing," in *Advances in Physiological Computing Human–Computer Interaction Series*, eds S. H. Fairclough, and K. Gilleade (London: Springer), 17–38. doi: 10.1007/978-1-4471-6392-3_2

Novak, D., Mihelj, M., and Munih, M. (2012). A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Interact. Comput.* 24, 154–172. doi: 10.1016/j.intcom.2012.04.003

Oliver, R. A., Cancio, J. M., Rábago, C. A., and Yancosek, K. E. (2019). Impact of fire arms training in a virtual reality environment on occupational performance (Marksmanship) in a polytrauma population. *Military Med.* 184, 832–838. doi: 10.1093/milmed/usz010

Oudejans, R. R. D. (2008). Reality-based practice under pressure improves handgun shooting performance of police officers. *Ergonomics* 51, 261–273. doi: 10.1080/00140130701577435

Panicker, S. S., and Gayathri, P. (2019). A survey of machine learning techniques in physiology based mental stress detection systems. *Biocybernet. Biomed. Eng.* 39, 444–469. doi: 10.1016/j.bbe.2019.01.004

Park, J.-R., Lim, D.-W., Lee, S.-Y., Lee, H.-W., Choi, M.-H., and Chung, S.-C. (2008). Long-term study of simulator sickness: differences in EEG response due to individual sensitivity. *Int. J. Neurosci.* 118, 857–865. doi: 10.1080/00207450701239459

Parnandi, A., and Gutierrez-Osuna, R. (2014). A comparative study of game mechanics and control laws for an adaptive physiological game. *J. Multimodal User Interf.* 9, 31–42. doi: 10.1007/s12193-014-0159-y

Pereira, M., Argelaguet, F., Millán, J., del, R., and Lécuyer, A. (2018). Novice shooters with lower pre-shooting alpha power have better performance during competition in a virtual reality scenario. *Front. Psychol.* 9:527. doi: 10.3389/fpsyg.2018.00527

Plews, D. J., Scott, B., Altini, M., Wood, M., Kilding, A. E., and Laursen, P. B. (2017). Comparison of heart-rate-variability recording with smartphone photoplethysmography. Polar H7 chest strap, and electrocardiography. *Int. J. Sports Physiol. Perform.* 12, 1324–1328. doi: 10.1123/ijspp.2016-0668

Pope, A. T., Stephens, C. L., and Gilleade, K. (2014). "Biocybernetic adaptation as biofeedback training method," in *Advances in Physiological Computing*, Ed. K. Gilleade (Berlin: Springer), 91–115. doi: 10.1007/978-1-4471-6392-3_5

Pugnetti, L., Meehan, M., and Mendozzi, L. (2001). Psychophysiological correlates of virtual reality: a review. *Presence* 10, 384–400. doi: 10.1162/1054746011470244

Pugnetti, L., Mendozzi, L., Barberi, E., Rose, F. D., and Attree, E. A. (1996). "Nervous system correlates of virtual reality experience," in *Proceedings of the European Conference on Disability, Virtual Reality and Associated Technology*, Maidenhead.

Rizzo, A., and Shilling, R. (2017). Clinical virtual reality tools to advance the prevention, assessment, and treatment of PTSD. *Eur. J. Psychotraumatol.* 8:1414560. doi: 10.1080/20008198.2017.1414560

Sajnog, C. (2013). *How to Shoot Like a Navy Seal: Combat Marksmanship Fundamentals.* San Diego, CA: Center Mass Group, LLC.

Sanei, S., and Chambers, J. A. (2013). *EEG Signal Processing.* Hoboken, NJ: John Wiley & Sons.

Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A. D., Rizzo, A., and Loetscher, T. (2020). Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. *Front. Hum. Neurosci.* 14:96. doi: 10.3389/fnhum.2020.00096

Zhang, M., Li, X., Yu, Z., Chen, Z., Sun, Y., and Li, Y. (2019). "Virtual range training based on virtual reality," in *Proceedings of the 4th International Conference on Contemporary Education, Social Sciences and Humanities (ICCESSH 2019)* (Moscow: Atlantis Press), doi: 10.2991/iccessh-19.2019.339

Check for updates

Well Done! Effects of Positive Feedback on Perceived Self-Efficacy, Flow and Performance in a Mental Arithmetic Task

Corinna Peifer[1]*, Pia Schönfeld[2], Gina Wolters[1], Fabienne Aust[1] and Jürgen Margraf[2]

[1] Faculty of Psychology, Applied Psychology in Work, Health, and Development, Ruhr University Bochum, Bochum, Germany, [2] Faculty of Psychology, Mental Health Research and Treatment Center, Ruhr University Bochum, Bochum, Germany

Self-efficacy is a well-known psychological resource, being positively associated with increased performance. Furthermore, results from field studies suggest a positive impact of self-efficacy on flow experience, which has not yet been tested experimentally. In this study, we manipulated self-efficacy by means of positive feedback and investigated whether self-efficacy serves as a mediator in the relationship between positive feedback and flow and in the relationship between positive feedback and performance. Our sample consisted of 102 participants (63 female, 39 male). The experimental group received positive feedback after completing 5 min of mental arithmetic tasks on a computer, whereas the control group received no feedback. A second session of a mental arithmetic task was then completed for 5 min. Mediation analyses confirmed that specific self-efficacy mediated a positive effect of positive feedback on flow as well as on both performance measures (quality and quantity) in a subsequent task. However, direct effects of feedback on flow and on performance were not significant, which suggests the presence of other mechanisms that remain to be investigated.

Keywords: feedback, self-efficacy, flow, performance, mental arithmetic

INTRODUCTION

"Well done!" – Positive feedback has been found not only to enhance performance (e.g., Kluger and DeNisi, 1996; Hattie and Timperley, 2007), but also to be an efficient intervention to manipulate perceived self-efficacy (e.g., Brown et al., 2012). Self-efficacy refers to the judgment of one's own abilities to successfully cope with future demands (Bandura, 1983). It can either refer to a general judgment, called general self-efficacy, or it can refer to more specific domains, such as mathematical skills, then called specific self-efficacy. Self-efficacy is well-known as a psychological resource protecting mental health and buffering the negative effects of stress (Bandura, 1977, 1986; Schönfeld et al., 2016). In line with Bandura's social cognitive theory (SCT), there is a broad basis for higher self-efficacy being associated with lower symptoms of depression and anxiety as well as with higher optimism and emotional well-being (Benight et al., 1999; Rottmann et al., 2010; Singh and Bussey, 2011; Wang et al., 2014).

While self-efficacy has been found to enhance performance (e.g., Bandura and Locke, 2003), contradictory findings also exist, suggesting that effects may differ with respect to different performance outcomes such as performance quality and quantity (Vancouver et al., 2014). The first part of our study contributes to the answer to this yet open research question by examining the effects of positive feedback as an intervention to manipulate self-efficacy and we test how self-efficacy affects performance quality and quantity. Further, we contribute to research on the mechanisms via which the effects of positive feedback are transmitted to performance by using self-efficacy as a mediator in the feedback-performance relationship.

The second part of this study examines flow – the experience of being fully absorbed in a task (Csikszentmihalyi, 1975) – in relation to positive feedback and self-efficacy. Feedback has been described as a central antecedent of flow (e.g., Landhäußer and Keller, 2012), but effects of positive feedback on flow have not yet been investigated experimentally. Another antecedent of flow as identified in field studies (e.g., Zubair and Kamal, 2015a,b) is self-efficacy. This study aims to replicate the positive relationship between self-efficacy and flow in an experimental setting, in which self-efficacy is manipulated using positive feedback. Finally, and adding to the existing research, we aim to test whether self-efficacy transmits effects of positive feedback on flow experience.

Positive Feedback and Performance

Feedback can be defined as the "provision of information regarding some aspect(s) of one's task performance" (Kluger and DeNisi, 1996, p. 255). Meta-analyses show impressive effects of feedback on increased performance, with average effect sizes of $d = 0.40$ (Kluger and DeNisi, 1996) and $d = 0.79$ (Hattie and Timperley, 2007). Research has identified moderators of the feedback-performance relationship, with findings suggesting that positive feedback is more efficient than negative feedback. For example, Arbel et al. (2014) found that positive feedback improved learning performance more than negative feedback. Furthermore, it has been found that feedback after good trials enhanced learning in comparison to feedback after poor trials (Chiviacowsky and Wulf, 2007). In line with these findings, the meta-analysis of Kluger and DeNisi (1996) found that feedback following correct results, that is, positive feedback, was more effective than feedback following incorrect results. Furthermore, feedback was more effective when it was provided by a computer ($d = 0.41$) vs. not ($d = 0.23$; Kluger and DeNisi, 1996).

One particular type of feedback is normative feedback, which refers to information on one's performance compared to referenced others, allowing comparative inferences (Hartwell and Campion, 2016). Accordingly, positive normative feedback is the information that one's performance was better than that of referenced others, such as feedback indicating above-average performance. Studies suggest that such positive normative feedback – even if it is false feedback – leads to increased performance compared to negative normative feedback (i.e., indicating below-average performance; Bandura and Jourden, 1991; Wulf et al., 2010). Based on these findings, we expect to replicate earlier studies by finding positive effects of positive normative feedback on performance.

Positive Feedback as an Intervention to Increase Self-Efficacy

Previous research has shown that false normative positive feedback not only affects performance, but also self-efficacy, and such feedback has been successfully applied to manipulate self-efficacy (e.g., Reynolds, 2006; Beattie et al., 2016; Brown et al., 2016; Dimotakis et al., 2017). This has also been used in experiments with mental arithmetic tasks (Weinberg et al., 1979; Wright and Gregorich, 1989; Eden and Zuk, 1995; Brown et al., 2012). The approach is in line with Bandura's SCT (Bandura, 1977). Bandura (1977) pointed out that there are different kinds of information that lead to expectations about personal efficacy: performance accomplishment, vicarious experience, verbal persuasion and psychological states. In line with that, external persuasion through positive feedback to induce an experience of success should be an effective strategy to manipulate self-efficacy (Achterkamp et al., 2015). In line with this, we expect to replicate positive effects of positive normative feedback on self-efficacy.

Self-Efficacy and Performance

Successful mastery experiences contribute to the development of efficacy beliefs and increase the investment of effort and the level of performance (Bandura, 1997). Perceived self-efficacy is a key dynamic and malleable factor affecting behavior (Gist and Mitchell, 1992; Hardy, 2014), and some evidence indicates that higher self-efficacy leads to better performance in cognitive and sports tasks (e.g., Beattie et al., 2014; Niemiec and Lachowicz-Tabaczek, 2015). At the same time, divergences in social cognitive and control theories lead to different assumptions about the effects of self-efficacy (see Bandura and Locke, 2003; Bandura, 2012; Schönfeld et al., 2017). For example, Powers' (1973, 1991) perceptual control theory assumes that the discrepancy between one's personal goal and one's perceived progress in handling a situation successfully regulates the performed action (e.g., Vancouver et al., 2002). In case of a low discrepancy, the person will invest fewer resources in achieving the goal, and successful performance is assumed to be easy. In situations in which perceived progress is ambiguous, perceived capabilities (i.e., self-efficacy) can be used as an indicator of progress. As a consequence, high perceived skills will lead to a decreased perceived discrepancy between goal and progress. Thus, according to perceptual control theory, high self-efficacy would undermine performance and motivation. Initial empirical findings support these assumptions (e.g., Vancouver et al., 2001, 2002, 2008, 2014; Vancouver and Kendall, 2006; Vancouver, 2012; Beattie et al., 2014). A study by Vancouver et al. (2014), for example, found that self-efficacy was negatively related to performance quality, while it was positively related to performance quantity. They assumed that individuals with high self-efficacy allocate less effort per task, which leads to faster progress (performance quantity), but lower quality of results in the form of more mistakes. Yet, more research

is needed to disentangle effects of self-efficacy on different performance measures.

In accordance with the findings of Vancouver et al. (2014), our study differentiates between performance quantity and performance quality in that we assume that self-efficacy has positive effects on performance quantity, but negative effects on performance quality.

A general limitation of research on self-efficacy and performance is that it is largely based on observational rather than experimental designs, so no conclusions can be drawn about the direction of effects. While most studies assume positive effects of self-efficacy on performance, Bandura (1977) in fact already identified performance accomplishment as an antecedent of self-efficacy. Accordingly, experimental research on the relationship between self-efficacy and performance is necessary to disentangle potential bidirectional effects.

Consequently, by differentiating between performance quantity and quality, and by applying an experimental design in which we manipulate self-efficacy by means of positive feedback, we aim to contribute to a better understanding of the relationship between self-efficacy and performance.

Self-Efficacy as a Mechanism That Transmits Effects of Positive Feedback on Performance

As mentioned above, the induction of positive feedback including a favorable comparison to others has been found to be a suitable method to enhance the level of self-efficacy, which in turn, affects performance (see Zinken et al., 2008). Based on a comparative appraisal, the individual is persuaded that he or she has performed successfully, which is in line with SCT. Using feedback-manipulation as a strategy to increase a person's appraisal of his or her capabilities, beneficial effects have also been demonstrated in the context of emotional learning processes (Zlomuzica et al., 2015). However, to the best of our knowledge, the mediation hypothesis of self-efficacy has not yet been tested. Integrating the described relationships, we propose that self-efficacy acts as a mediator, transmitting positive effects of positive feedback on performance. Taking the differential expectations for the relationship of self-efficacy with performance quantity and performance quality into account, *we expect that self-efficacy acts as a mediator, transmitting positive effects of positive feedback on performance quantity (Hypothesis 1a), but transmitting negative effects of positive feedback on performance quality (Hypothesis 1b).*

Effects of Feedback on Flow-Experience

Flow is the positive experience of being fully absorbed in an optimally challenging task. While in flow, individuals are completely concentrated on the task at hand, which is experienced as rewarding in itself. Individuals perceive clear goals and feedback and a high level of control over the demands, thereby experiencing a merging of action and awareness, and a loss of self-consciousness, along with a distorted sense of time (Csikszentmihalyi, 1975). Flow can be experienced in different activities and tasks, among them cognitive tasks such as solving

math calculations – even under laboratory conditions (e.g., Harmat et al., 2015; Ulrich et al., 2016).

Feedback has been described as one of the core antecedents fostering flow-experience (Bakker, 2005; Demerouti, 2006; Landhäußer and Keller, 2012; Nakamura and Csikszentmihalyi, 2014). This conceptualization is in line with findings based on the job characteristics model (Hackman et al., 1975), showing that feedback along with four other core job characteristics is positively related to flow experience (Bakker, 2005; Demerouti, 2006; Maeran and Cangiano, 2013). Studies that have specifically examined the relationship between feedback and flow have confirmed a positive link between the two (Rau and Riedel, 2004; Maeran and Cangiano, 2013). However, these studies have focused on feedback in general, and the effects of specifically positive feedback on flow have not yet been studied using quantitative research. Qualitative research has provided first indications that positive feedback – but not negative feedback – is an antecedent of flow (Jackson, 1995; Swann et al., 2015). In line with this, positive normative feedback has been found to have positive effects on positive affect during a challenging task (Hutchinson et al., 2008) – which is often linked to flow experience. Furthermore, positive normative feedback has been suggested to have energizing and reinforcing effects (Kühn et al., 2008) – both typical characteristics of the experience of flow. Bringing together theoretical and empirical evidence, we expect to find positive effects of positive normative feedback on flow experience.

Self-Efficacy and Flow-Experience

A central component of flow is the perceived balance between the demands of the task and the individual's skills (e.g., Csikszentmihalyi, 1975; Landhäußer and Keller, 2012). The level of self-efficacy is an individuals' evaluation of his/her skills and therefore has a substantial impact on how the balance between skills and task demands is perceived. High levels of self-efficacy should thus positively impact flow. Empirical studies support this assumption: For example, Zubair and Kamal (2015a,b) investigated the relationship between the dimensions of psychological capital (Luthans et al., 2004) and flow experience and found that all dimensions, including self-efficacy, were positively related to flow. In a two-wave longitudinal study design, Salanova et al. (2006) found that work-specific self-efficacy beliefs facilitated the experience of work-related flow. In another longitudinal study, Rodríguez-Sánchez et al. (2011) found that teachers' work-related self-efficacy positively affected their flow experience. Furthermore, collective efficacy beliefs are associated with higher flow. In a longitudinal study with small groups, it was found that collective efficacy beliefs can lead to higher collective flow, which in turn leads to higher collective self-efficacy in the future, forming a reciprocal relationship (Salanova et al., 2014). Furthermore, Pineau et al. (2014) found that self-efficacy as well as team-efficacy are significantly related to dispositional flow. All in all, abundant research supports the hypothesis that self-efficacy beliefs are positively associated with flow experience. Following the existing literature, we postulate that self-efficacy facilitates flow experience.

Self-Efficacy as a Mechanism That Transmits Effects of Positive Feedback on Flow

While results from field studies, including long-term studies, suggest a reciprocal and positive relationship between self-efficacy and flow, this relationship has not yet been tested experimentally. As outlined above, positive feedback is an established intervention to positively affect self-efficacy. Thus, we use positive feedback to manipulate self-efficacy with the aim to test the effects of self-efficacy on flow experimentally. Furthermore, and as described above, theoretical considerations and qualitative research suggest also direct effects of positive feedback on flow. We suggest that these effects of positive feedback on flow can be explained at least partially by increased self-efficacy. Accordingly, *we propose that self-efficacy acts as a mediator, transmitting positive effects of positive feedback on flow experience (Hypothesis 2).*

MATERIALS AND METHODS

Participants and Design

The sample was recruited at Ruhr University Bochum (Germany) through postings in social media networks, such as student groups in Facebook or via announcements on notice boards. The total sample consisted of 134 subjects (82 females, 52 males). Due to missing values, data from 23 subjects were excluded from the analyses: Eighteen participants did not complete the Flow-Short-Scale, one participant did not complete the self-efficacy scale and four participants did not perform the mental arithmetic task. The data was z-transformed and due to outliers on the study variables (flow, specific self-efficacy, performance quantity, and performance quality) another nine subjects were excluded[1]. The final analysis included data from 102 participants, of which 63 were female and 39 were male, with a mean age of 22.51 (SD_{age} = 3.13). Participants were mainly undergraduate students (76.5%). Another 20.6% were students with a bachelor's degree and 2.9% held a secondary school degree. Participants rated their ability in mental arithmetic on a 100-point scale on average at $M = 54.22$ (SD = 17.32). Furthermore, they rated the difficulty of the experimental task on an 8-point Likert Scale from 0 = "not difficult at all" to 7 = "very difficult" to be at an average level, with the second task being slightly more difficult than the first (M_{task1} = 3.87, SD_{task1} = 1.60; M_{task2} = 4.43, SD_{task2} = 1.63).

Participants were randomly assigned to one of two conditions: the positive feedback condition (n = 53, 33 female, M_{age} = 22.43, SD_{age} = 3.17) or the non-feedback condition (n = 49, 30 female, M_{age} = 22.59, SD_{age} = 3.12). All participants provided written informed consent and received course credit for participation. The study was approved by the local Ethics Committee of the Faculty of Psychology at Ruhr University Bochum, Germany.

Task

The experimental task was a computer-based mental arithmetic task, which lasted 5 min per block, with the participant sitting

[1]We performed all analyses with and without outliers and results did not change.

alone in the laboratory in front of a computer screen. A computer program written in VB.NET (Microsoft Visual Studio [Software], 2015) was used to generate the task on the screen. Participants typed the calculated numbers into the computer and pressed "enter" after each calculation. All the previously calculated numbers could be seen on the screen while working on the task. After 5 min, the task stopped automatically. In the first block of the mental arithmetic task, participants were asked to subtract the number 12, starting at 2000, consecutively with maximal accuracy and rapidness for 5 min. In the second block of the mental arithmetic task, participants were asked to subtract the number 17, starting at 2043, consecutively with maximal accuracy and rapidness for 5 min. The mental arithmetic task was constructed based on the Trier Social Stress Test (Kirschbaum et al., 1993), which uses a similar mental arithmetic task as part of the protocol. Importantly, and in contrast to the Trier Social Stress Test, there was no social stress component in our mental arithmetic task.

Feedback Manipulation

After the first block of the mental arithmetic task, the feedback group saw a note on their computer screen stating that their performance had been evaluated in terms of accuracy and rapidness. According to this analysis, he or she had performed better than the average of the previous participants, and that compared to the average participant, he or she was better able to follow new instructions and to manage mathematical problems spontaneously. The control group also saw a note on their screen, simply stating that time was up.

Procedure

The experiment was conducted in a laboratory room of Ruhr University Bochum. The participant was seated in front of a computer and asked to read and sign the informed consent form. Self-report measures and subsequent instructions were presented on the computer screen (see **Figure 1**). After a baseline measure of self-efficacy, the participant was asked to complete the first block of the mental arithmetic task (2000–12). Participants in the feedback group received positive normative feedback after the task was accomplished, while participants in the no-feedback group received no feedback. Right after this feedback, specific self-efficacy was assessed. Participants then completed the second block of mental arithmetic tasks (2043–17) for 5 min and finally were asked to answer questionnaires on flow and specific self-efficacy with respect to that task. Participants were then debriefed regarding the purpose of the study and the feedback manipulation.

Measures
Specific Self-Efficacy
Based on a guide for constructing self-efficacy scales developed by Bandura (2006), participants rated their ability to complete mental arithmetic tasks on a 10-point scale from 0 = "cannot do at all" to 100 = "can do very well" to measure their level of specific self-efficacy (M_{t1} = 52.84, SD_{t1} = 18.21) for mental arithmetic tasks. This item was assessed at baseline level (t0), after the first task following the feedback manipulation (t1), and

FIGURE 1 | Study procedure.
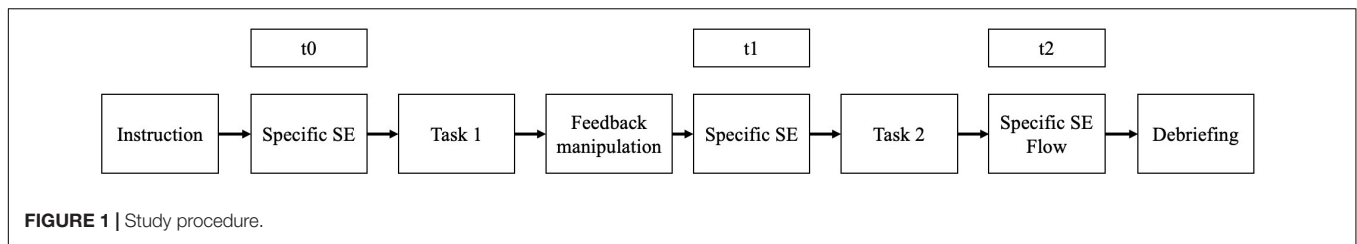
after the second task (compare **Figure 1**). The three measurement points t0, t1, and t2 were used for the manipulation check, that is, to test if the feedback manipulation increased specific self-efficacy over time. Specific self-efficacy at t1 was used to test the hypothesized mediation effects.

Performance

Performance quantity (M_{QN} = 28.33, SD_{QN} = 9.28) was assessed using the number of calculated results in the given time (5 min). Performance quality (M_{QL} = 0.90, SD_{QL} = 0.10) was assessed using the ratio between the number of correctly calculated results and the total number of calculated results.

Flow

Flow (M_{t2} = 4.50, SD_{t2} = 1.18) was measured with the Flow-Short-Scale (Rheinberg et al., 2003), which comprises ten items measuring absorption ("I did not notice time passing") and fluency ("My thoughts/activities ran fluidly and smoothly") as experienced during the task on a 7-point Likert Scale. The reliability of the scale was very good with a Cronbach's Alpha of 0.92. The scale was administered after the second task (at t2; compare **Figure 1**).

Data Analysis

Data were analyzed with the IBM SPSS statistics package. To analyze the efficiency of our manipulation, we tested whether participants' specific self-efficacy increased over time in the feedback group using a repeated measures ANOVA. As the Mauchly test of sphericity was significant, we used the Greenhouse-Geisser correction procedure. The mediation analyses were conducted with the SPSS macro Process (Hayes, 2013). For the mediation analyses, all variables were z-standardized. To estimate if self-efficacy served as the mediator, the indirect effect *ab* was estimated (Preacher and Hayes, 2008).

We report 95% confidence bootstrap intervals for the indirect effect ($n_{bootstrap}$ = 5000).

RESULTS

Table 1 shows means, standard deviations and correlations of all study variables.

Manipulation Check

With regard to the experimental manipulation of specific self-efficacy, a significant main effect of time [$F_{(1.82, 181.59)}$ = 4.97, p = 0.010, η_p^2 = 0.047] and an interaction effect [$F_{(1.82, 181.59)}$ = 6.23, p = 0.003, η_p^2 = 0.059] were found for specific self-efficacy for mental arithmetic tasks. As can be seen in **Figure 2**, in contrast to the feedback group, the control group showed a decrease in specific self-efficacy for mental arithmetic tasks after Task 1.

FIGURE 2 | Manipulation Check: Level of specific self-efficacy regarding mental arithmetic tasks at measurement points t0, t1, and t2 for the control group (no feedback) and the experimental group (positive feedback).
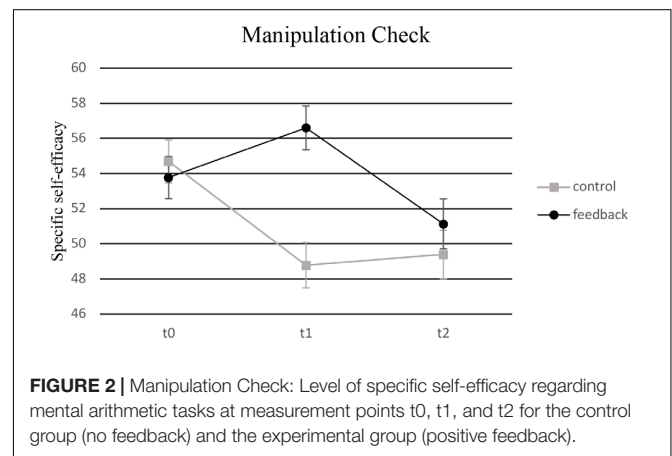
TABLE 1 | Shows means, standard deviations and correlations of all study variables.

Variable	MW	SD	MW$_{CG}$	SD$_{CG}$	MW$_{EG}$	SD$_{EG}$	1	2	3	4	5	6
1 Group	–	–	–	–	–	–	1					
2 Specific SE (t0)	54.22	17.32	54.69	16.85	53.77	17.89	−0.03	1				
3 Specific SE (t1)	52.84	18.21	48.78	17.75	56.60	17.96	0.22*	0.76**	1			
4 Specific SE (t2)	50.29	19.47	49.39	19.08	51.13	19.97	0.05	0.69**	0.79*	1		
5 Flow	4.50	1.18	4.43	1.14	4.57	1.23	0.06	0.46**	0.56**	0.67**	1	
6 Performance QL	0.90	0.10	0.90	0.09	0.90	0.11	−0.00	0.29**	0.30**	0.45**	0.40**	1
7 Performance QN	28.33	9.28	28.69	9.70	28.00	8.96	−0.04	0.39**	0.37**	0.45**	0.20*	0.32**

*p < 0.05; **p < 0.01.

FIGURE 3 | Mediation models of the effect of positive feedback on performance quantity **(A)** and performance quality **(B)** via specific self-efficacy. The arrows represent direct effects. The indirect effects were **(A)** β = 0.17 (SE = 0.10, 0.01 < CI < 0.39) for performance quantity and **(B)** β = 0.13 (SE = 0.08, 0.00 < CI < 0.32) for performance quality. *p < 0.05; **p < 0.01; ***p < 0.001.
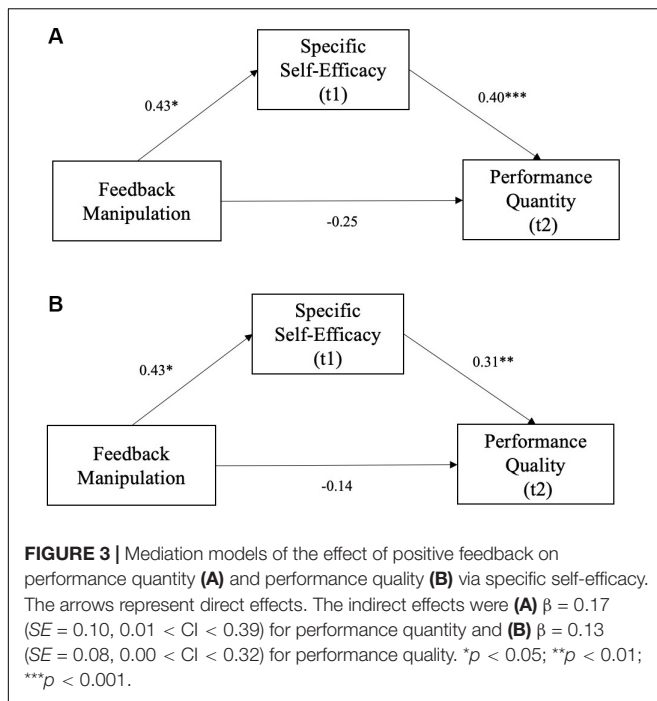
Testing of Hypotheses

To test *Hypothesis 1a*, we z-standardized all study variables and performed a mediation procedure with feedback as the independent variable, specific self-efficacy as the mediator, and performance quantity as the dependent variable. The *a*- (β = 0.43, SE = 0.19, p = 0.029) and *b*-path (β = 0.40, SE = 0.09, p < 0.001) were significant. The indirect effect was β = 0.17 (SE = 0.10, 0.01 < CI < 0.39) and significant (compare **Figure 3A**). The total effect was β = −0.07 (SE = 0.20, t = −0.38, p = 0.708) and unexpectedly *not* significant. The direct effect was also not significant (β = −0.25, SE = 0.19, t = −1.30, p = 0.195) when controlling for the indirect effect. However, as the indirect effect was significant, specific self-efficacy appeared to transmit a positive effect of positive feedback on performance quantity as hypothesized.

Regarding *Hypothesis 1b*, the *a*- (β = 0.43, SE = 0.19, p = 0.029) and *b*-path (β = 0.31, SE = 0.10, p = 0.002) of the model with specific self-efficacy as the mediator, group as the independent variable, and performance quality as the dependent variable were both significant. The indirect effect was β = 0.13 (SE = 0.08, 0.00 < CI < 0.32) and significant (compare **Figure 3B**). The total effect was β = −0.00 (SE = 0.20, t = −0.02, p = 0.988) and *not* significant. The direct effect was not significant (β = −0.14, SE = 0.20, t = −0.70, p = 0.483) when controlling for the indirect effect. As the b-path in this model was significantly positive, these results run counter to our expectation to find negative effects of increased self-efficacy on performance quality. We further found that self-efficacy appeared to transmit a positive effect of positive feedback on performance quality – while we had expected that a negative effect would be transmitted.

To conclude – in line with *Hypothesis 1a,* but in conflict with *Hypothesis 1b –* specific self-efficacy mediated a positive effect

FIGURE 4 | Mediation model of the effect of positive feedback on flow via specific self-efficacy. The arrows represent direct effects. The indirect effect was β = 0.25 (SE = 0.11, 0.02 < CI < 0.47). *p < 0.05; ***p < 0.001.

of positive feedback on both performance quantity and quality. Contrary to our expectations, however, the total effect of positive feedback on both performance measures was not significant.

Regarding *Hypothesis 2* the *a*- (β = 0.43, SE = 0.19, p = 0.029) and *b*-path (β = 0.57, SE = 0.09, p < 0.001) of the z-standardized mediation model with specific self-efficacy as the mediator, feedback as the independent variable, and flow as the dependent variable were signif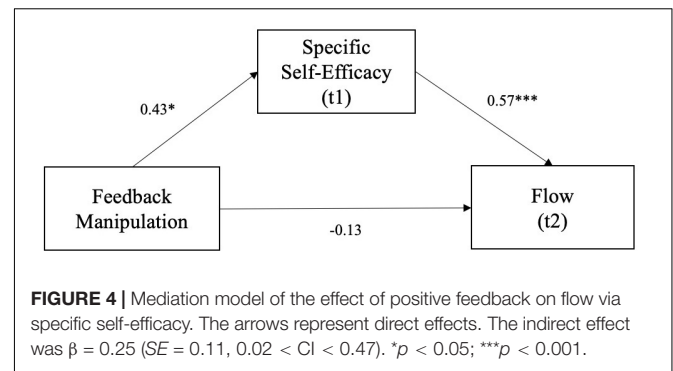icant. In support of *Hypothesis 2,* the indirect effect was β = 0.25 (SE = 0.11, 0.02 < CI < 0.47) and significant. The total effect was β = 0.12 (SE = 0.20, t = 0.61, p = 0.544) and *not* significant. The direct effect was not significant (β = −0.13, SE = 0.17, t = −0.74, p = 0.461) when controlling for the indirect effect (compare **Figure 4**). The significance of the indirect effect confirms *Hypothesis 2*, that specific self-efficacy would transmit a positive effect of positive feedback on flow. However, and contrary to our expectations, the total effect of positive feedback on flow was not significant.

DISCUSSION

In this study we aimed to test the postulated effect of self-efficacy on flow-experience and performance experimentally. To manipulate self-efficacy, we used a well-established paradigm, that is, false normative positive feedback about performance on a mental arithmetic task. Using a bootstrap procedure to conduct mediation analyses, we found that positive feedback enhances specific self-efficacy, which, in turn, enhances performance (quality and quantity) and flow experience in a subsequent task. In the following we discuss our results in more detail:

First of all, given our successful manipulation check, we could replicate earlier studies showing that false positive normative feedback is an efficient intervention to promote self-efficacy (Weinberg et al., 1979; Wright and Gregorich, 1989; Eden and Zuk, 1995; Brown et al., 2012). It can therefore be recommended as an experimental manipulation of self-efficacy in future studies.

Furthermore, *Hypothesis 1a* was confirmed: The results suggested that positive feedback has an indirect positive effect on performance quantity via self-efficacy. Considering the positive *b*-path in our model, that is, the positive effect of self-efficacy on performance quantity, our results replicate earlier studies that also found that performance increases with increased self-efficacy (e.g., Bandura and Locke, 2003; Vancouver et al., 2014).

As previous research on the relationship between self-efficacy and performance is largely based on observational data, a strength of our study is the experimental approach to manipulate self-efficacy using positive feedback.

By differentiating between performance quantity and quality, our study further contributes to the debate regarding whether – and for which performance parameters – self-efficacy has positive or maybe even negative effects on performance: some researchers argue that high self-efficacy might undermine motivation, as a person might believe that effort is not necessary to successfully cope with low demands compared to high abilities, which leads to an increase in performance quantity (as also found in our study), but to a decrease in performance quality (see, e.g., Vancouver et al., 2001, 2002, 2014; Vancouver, 2012). That is why we assumed in *Hypothesis 1b* that we would find a negative indirect effect of positive feedback via self-efficacy on performance quality. However, contrary to *Hypothesis 1b*, we found this indirect effect to be *positive*, with a positive effect of self-efficacy on performance quality. A possible explanation for the contradictory findings in the literature could be that the undermining effect of self-efficacy on performance quality only occurs if self-efficacy is very high. In our case, we told participants in the feedback condition that their performance was "above average", which is a relatively moderate manipulation. Thus, while we successfully increased self-efficacy levels, they only increased to a moderate but not to a very high level. This explanation would be in line with the proposed inverted u-shaped relationship between self-efficacy and performance (cf. Schönfeld et al., 2017): there could be an increase in performance until self-efficacy is moderately high and a decrease in performance if self-efficacy further increases.

Hypothesis 2 was also confirmed, with our results suggesting that positive feedback increases flow-experience via increased self-efficacy. The finding that experimentally induced self-efficacy increased flow-experience supports earlier cross-sectional field studies that have found positive associations between the two (Zubair and Kamal, 2015a,b) and longitudinal studies (Salanova et al., 2006; Rodríguez-Sánchez et al., 2011) with first evidence for a causal effect of self-efficacy on flow experience. By using an experimental manipulation to increase self-efficacy, we add further evidence to the existing literature that self-efficacy can causally increase flow-experience.

However, our study did not test the opposite causal direction, that experiencing flow would increase self-efficacy. Flow provides an enjoyable feeling of control over the activity at hand, while applying one's skills. Thus, it is likely that flow also enhances self-efficacy, and that the effects are bidirectional. This reciprocity suggests that an upward spiral of self-efficacy and flow can occur – as supported by earlier results from field studies (Salanova et al., 2006, 2014). Future experimental studies should look at both causal directions, replicating and complementing previous results.

Our results further showed an indirect effect of positive feedback on flow. This is in line with earlier theory and research: In his nine components of flow-experience, Csikszentmihalyi (1990) had already named "clear goals and feedback" as one of the components. Later operationalizations of flow distinguished between antecedents, characteristics and consequences of flow

and considered feedback as an antecedent (Nakamura and Csikszentmihalyi, 2002; Landhäußer and Keller, 2012). Cross-sectional field studies on feedback and flow supported this assumption (Rau and Riedel, 2004; Maeran and Cangiano, 2013). To the best of our knowledge, ours is the first study to show an indirect effect of *positive* feedback on flow in an experimental design, thereby providing insights into a mechanism that can transmit the effects of positive feedback to flow: self-efficacy. Accordingly, providing positive feedback enhances self-efficacy, which presumably enhances the feeling of competency and control in the respective task – two characteristics of flow experience.

However, it needs to be stated that while the indirect effect of positive feedback via self-efficacy on flow was significant, the total effect was not significant, and neither was the direct effect of positive feedback on flow when self-efficacy was included as a mediator. According to Hayes (2009, 2013), finding an indirect effect confirms mediation, while a missing total effect does not contradict mediation: "A failure to test for indirect effects in the absence of a total effect can lead to you miss some potentially interesting, important, or useful mechanisms by which X exerts some kind of effect on Y (Hayes, 2009; p. 415)." In cases in which the total effect is not significant, it is likely that different mechanisms play a role in the relationship between independent and dependent variable. This means that in addition to a positive effect of positive feedback on flow via enhanced self-efficacy, it is likely that there are counteracting mediators in this process, transmitting negative effects of positive feedback on flow. Such possible mediators should be investigated in future studies.

Similarly, the indirect effect of positive feedback via self-efficacy on performance quantity and quality were significant, while the total and direct effects were not significant. This again underlines the possibility of counteracting mechanisms between positive feedback and performance. While positive feedback via self-efficacy positively impacts performance, positive feedback might lead to the assumption that less effort is necessary, which could have a counteracting negative impact on performance (e.g., Vancouver et al., 2014). Such counteracting mechanisms of action should be examined in future studies, for example with the use of physiological indicators of mental effort, such as high frequency heart rate variability. Not only short-term but also long-term perspectives are relevant: an increase in self-efficacy through regular positive feedback might have long-term consequences on flow, and – through extensive application – on skill acquisition and future performance.

Another possible explanation for the lack of a total or direct effect of positive feedback on flow is that the feedback manipulation was potentially not salient enough: We told participants that they were better than average – which is what most people would assume anyway: the "better-than-average" effect is a robust finding in research on social comparisons (compare Alicke and Govorun, 2005). Accordingly, our positive feedback was potentially a confirmation of an existing assumption (i.e., neutral feedback) rather than a positive deviation. Future studies should use stronger positive feedback manipulations in order to investigate its effects on flow and performance.

Yet another explanation for the missing findings could be the kind of feedback that we used: Prior studies that addressed the relationship between flow mostly referred to either task-inherent feedback or supervisor feedback (mostly according to the Job Characteristics Model, Hackman et al., 1975). In their feedback measure, these studies did not differentiate between a normative vs. an individual reference norm (e.g., Bakker, 2005; Demerouti, 2006). Regarding performance there is research differentiating between normative and individual feedback: Brunstein and Hoyer (2002), for example, investigated the effects of positive vs. negative feedback with normative vs. individual reference norm on performance. They did not find any main effects of positive normative feedback on performance (i.e., d2-test for concentration). Descriptively, positive normative feedback was even associated with lower performance (higher reaction times). In addition, Brunstein and Hoyer (2002) found significant interactions with the achievement motive: A higher implicit (but not explicit) achievement motive was associated with higher performance after feedback. Interestingly, it was not positive normative feedback but negative individual feedback that spurred achievement-motivated individuals' performance. These results show that future studies should explicitly differentiate between normative and individual feedback to get deeper insights into the mechanisms that facilitate or hinder flow experience and performance. As moderating variables, implicit and explicit achievement motives should be controlled, as differential effects might occur. Furthermore, feedback with an individual reference norm may be more appropriate to foster performance and flow (Brunstein and Hoyer, 2002; Brunstein and Maier, 2005).

While Brunstein and Hoyer found that performance was enhanced by negative feedback at least for individuals scoring high on the implicit achievement motive, we do not know how negative feedback impacts flow. In theoretical models (Nakamura and Csikszentmihalyi, 2002; Landhäußer and Keller, 2012) as well as in previous (field) studies (Rau and Riedel, 2004; Maeran and Cangiano, 2013) on the relationship between feedback and flow, positive and negative feedback were not distinguished. However, there is evidence from qualitative research that positive feedback is especially beneficial for flow (Jackson, 1995; Swann et al., 2015). Thus, it will be interesting for future studies to look at the differential effects of positive vs. negative feedback, differentiating between task-inherent, normative and individual feedback, and controlling for an individual's achievement motive.

A potential limitation of our results refers to the fact that we did not use change scores of performance in the mental arithmetic tasks to control for baseline scores. However, based on the design of our study, the interpretation of change scores is problematic: The two tasks differed very much in difficulty: while the first task was very easy (continuously subtracting 12 from 2000), the second task was more difficult (continuously subtracting 17 from 2043). Our results clearly support this, as both performance indicators decreased significantly from t1 to t2. Furthermore, the low difficulty in the first task likely led to a ceiling effect of performance in both groups, reducing the systematic variance in the data. These circumstances make the change scores very hard to interpret. At the same time, as we had randomly assigned our participants to the experimental

conditions, we believe that mental arithmetic skills are equally distributed between groups. Accordingly, we refrained from using change scores in our data analysis. Detailed results using change scores in the analyses can be found in the **Supplementary Material**.

One more possible limitation of our study refers to the measurement of flow: By using the flow short scale (Engeser and Rheinberg, 2008), we applied a widely used componential approach to assess flow (compare Moneta, 2012). This scale measures flow as a continuous phenomenon: the more its components are pronounced, the higher flow values. The components used in the flow short scale reflect those proposed by Csikszentmihalyi (1975, see flow definition above). However, an ongoing discussion in flow research is whether flow is a continuous phenomenon or if it is rather a yes-or-no phenomenon (compare Engeser, 2012; Peifer and Engeser, in press), with an individual either being in flow or not. A cut-off-value for flow when measured with the componential approach has not yet been identified and is a challenge for future research. With what we know today, the current study is limited insofar as it cannot differentiate between flow or not-flow, but rather measures more or less pronounced flow components.

Practical Implications

Based on our results, we can recommend positive feedback as an intervention to enhance self-efficacy. This recommendation can be applied to different contexts such as work or schools. In the work context, positive feedback could be given by supervisors in annual performance reviews and in regular meetings. A precondition to providing clear feedback is goal setting, as expectations are made transparent to the employee. Goals should be realistic and achievable in due time in order to provide the opportunity for positive feedback on a regular basis. In general, positive feedback can also come from sources other than just the supervisor – for example from customers or colleagues. Positive feedback has further been found to be positively related to job satisfaction (Yukl et al., 2002) and well-being (Stocker et al., 2014). Thus, an organizational climate of mutual appreciation is a good basis for employees' self-efficacy, satisfaction and well-being. As shown in the current study, self-efficacy is also positively related to flow and performance and acts as a mediator transmitting positive effects of positive feedback to flow and performance.

While we manipulated self-efficacy in this study using positive feedback, there are other interventions that have been successfully used to enhance self-efficacy, for example interventions to increase psychological capital (Luthans et al., 2006, 2008, 2010). These interventions could therefore also be used to increase self-efficacy, and, in turn, performance and flow-experience.

In this study, we found short-term and immediate effects of self-efficacy on flow-experience. Future studies should have a look at spillover effects and at long-term effects of self-efficacy interventions: Enhancing flow-experience has been found to affect future flow and performance in similar tasks (Christandl et al., 2018). Furthermore, flow has been found to lead to long-term increased performance via increased motivation to

practice (Engeser et al., 2005; Schüler, 2007; Schüler and Brunner, 2009). Thus, interventions to foster the pleasant experience of flow are a valuable endeavor for institutions (e.g., organizations, schools) as well as for individuals (e.g., employees, students).

CONCLUSION

"Well done?" – our study investigated the relationships between positive normative feedback, self-efficacy, performance, and flow experience. Our results provide experimental evidence that positive feedback enhances self-efficacy. Further, we found an indirect effect of feedback via self-efficacy on performance quantity and quality, as well as on flow experience. However, mutually opposing counter-mechanisms were potentially also active as we did not find a total effect of positive normative feedback on performance and flow, calling for further research on this issue.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with

the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

INFORMED CONSENT

Informed consent was obtained from all individual participants included in the study.

AUTHOR CONTRIBUTIONS

CP and PS conceived of the presented idea. PS carried out the experiment. CP, PS, and FA developed the theory. PS and GW wrote the methods. GW and FA performed the computations and wrote the results part. CP wrote the discussion. JM supervised the concept and findings of this work. All authors discussed the results and contributed to the final manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01008/full#supplementary-material

REFERENCES

Achterkamp, R., Hermens, H. J., and Vollenbroek-Hutten, M. M. R. (2015). The influence of success experience on self-efficacy when providing feedback through technology. *Comput. Hum. Behav.* 52, 419–423.

Alicke, M. D., and Govorun, O. (2005). "The better-than-average effect," in *The Self in Social Judgment*, eds M. D. Alicke, D. A. Dunning, and J. Krueger (New York, NY: Psychology Press), 85–106.

Arbel, Y., Murphy, A., and Donchin, E. (2014). On the utility of positive and negative feedback in a paired-associate learning task. *J. Cogn. Neurosci.* 26, 1445–1453. doi: 10.1162/jocna00617

Bakker, A. B. (2005). Flow among music teachers and their students: the crossover of peak experiences. *J. Vocat. Behav.* 66, 26–44. doi: 10.1016/j.jvb.2003.11.001

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioural change. *Psychol. Rev.* 84, 191–215. doi: 10.1037//0033-295x.84.2.191

Bandura, A. (1983). Self-efficacy determinants of anticipated fears and calamities. *J. Pers. Soc. Psychol.* 45, 464–469. doi: 10.1037/0022-3514.45.2.464

Bandura, A. (1986). *Social Foundations of thought and Action.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control.* New York, NY: Freeman.

Bandura, A. (2006). "Guide for constructing self-efficacy scales," in *Self-Efficacy Beliefs of Adolescents*, Vol. 5, eds F. Pajares, and T. Urdan (Greenwich, CT: Information Age Publishing), 307–337.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *J. Manage.* 38, 9–44. doi: 10.1177/0149206311410606

Bandura, A., and Jourden, F. J. (1991). Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *J. Pers. Soc. Psychol.* 60, 941–951.

Bandura, A., and Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *J. Appl. Psychol.* 88, 87–99. doi: 10.1037/0021-9010.88.1.87

Beattie, S., Fakehy, M., and Woodman, T. (2014). Examining the moderating effects of time on task and task complexity on the within person self-efficacy and performance relationship. *Psychol. Sport Exerc.* 15, 605–610.

Beattie, S., Woodman, T., Fakehy, M., and Dempsey, C. (2016). The role of performance feedback on the self-efficacy—performance relationship. *Sport Exerc. Perform. Psychol.* 5:1.

Benight, C. C., Ironson, G., Klebe, K., Carver, C. S., Wynings, C., Burnett, K., et al. (1999). Conservation of resources and coping self-efficacy predicting distress following a natural disaster: a causal model analysis where the environment meets the mind. *Anxiety Stress Coping* 12, 107–126. doi: 10.1080/10615809908248325

Brown, A. D., Dorfman, M. L., Marmar, C. R., and Bryant, R. A. (2012). The impact of perceived self-efficacy on mental time travel and social problem solving. *Conscious. Cogn.* 21, 299–306. doi: 10.1016/j.concog.2011.09.023

Brown, G. T. L., Peterson, E. R., and Yao, E. S. (2016). Student conceptions of feedback: impact on self-regulation, self-efficacy, and academic achievement. *Br. J. Educ. Psychol.* 86, 606–626. doi: 10.1111/bjep.12126

Brunstein, J. C., and Hoyer, S. (2002). Implizites versus explizites Leistungsstreben: befunde zur Unabhängigkeit zweier Motivationssysteme. *Z. Pädagog. Psychol.* 16, 51–62.

Brunstein, J. C., and Maier, G. W. (2005). Implicit and self-attributed motives to achieve: two separate but interacting needs. *J. Pers. Soc. Psychol.* 89, 205–222. doi: 10.1037/0022-3514.89.2.205

Chiviacowsky, S., and Wulf, G. (2007). Feedback after good trials enhances learning. *Res. Q. Exerc. Sport* 78, 40–47. doi: 10.1080/02701367.2007.10599402

Christandl, F., Mierke, K., and Peifer, C. (2018). Time flows: manipulations of subjective time progression affect recalled flow and performance in a subsequent task. *J. Exp. Soc. Psychol.* 74, 246–256.

Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*. San Francisco, CA: Jossey-Bass Publishers.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York, NY: Harper and Row.

Demerouti, E. (2006). Job characteristics, flow, and performance: the moderating role of conscientiousness. *J. Occup. Health Psychol.* 11, 266–280. doi: 10.1037/1076-8998.11.3.266

Dimotakis, N., Mitchell, D., and Maurer, T. (2017). Positive and negative assessment center feedback in relation to development self-efficacy, feedback seeking, and promotion. *J. Appl. Psychol.* 102, 1514–1527. doi: 10.1037/apl0000228

Eden, D., and Zuk, Y. (1995). Seasickness as a self-fulfilling prophecy – raising self-efficacy to boost performance at sea. *J. Appl. Psychol.* 80, 628–635. doi: 10.1037/0021-9010.80.5.628

Engeser, S. (2012). "Theoretical integration and future lines of flow research," in *Advances in Flow Research*, ed. S. Engeser (New York, NY: Springer), 187–199.

Engeser, S., and Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motiv. Emot.* 32, 158–172. doi: 10.1007/s11031-008-9102-4

Engeser, S., Rheinberg, F., Vollmeyer, R., and Bischoff, J. (2005). Motivation, Flow-Erleben und Lernleistung in universitären Lernsettings [motivation, flow-experience and learning performance in universities' learning settings]. *Z. Pädagog. Psychol.* 19, 159–172. doi: 10.1024/1010-0652.19.3.159

Gist, M. E., and Mitchell, T. R. (1992). Self-efficacy: a theoretical analysis of its determinants and malleability. *Acad. Manage. Rev.* 17, 183–211.

Hackman, J. R., Oldham, G., Janson, R., and Purdy, K. (1975). A new strategy for job enrichment. *Calif. Manage. Rev.* 17, 57–71. doi: 10.1016/j.pbb.2015.09.009

Hardy, J. H. III (2014). Dynamics in the self-efficacy–performance relationship following failure. *Pers. Individ. Dif.* 71, 151–158.

Harmat, L., de Manzano, Ö., Theorell, T., Högman, L., Fischer, H., and Ullén, F. (2015). Physiological correlates of the flow experience during computer game playing. *Int. J. Psychophysiol.* 97, 1–7. doi: 10.1016/j.ijpsycho.2015.05.001

Hartwell, C. J., and Campion, M. A. (2016). Getting on the same page: the effect of normative feedback interventions on structured interview ratings. *J. Appl. Psychol.* 101, 757–778. doi: 10.1037/apl0000099

Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487

Hayes, A. F. (2009). Beyond Baron and Kenny: statistical mediation analysis in the new millennium. *Commun. Monogr.* 76, 408–420.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.

Hutchinson, J. C., Sherman, T., Martinovic, N., and Tenenbaum, G. (2008). The effect of manipulated self-efficacy on perceived and sustained effort. *J. Appl. Sport Psychol.* 20, 457–472. doi: 10.1080/10413200802351151

Jackson, S. A. (1995). Factors influencing the occurrence of flow state in elite athletes. *J. Appl. Sport Psychol.* 7, 138–166. doi: 10.1080/10413209508406962

Kirschbaum, C., Pirke, K. M., and Hellhammer, D. H. (1993). The Trier Social Stress Test - A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28, 76–81. doi: 10.1159/000119004

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

Kühn, A. A., Brücke, C., Hübl, J., Schneider, G. H., Kupsch, A., Eusebio, A., et al. (2008). Motivation modulates motor-related feedback activity in the human basal ganglia. *Curr. Biol.* 18, R648–R650. doi: 10.1016/j.cub.2008.06.003

Landhäußer, A., and Keller, J. (2012). "Flow and its affective, cognitive, and performance-related consequences," in *Advances in Flow Research*, ed. S. Engeser (Berlin: Springer), 65–85.

Luthans, F., Avey, J. B., Avolio, B. J., Norman, S. M., and Combs, G. M. (2006). Psychological capital development: toward a micro-intervention. *J. Organ. Behav.* 27, 387–393. doi: 10.1002/job.373

Luthans, F., Avey, J. B., Avolio, B. J., and Peterson, S. J. (2010). The development and resulting performance impact of positive psychological capital. *Hum. Resour. Dev. Q.* 21, 41–67. doi: 10.1002/hrdq.20034

Luthans, F., Avey, J. B., and Patera, J. L. (2008). Experimental analysis of a web-based training intervention to develop positive psychological capital. *Acad. Manage. Learn. Educ.* 7, 209–221. doi: 10.5465/AMLE.2008.32712618

Luthans, F., Luthans, K. W., and Luthans, B. C. (2004). Positive psychological capital: beyond human and social capital. *Bus. Horiz.* 47, 45–50. doi: 10.1016/j.bushor.2003.11.007

Maeran, R., and Cangiano, F. (2013). Flow experience and job characteristics: analyzing the role of flow in job satisfaction. *TPM Test. Psychometr. Methodol. Appl. Psychol.* 20, 13–26.

Microsoft Visual Studio [Software] (2015). Available online at: https://www.visualstudio.com/ (accessed April 12, 2015).

Moneta, G. B. (2012). "On the measurement and conceptualization of flow," in *Advances in Flow Research*, ed. S. Engeser (New York, NY: Springer), 23–50.

Nakamura, J., and Csikszentmihalyi, M. (2002). "The concept of flow," in *The Handbook of Positive Psychology*, eds C. R. Snyder and S. J. Lopez (New York, NY: Oxford University Press), 89–105. doi: 10.1007/978-94-017-9088-8_16

Nakamura, J., and Csikszentmihalyi, M. (2014). "The concept of flow," in *Flow and the Foundations of Positive Psychology*, ed. M. Csikszentmihalyi (Dordrecht: Springer, 239–263.

Niemiec, T., and Lachowicz-Tabaczek, K. (2015). The moderating role of specific self-efficacy in the impact of positive mood on cognitive performance. *Motiv. Emot.* 39, 498–505. doi: 10.1007/s11031-014-9469-3

Peifer, C., and Engeser, S. (in press). "Theoretical Integration and future lines of flow research," in *Advances in Flow Research*, eds C. Peifer and S. Engeser (New York, NY: Springer).

Pineau, T. R., Glass, C. R., Kaufman, K. A., and Bernal, D. R. (2014). Self- and team-efficacy beliefs of rowers and their relation to mindfulness and flow. *J. Clin. Sport Psychol.* 8, 142–158. doi: 10.1123/jcsp.2014-0019

Powers, W. T. (1973). *Behaviour: The Control of Perception*. Chicago, IL: Aldine.

Powers, W. T. (1991). Commentary on Bandura's "human agency". *Am. Psychol.* 46, 151–153.

Preacher, K. J., and Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods* 40, 879–891. doi: 10.3758/BRM.40.3.879

Rau, R., and Riedel, S. (2004). Besteht ein Zusammenhang zwischen dem Auftreten von positivem Arbeitserleben unter Flow-Bedingungen und Merkmalen der Arbeitstätigkeit? *Z. Arbeits Organisationspsychol.* 48, 55–66. doi: 10.1026/0932-4089.48.2.55

Reynolds, D. (2006). To what extent dies performance-related feedback affect managers' self-efficacy? *Int. J. Hosp. Manage.* 25, 54–68. doi: 10.1016/j.ijhm.2004.12.007

Rheinberg, F., Vollmeyer, R., and Engeser, S. (2003). "Die Erfassung des Flow-Erlebens [Measuring flow-experience]," in *Diagnostik von Motivation und Selbstkonzept*, eds J. Stiensmeier-Pelster and F. Rheinberg (Göttingen: Hogrefe), 261–279.

Rodríguez-Sánchez, A., Salanova, M., Cifre, E., and Schaufeli, W. B. (2011). When good is good: a virtuous circle of self-efficacy and flow at work among teachers. *Rev. Psicol. Soc.* 26, 427–441.

Rottmann, N., Dalton, S. O., Christensen, J., Frederiksen, K., and Johansen, C. (2010). Self-efficacy, adjustment style and well-being in breast cancer patients: a longitudinal study. *Qual. Life Res.* 19, 827–836. doi: 10.1007/s11136-010-9653-1

Salanova, M., Bakker, A. B., and Llorens, S. (2006). Flow at work: evidence for an upward spiral of personal and organizational resources. *J. Happiness Stud.* 7, 1–22. doi: 10.1007/s10902-005-8854-8

Salanova, M., Rodríguez-Sánchez, A. M., Schaufeli, W. B., and Cifre, E. (2014). Flowing together: a longitudinal study of collective efficacy and collective flow among workgroups. *J. Psychol.* 148, 435–455. doi: 10.1080/00223980.2013.806290

Schönfeld, P., Brailovskaia, J., Bieda, A., Zhang, X. C., and Margraf, J. (2016). The effects of daily stress on positive and negative mental health: mediation through self-efficacy. *Int. J. Clin. Health Psychol.* 16, 1–10. doi: 10.1016/j.ijchp.2015.08.005

Schönfeld, P., Preusser, F., and Margraf, J. (2017). Costs and benefits of self-efficacy: differences of the stress response and clinical implications. *Neurosci. Biobehav. Rev.* 75, 40–52. doi: 10.1016/j.neubiorev.2017.01.031

Schüler, J. (2007). Arousal of flow experience in a learning setting and its effects on exam performance and affect. *Z. Pädagog. Psychol.* 21, 217–227. doi: 10.1024/1010-0652.21.3.217

Schüler, J., and Brunner, S. (2009). The rewarding effect of flow experience on performance in a marathon race. *Psychol. Sport Exerc.* 10, 168–174. doi: 10.1016/j.psychsport.2008.07.001

Singh, P., and Bussey, K. (2011). Peer victimization and psychological maladjustment: the mediating role of coping self-efficacy. *J. Res. Adolesc.* 21, 420–433.

Stocker, D., Jacobshagen, N., Krings, R., Pfister, I. B., and Semmer, N. K. (2014). Appreciative leadership and employee well-being in everyday working life. *Z. Personalforsch.* 28, 73–95.

Swann, C., Crust, L., Keegan, R., Piggott, D., and Hemmings, B. (2015). An inductive exploration into the flow experiences of European tour golfers. *Qual. Res. Sport Exerc. Health* 7, 210–234. doi: 10.1080/2159676X.2014.926969

Ulrich, M., Keller, J., and Grön, G. (2016). Dorsal raphe nucleus down-regulates medial prefrontal cortex during experience of flow. *Front. Behav. Neurosci.* 10:169. doi: 10.3389/fnbeh.2016.00169

Vancouver, J. B. (2012). Rhetorical reckoning a response to Bandura. *J. Manage.* 38, 465–474.

Vancouver, J. B., Gullekson, N. L., Morse, B. J., and Warren, M. A. (2014). Finding a between-person negative effect of self-efficacy on performance: not just a within-person effect anymore. *Hum. Perform.* 27, 243–261.

Vancouver, J. B., and Kendall, L. N. (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *J. Appl. Psychol.* 91, 1146–1153. doi: 10.1037/0021-9010.91.5.1146

Vancouver, J. B., More, K. M., and Yoder, R. J. (2008). Self-efficacy and resource allocation: support for a nonmonotonic, discontinuous model. *J. Appl. Psychol.* 93, 35–47. doi: 10.1037/0021-9010.93.1.35

Vancouver, J. B., Thompson, C. M., Tischner, E. C., and Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *J. Appl. Psychol.* 87, 506–516. doi: 10.1037/0021-9010.87.3.506

Vancouver, J. B., Thompson, C. M., and Williams, A. A. (2001). The changing signs in the relationships among self-efficacy, personal goals, and performance. *J. Appl. Psychol.* 86, 605–620. doi: 10.1037//0021-9010.86.4.605

Wang, Y., Yao, L., Liu, L., Yang, X., Wu, H., Wang, J., et al. (2014). The mediating role of self-efficacy in the relationship between big five personality and depressive symptoms among Chinese unemployed population: a cross-sectional study. *BMC Psychiatry* 14:61. doi: 10.1186/1471-244X-14-61

Weinberg, R. S., Gould, D., and Jackson, A. (1979). Expectations and performance: an empirical test of Bandura's self-efficacy theory. *J. Sport Psychol.* 1, 320–331.

Wright, R. A., and Gregorich, S. (1989). Difficulty and instrumentality of imminent behaviour as determinants of cardiovascular response and self-reported energy. *Psychophysiology* 26, 586–592. doi: 10.1111/j.1469-8986.1989.tb00715.x

Wulf, G., Chiviacowsky, S., and Lewthwaite, R. (2010). Normative feedback effects on learning a timing task. *Res. Q. Exerc. Sport* 81, 425–431. doi: 10.1080/02701367.2010.10599703

Yukl, G., Gordon, A., and Raber, T. (2002). A hierarchical taxonomy of leadership behavior: integrating a half century of behavior research. *J. Leadersh. Organ. Stud.* 9, 15–32.

Zinken, K. M., Cradock, S., and Skinner, T. C. (2008). Analysis system for self-efficacy training (ASSET): assessing treatment fidelity of self-management interventions. *Pat. Educ. Couns.* 72, 186–193. doi: 10.1016/j.pec.2008.04.006

Zlomuzica, A., Preusser, F., Schneider, S., and Margraf, J. (2015). Increased perceived self-efficacy facilitates the extinction of fear in healthy participants. *Front. Behav. Neurosci.* 9:270. doi: 10.3389/fnbeh.2015.00270

Zubair, A., and Kamal, A. (2015a). Authentic leadership and creativity: mediating role of work-related flow and psychological capital. *J. Behav. Sci.* 25:150.

Zubair, A., and Kamal, A. (2015b). Work related flow, psychological capital, and creativity among employees of software houses. *Psychol. Stud.* 60, 321–331. doi: 10.1007/s12646-015-0330-x

frontiers
in Human Neuroscience

Dissociable Effects of Reward on P300 and EEG Spectra Under Conditions of High vs. Low Vigilance During a Selective Visual Attention Task

Jia Liu [1,2], Chi Zhang [1], Yongjie Zhu [1,2], Yunmeng Liu [1], Hongjin Sun [3], Tapani Ristaniemi [2], Fengyu Cong [1,2,4,5]* and Tiina Parviainen [6]*

[1]School of Biomedical Engineering, Dalian University of Technology, Dalian, China, [2]Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland, [3]Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, ON, Canada, [4]School of Artificial Intelligence, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China, [5]Key Laboratory of Integrated Circuit and Biomedical Electronic System, Dalian University of Technology, Dalian, China, [6]Centre for Interdisciplinary Brain Research, Department of Psychology, Faculty of Education and Psychology, University of Jyvaskyla, Jyvaskyla, Finland

The influence of motivation on selective visual attention in states of high vs. low vigilance is poorly understood. To explore the possible differences in the influence of motivation on behavioral performance and neural activity in high and low vigilance levels, we conducted a prolonged 2 h 20 min flanker task and provided monetary rewards during the 20- to 40- and 100- to 120-min intervals of task performance. Both the behavioral and electrophysiological measures were modulated by prolonged task engagement. Moreover, the effect of reward was different in high vs. low vigilance states. The monetary reward increased accuracy and decreased the reaction time (RT) and number of omitted responses in the low but not in the high vigilance state. The fatigue-related decrease in P300 amplitude recovered to its level in the high vigilance state by manipulating motivation, whereas the fatigue-related increase in P300 latency was not modulated by reward. Additionally, the fatigue-related increase in event-related spectral power at 1–4 Hz was sensitive to vigilance decrement and reward. However, the spectral power at 4–8 Hz was only affected by the decrease in vigilance. These electrophysiological measures were not influenced by motivation in the state of high vigilance. Our results suggest that neural processing capacity, but not the timing of processing, is sensitive to motivation. These findings also imply that the fatigue-related impairments in behavioral performance and neural activity underlying selective visual attention only partly recover after manipulating motivation. Furthermore, our results provide evidence for the dissociable neural mechanisms underlying the fatigue-related decrease vs. reward-related increase in attentional resources.

Keywords: vigilance, mental fatigue, motivation, selective visual attention, event-related potential, event-related spectral perturbation

HIGHLIGHTS

- Time-on-task impairs performance in a selective visual attention task.
- Lower vigilance is associated with decreased amplitude and increased latency of P300.
- Monetary rewards improve P300 amplitude, but not P300 latency.
- The event-related spectral power at 1–4 Hz is sensitive to vigilance decrement and reward, whereas the spectral power at 4–8 Hz is sensitive only to vigilance decrement.
- Reward improves P300 amplitude and spectral power at 1–4 Hz only in the low and not in the high vigilance state.

INTRODUCTION

Although we are subjected to constant visual information in daily life, our visual capacity to process this information is limited. To perform in an efficient and goal-directed manner, we need to continuously distinguish the relevant information from the visual environment and allocate our limited attentional capacity to the selected target objects, a phenomenon referred to as selective visual attention (Moore and Zirnsak, 2017). As outlined by Robert and Duncan (1995), selective visual attention is characterized by two basic phenomena: the ability to filter out task-irrelevant stimuli and the limited capacity for task-relevant information processing, both of which leading to reduced accuracy when the target number increases. We experience selective visual attention in many daily activities. For example, customers find the target objects among colorful irrelevant sales; car drivers filter out irrelevant surroundings and detect the relevant road marks and traffic lights. However, prolonged engagement in selective attention tasks inevitably leads to increased errors, deactivated performance goals, diminished motivation to continue performing the task (Boksem et al., 2005), and an increase in mental fatigue (Kok, 2001; Lal and Craig, 2001; Gergelyfi et al., 2015; Benoit et al., 2019).

Mental fatigue is caused by prolonged cognitive task performance (Gergelyfi et al., 2015). It is considered a related concept but distinct from arousal, which often refers to a physiological state and is closely linked with the transition between wakefulness and sleep (Shen et al., 2006). Mental fatigue is a cumulative process, accompanied by a feeling of indolence, reduced motivation, and impaired performance (Lal and Craig, 2001). What is more, mental fatigue exhibits more cognitive elements than arousal. Based on different causal factors, two types of mental fatigue can be identified: sleep- and task-related (May and Baldwin, 2009). The former results from accumulated sleep debt, whereas the latter from prolonged task engagement (May and Baldwin, 2009). In the present study, we aimed to examine the task-related mental fatigue, specifically related to attentional resources. The attention-requiring task performance over a prolonged duration pointedly refers to vigilance decrement (Mackworth et al., 1964), which is likely identical or very closely related to mental fatigue (Oken et al., 2006). For this reason, both terms have been used interchangeably in previous studies (Taya et al., 2018; Reteig et al., 2019).

Vigilance decrement has been reported as a major factor in a large proportion of road crashes due to the reduction of attentional resources. Although the risks of vigilance decrement have received much attention, the underlying neurophysiological mechanisms have not yet been established (Lorist et al., 2005; Tops et al., 2006; Benoit et al., 2019). In earlier research, three core concepts around vigilance decrement or mental fatigue have emerged, namely, active fatigue, passive fatigue, and motivational control. Active fatigue is a result of an excessive workload—needed to carry out a task over a prolonged duration, resulting in the depletion of cognitive resources (Helton and Warm, 2008). Passive fatigue is a result of a lower workload—needed to engage in prolonged, but relatively easy tasks (May and Baldwin, 2009). Motivational control plays an important role in vigilance decrement, as it reflects the level of willingness to perform a task. Motivational control is linked with the process of subconscious balancing between costs and benefits to expend or conserve energy (Kurzban et al., 2013a). For instance, Kurzban and colleagues suggested that people experienced performance reductions over time when the costs outweighed the benefits (Kurzban et al., 2013b). Recent studies recognize that these three core concepts are not mutually exclusive, and there are still limitations in the core concepts account for changes induced by fatigue (Boksem and Tops, 2008; Seli et al., 2015; Thomson et al., 2015). Therefore, the hybrid models synthesizing different concepts have emerged to complement the limitations. For example, Boksem and Tops (2008) proposed a framework of mental fatigue that integrated the motivational control and energetical costs, suggesting that people would no longer maintain their performance when the energetical resources depleted, although the costs outweighed the benefits. All in all, it is still unclear why task performance deteriorates with time-on-task.

The influence of motivation on prolonged task performance has been studied by subsequently providing monetary rewards. The effects on response selection (Möckel et al., 2015), action monitoring (Boksem et al., 2006), and sustained attention (Reteig et al., 2019) have been previously shown. Although numerous studies have demonstrated that monetary rewards can improve performance when provided after long-term tasks (Lorist et al., 2005; Boksem et al., 2006; Hopstaken et al., 2015), the neural mechanisms upon which this improvement builds on are not established. Moreover, the effect of reward on performance in different (i.e., high vs. low) vigilance states has rarely been approached.

To explore the effects of motivation on behavioral performance and brain electrophysiology in high and low vigilance states, we conducted a 140-min selective visual attention task and provided monetary rewards for successful task performance in the early stage (during the 20- to 40-min interval) and in the late stage (during the 100- to 120-min interval; **Figure 1**). By utilizing brain electrophysiological measures derived from high-temporal-resolution electroencephalograms (EEGs), we focused on time domain [event-related potential (ERP) P300 amplitude and latency] and time-frequency

domain [event-related spectral perturbations (ERSPs)] variables as electrophysiological markers of visually induced neural activations. We further quantified the degree of recovery of behavioral and electrophysiological measures in the low vigilance state after motivation manipulation.

The stimulus-locked ERP component P300 has received much attention as a potential indicator of mental workload in a selective visual attention task (Faber et al., 2012). The amplitude of P300 was proved to be a useful measure of processing capacity that correlates positively with the accuracy of the memory search task (Kok, 2001). Furthermore, the latency of P300 was suggested to be an indicator of mental chronometry as demonstrated by its positive correlation with reaction time (RT; Verleger, 1997). While reports about the effect of time-on-task on the P300 component are diverse, the study of Faber et al. (2012) did not find a significant decrease in the P3b amplitude during prolonged engagement in a selective visual attention task. Boksem et al. (2006) also showed that the P300 amplitude did not change with time-on-task, but the P300 latency increased with vigilance decrement. Although the P300 amplitude and latency have been widely used in studies on vigilance (Kato et al., 2009; Käthner et al., 2014; Hopstaken et al., 2015), most results are limited to conventional ERP analysis.

It is also valuable to explore how the oscillatory dynamics reflect changes in attentional allocation and information processing during a selective visual attention task. Frontal theta oscillations have been shown to be related to the allocation of attention to task-relevant visual and auditory stimuli (Keller et al., 2017). Oscillations in the delta band have been implicated in attention and salience detection and are associated with vigilance levels and motivation (Knyazev, 2012). It has also been suggested that EEG delta oscillations are an indicator of attention to internal processing during the performance of mental tasks (Harmony et al., 1996). Compared with traditional time- and phase-locked ERP analysis, the changes in spectral power provided by two-dimensional time-frequency analysis could provide a better account of the neural mechanisms involved in selective visual attention. In the current study, besides the evoked P300 component, we will analyze the ERSPs.

We hypothesize that vigilance decrement induced by prolonged engagement in a selective visual attention task impairs behavioral performance and neural activity and is evident in P300 latency and amplitude. We further hypothesize that monetary rewards improve the behavioral performance and neural activity in the low vigilance state. We apply a variant of the Eriksen Flanker Task conducted over 2 h 20 min (seven blocks) and assume that the subjects are in a lower vigilance state at the end of the task (blocks 5 and 6) than at the beginning (blocks 1 and 2). To compare the effects of motivation on performance in states of high vs. low vigilance, we introduce rewards in block 2 (during 20–40 min after task onset) and block 6 (during 100–120 min after task onset). The behavioral performance, evoked ERPs, and ERSPs were compared between high and low vigilance states with and without rewards.

MATERIALS AND METHODS

Subjects

Twenty healthy participants (eight males), ranging from 18 to 28 [mean = 21.9, standard deviation (SD) = 2.4] years of age, were recruited from the university population. Participants reported that they had no history of smoking, sleep problems, or use of prescription medication. None worked the night shift. Furthermore, they all had normal or corrected-to-normal visual acuity, and they were right-handed according to their own report. The participants were compensated for their participation. The study was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee of Liaoning Normal University. Informed consent was obtained from each subject prior to the study.

Measures

Task and Stimuli

A version of the Eriksen Flanker Task (Eriksen and Eriksen, 1974) was adopted. A five-letter string stimuli with a central target letter (M/N) and four-remaining flanker letters (N/M) were used. The letters M and N were more similar with increased complexity in comparison to the original version with the letters H and S (Gulbinaite et al., 2014). In congruent trials (MMMMM or NNNNN), the target letter (the middle letter in the five-letter string) was identical to the flankers, whereas in incongruent trials (MMNMM or NNMNN), the target letter differed from the flankers. The participants were instructed to press the left button with the left index finger if the target was M and the right button with the right index finger if the target was N as soon as possible while maintaining a high level of accuracy.

All stimuli were presented as white against a black background on a computer screen. At the beginning of the task, there was a fixation cross in the center of the screen (0.32° × 0.32°). Each letter of the string had a height and width of 0.24° visual angle. The letters were 0.05° apart to increase the error rates (Boksem et al., 2008). After 1,000 ms, the fixation cross was replaced by the five-letter string. The stimuli disappeared after 200 ms and—for the subjects to provide the response—were followed by a time interval, which elapsed until the response button was pressed or until 600 ms. An additional 200-ms interval was provided for the subjects to realize a possible erroneous response. Finally, the feedback indicating task performance was presented for 1,000–1,500 ms, depending on the response time. Feedback was presented with given responses ("Correct," "Error," or "Miss") at a width of 0.5 cm. Each trial lasted 3 s in total. The trial structures are depicted in **Figure 1**. Congruent (60%) and incongruent (40%) trials were presented in random order (Tops et al., 2006).

Reward

Although individuals present differences in sensitivity to reward, the monetary reward has been corroborated to be an effective means of motivation manipulation (Paschke et al., 2015). Participants were told that in one or some blocks, for each correct response, they would receive bonus money, and they would not lose money for errors or misses. Participants could earn up to

FIGURE 1 | The procedure of experiment design and the trial structure of flanker task.
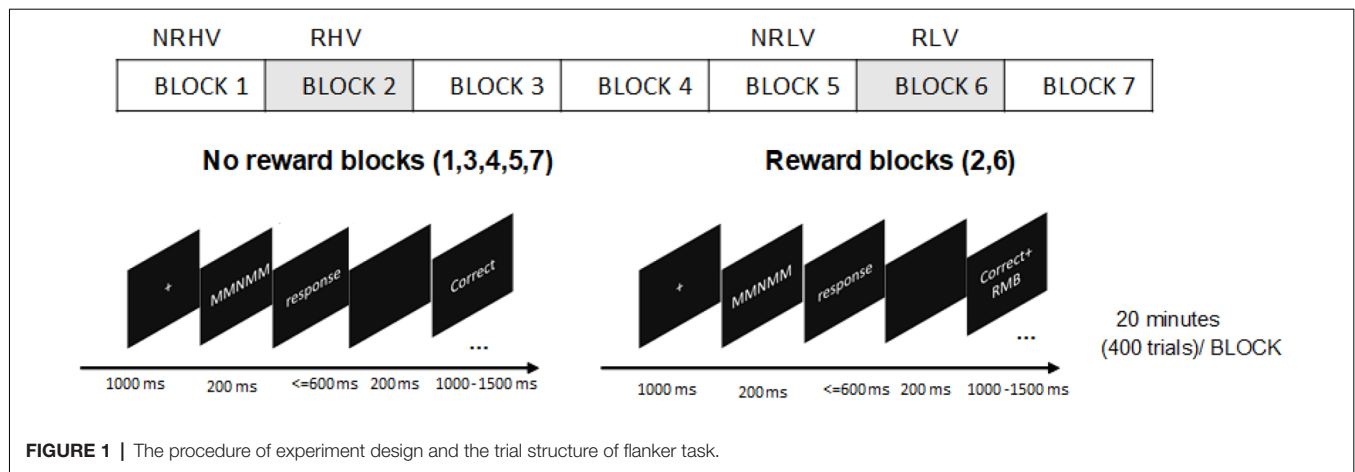
100 RMB (approximately 12.8€) in addition to a basic sum of 50 RMB (approximately 6.4€) payment. The amount of money was evaluated proportionally to students' monthly expenses when manipulating motivation. To maintain the effectiveness of the reward, it was stressed that they would receive the bonus if the average accuracy of the reward blocks was more than 90%; otherwise, they would lose it. For the feedback in the reward blocks, the correct responses "Correct" coupled with "+ RMB" were 1 cm in width, and the "Error" or "Miss" responses were similar to the nonreward blocks (**Figure 1**).

Procedure

The participants were informed that they should abstain from alcohol, tea, and coffee for 24 h before the experiment. After arriving at the laboratory, they were given the written task instructions. They were asked to leave their watches and mobile phones outside the laboratory so that they had no indication of time during the measurement. The participants were then seated in front of a 19-inch PC monitor (1,280 × 1,024 pixels) at a distance of 0.9 m in a dimly lit, sound-attenuated, and electrically shielded room. Participants practiced the task before the formal experiment day to achieve an accuracy of 90% (those with an accuracy of <90% were not included in this study). Moreover, the reward was introduced in the practice experiment to build the association between task performance and monetary reward already prior to the experiment to avoid different time of reward exposure in high vs. low vigilance states in the formal experiment. On the experiment day, prior to the start of the formal experiment, participants performed the task for 10 min (200 trials) to adapt to the task. In the formal experiment, they were instructed to respond to the target letter presented in seven blocks of 20 min, for a total of 2 h 20 min (2,800 trials). Among the seven blocks, the monetary reward was introduced in blocks 2 and 6. The procedures can be found in **Figure 1**. The task blocks 1–4 were performed to induce vigilance decrement. To avoid any anticipatory effect of experiment ending, the additional no-reward block 7 was performed after the rewarded block 6. There was no rest during the experiment or any subjective questionnaires to maintain task performance and avoid the

effects of short breaks alleviating fatigue. Prior studies have shown that even short breaks can increase task performance, making it difficult to evaluate whether the performance recovery results from motivation or the short break (Helton and Russell, 2015; Lim and Kwok, 2016). To maintain task performance, subjects were asked to focus their attention on the target letter presented in the center of the screen. The subjects were informed of the beginning and end of the reward blocks by instructions displayed on the screen. At the end of the task, the average accuracy of reward blocks was calculated to determine whether they would receive the bonus money or not.

EEG Recording and Processing

The EEGs were recorded using 64 Ag/AgCl electrodes attached to an electro cap according to the International 10-20 System. An ANT Neuro EEG amplifier was used to record EEG signals sampled at a digitization rate of 500 Hz. Horizontal and vertical electrooculograms were recorded from the outer canthi of the eyes and above and below the left eye. The electrode impedance was kept below 10 kΩ, and the EEG was online referenced to the CPz channel.

In the offline analysis, EEG data were notch filtered at 50 Hz. Next, a digital high-pass filter of 0.5 Hz and a low-pass filter of 30 Hz were applied. After removing the direct current (DC) component, the EEG signals were denoised using the wavelet threshold method (Zhang et al., 2018), wherein the wavelet coefficient threshold was set to *abs* (mean ± 3 × SD). If the absolute value of the wavelet coefficients exceeded the threshold, the coefficients were reset to one-quarter of the average value. The data were re-referenced to the average of the mastoid references (M1, M2). The ERP epochs from 200 ms before to 800 ms after stimulus onset were extracted. Finally, by using the Icasso software (Himberg and Hyvärinen, 2003), independent artifact components (e.g., blinks, movements, etc.) were removed through visual inspection.

Data Analysis

To study the effects of the reward state (i.e., no-reward vs. reward) on the behavioral and electrophysiological measures

in the states of high vs. low vigilance, four blocks (blocks 1, 2, 5, and 6) were selected. The subjects were provided with monetary rewards in blocks 2 and 6. In both high (blocks 1 and 2) and low (blocks 5 and 6) vigilance states, the reward blocks were introduced after the no-reward blocks. In summary, the analysis was based on 2 × 2 comparisons, representing the no-reward high vigilance (NRHV) condition in block 1, reward high vigilance (RHV) condition in block 2, no-reward low vigilance (NRLV) condition in block 5, and reward low vigilance (RLV) condition in block 6.

Behavioral Performance
For each participant, the accuracy, mean RT, and number of omitted responses were calculated. Only responses occurring between 100 and 600 ms were included in the RT analysis. A response time equal to zero was regarded as an omitted response. The accuracy was calculated as the percentage of correct responses in each block. We addressed the main effects and interactions of the vigilance state and the reward state on task performance. In addition, the effect of congruency (congruent vs. incongruent) was also tested for accuracy, RT, and omitted responses.

Event-Related Potentials
ERPs were analyzed with MATLAB 2015b. First, the individual correct trials whose amplitude was out of range (max >75 μv, baseline max >30 μv) were rejected, and then the baseline 200 ms before stimulus onset was subtracted from the waveforms. Next, trials were averaged across blocks for each subject. The mean (with SD in parentheses) number of trials across all subjects for NRHV, RHV, NRLV, and RLV were 236 (82), 232 (65), 234 (64), and 238 (64), respectively. The P300 amplitude and latency were quantified for further analysis. Based on some earlier studies (Polich and Kok, 1995; Kuba et al., 2012; van Dinteren et al., 2014) and topographic activations in our study, eight electrodes (FC1, FC2, FCz, C1, C2, Cz, CP1, and CP2) were chosen for the P300 analysis. A time window of 440–660 ms for the P300 component was selected. The P300 latency values were calculated as the time of maximum amplitude within the time window of the P300 component (Luck, 2005).

EEG Spectra
The EEG spectral power was assessed by calculating the ERSP using the continuous wavelet transform (CWT; Zhang et al., 2018). The complex Morlet wavelet was adopted for the CWT analysis, by which the time-dependent signals were evaluated at each sampling instant with a central frequency band of 1.5 Hz covering frequencies from 1 to 30 Hz, with a frequency step of 0.5 Hz. Additionally, we normalized the power spectra with the subtraction change from −1,000- to 0-ms baseline. For quantifying the oscillatory dynamics, we focused on separate time windows in the analysis of two frequency bands (**Figure 5**). According to the maximum power of the different frequency bands, statistical analysis was performed within the time window of 440–660 ms for the delta band (1–4 Hz) and within the time window of 300–600 ms for the theta band (4–8 Hz). In order to account for the effect of phase-locked (evoked response)

activity in the induced oscillations, we also analyzed the induced activations by subtracting the averaged evoked response from each epoch prior to the wavelet analysis. The results of this analysis are provided in the **Supplementary Materials**.

Statistical Analysis
Data were analyzed using the IBM SPSS software (version 22.0), Chicago: SPSS Inc. The significance level $p < 0.05$ was used, and all results were reported under the 2-tailed condition. One-way repeated-measures analysis of variance (ANOVA) with the blocks 1, 3, 4, and 5 was used to test the hypothesis that behavioral performance deteriorates with time-on-task. Blocks 2 and 6 with an additional influence of motivation and block 7 with an effect of approaching the end of the task were excluded to capture the changes purely due to time-on-task. Moreover, behavioral, time domain, and time-frequency domain data were subjected to 2 × 2 [vigilance states (high and low) × reward states (no-reward and reward)] repeated-measures ANOVA. In case of significant interaction and/or main effects, a follow-up ANOVA was applied to separately test the effect of the vigilance state in no-reward and reward conditions (NRLV vs. NRHV indicates the effects of vigilance decrement) and the effect of reward in low and high vigilance states (RHV vs. NRHV and RLV vs. NRLV indicate the effects of motivation in the high and low vigilance states, respectively). The Greenhouse–Geisser correction was used as the adjusted report, and the effect size was determined using adjusted partial η^2 (η^2_{ap}; Mordkoff, 2019).

The effect of congruency was initially tested with 2 × 2 × 2 ANOVA (congruency, vigilance state, and reward state). However, as no interaction was found for congruency, the effects of the reward and vigilance states were tested with congruent and incongruent trials integrated together. The correlations between performance (accuracy, RT, and omitted response) and ERPs (P300 amplitude and latency) were calculated using the Pearson Correlation Coefficient to study the association between the behavioral and electrophysiological measures in different vigilance and reward states.

RESULTS

Behavioral Performance
Figure 2 illustrates the alterations of behavioral performance (accuracy, RT, and number of omitted responses) with time-on-task. Based on the one-way repeated-measures ANOVA, we found that the accuracy significantly decreased ($F_{(1.37,25.93)} = 4.44$, $p = 0.02$, $\eta^2_{ap} = 0.15$) with time-on-task. Meanwhile, the RT ($F_{(2.37,44.99)} = 3.97$, $p = 0.03$, $\eta^2_{ap} = 0.13$) and the number of omitted responses ($F_{(2.55,48.45)} = 4.12$, $p = 0.02$, $\eta^2_{ap} = 0.14$) significantly increased along with the prolonged task performance.

Accuracy
In the 2 (vigilance states) × 2 (reward states) ANOVA analysis, there was a significant main effect of the reward state ($F_{(1,19)} = 6.02$, $p = 0.03$, $\eta^2_{ap} = 0.21$) and a significant vigilance state × reward state interaction ($F_{(1,19)} = 7.38$, $p = 0.01$, $\eta^2_{ap} = 0.24$). When the vigilance states were contrasted separately for reward and no-reward conditions, the accuracy was lower
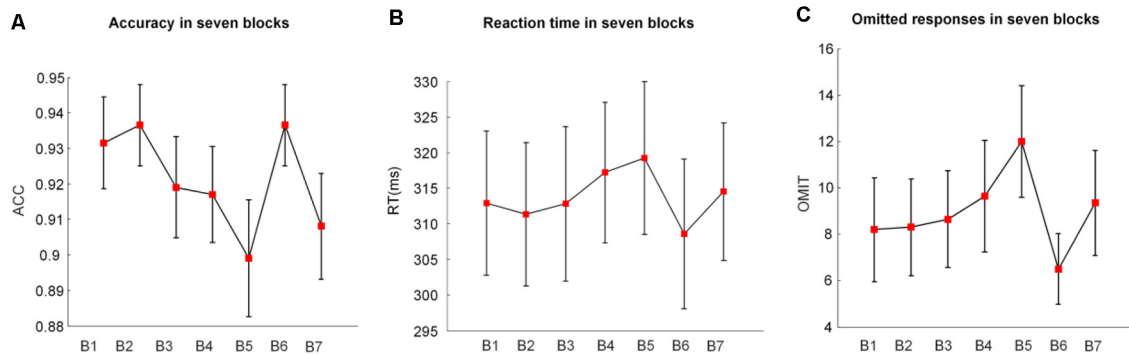
FIGURE 2 | The mean values and standard deviations (SDs)/sqrt (subjects) of **(A)** the accuracy (ACC), **(B)** the reaction time (RT), and **(C)** the omitted responses (OMIT) in seven blocks. B indicates block (e.g., B1 is block1); B2 and B6 were introduced with monetary rewards.
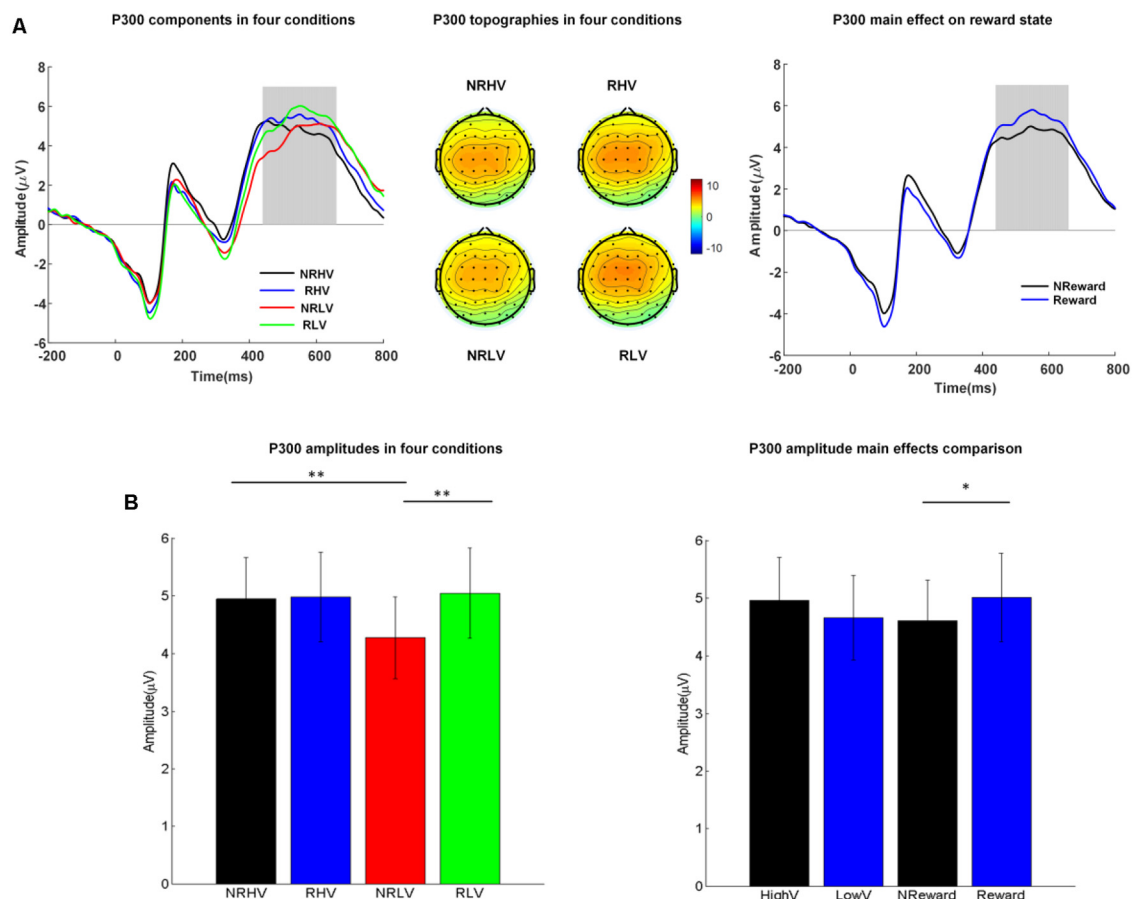
FIGURE 3 | (A) The P300 waveforms (left) averaged from electrodes of FC1, FC2, FCz, C1, C2, Cz, CP1, and CP2, the topographies (middle) in the four conditions, and P300 waveforms coalesced for the reward vs. non-reward states (right). **(B)** Mean values and standard error of the P300 amplitude in the four conditions (left) and in the two main factors (right). HighV (high vigilance blocks) = (NRHV + RHV)/2, LowV (low vigilance blocks) = (NRLV + RLV)/2, NReward (no-reward blocks) = (NRHV + NRLV)/2, and reward (reward blocks) = (RHV + RLV)/2. Analysis of variance (ANOVA) results were marked by *$p < 0.05$ and **$p < 0.01$.

in the low vigilance state than in the high vigilance state in the no-reward condition (NRLV: mean = 0.88, SD = 0.12, NRHV: mean = 0.93, SD = 0.06, $F_{(1,19)}$ = 5.24, p = 0.03, η_{ap}^2 = 0.17).

There was no difference between the rewarded low and high vigilance states (RLV: mean = 0.94, SD = 0.05, RHV: mean = 0.94, SD = 0.05, $F_{(1,19)}$ = 0.00, p = 1.00, η_{ap}^2 = 0.00). The monetary
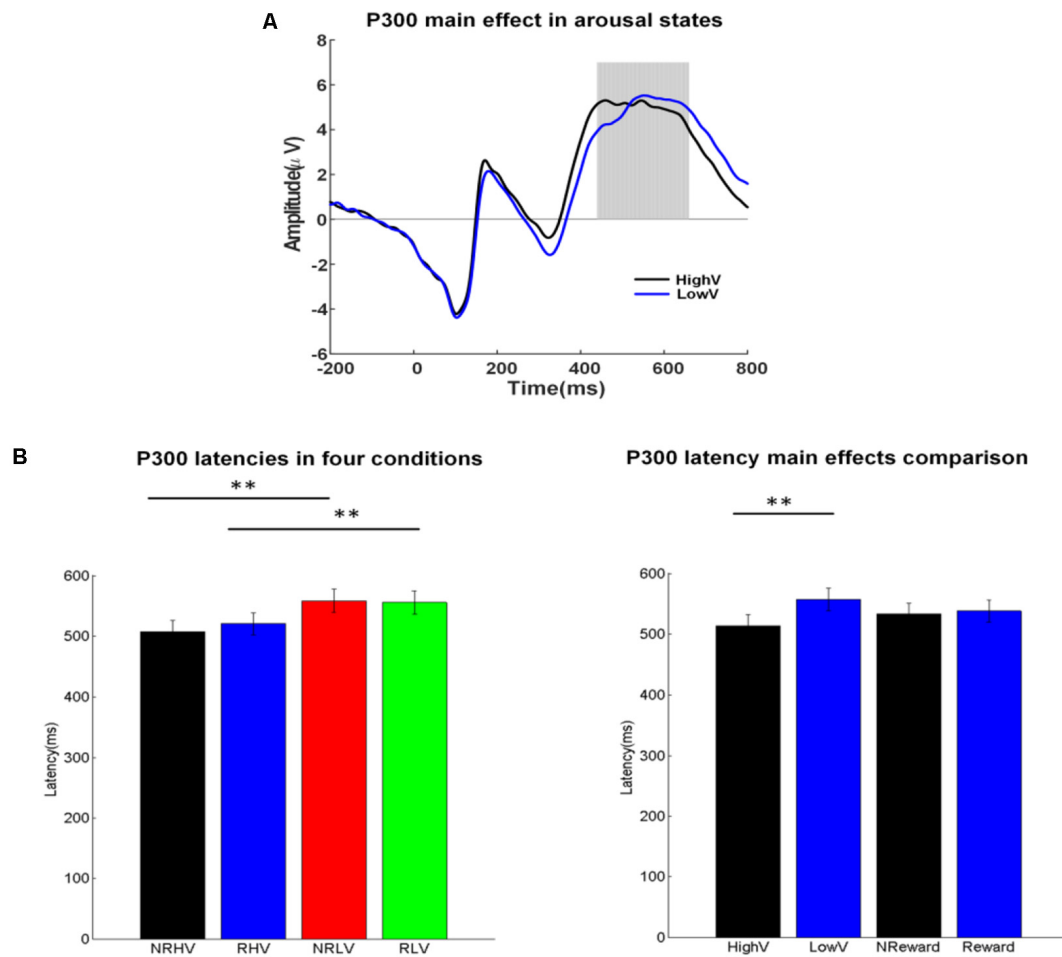
FIGURE 4 | **(A)** P300 waveforms coalesced for the high vs. low vigilance states. **(B)** Mean values and standard error for the P300 latency in the four conditions (left) and for the two main factors (right). The ANOVA results were marked by *$p < 0.05$ and **$p < 0.01$.

reward played a role only in the low vigilance state. The accuracy was higher in the rewarded than in the no-rewarded low vigilance condition ($F_{(1,19)} = 7.37$, $p = 0.01$, $\eta^2_{ap} = 0.24$), although there was no significant difference between the rewarded and the no-rewarded high vigilance conditions ($F_{(1,19)} = 0.47$, $p = 0.50$, $\eta^2_{ap} = -0.03$).

Reaction Time

There was a significant main effect of the reward state on the RT ($F_{(1,19)} = 10.95$, $p < 0.01$, $\eta^2_{ap} = 0.33$). The follow-up ANOVA indicated that the RT increased with vigilance decrement in the no-reward condition (NRLV: mean = 319.22, $SD = 49.15$, NRHV: mean = 311.02, $SD = 46.42$, $F_{(1,19)} = 5.52$, $p = 0.03$, $\eta^2_{ap} = 0.18$). There was no significant difference between the low and high vigilance states in the reward condition (RLV: mean = 308.58, $SD = 45.25$, RHV: mean = 309.83, $SD = 47.25$, $F_{(1,19)} = 0.21$, $p = 0.65$, $\eta^2_{ap} = -0.04$). When rewards were provided in the states of low and high vigilance, the RT was faster in the low vigilance state ($F_{(1,19)} = 8.38$, $p = 0.01$, $\eta^2_{ap} = 0.27$) but was not improved in the high vigilance state ($F_{(1,19)} = 1.75$, $p = 0.20$, $\eta^2_{ap} = 0.04$).

Omitted Responses

There was a significant main effect of the reward state on the number of omitted responses ($F_{(1,19)} = 9.22$, $p = 0.01$, $\eta^2 = 0.29$). The follow-up ANOVA revealed that the omitted responses increased with the decrease in vigilance in the no-reward condition ($F_{(1,19)} = 5.39$, $p = 0.03$, $\eta^2_{ap} = 0.18$). No difference was found between low and high vigilance states in the reward condition ($F_{(1,19)} = 0.07$, $p = 0.80$, $\eta^2_{ap} = -0.05$). The number of omitted responses decreased in the state of low vigilance ($F_{(1,19)} = 10.94$, $p = 0.01$, $\eta^2_{ap} = 0.33$) after motivation manipulation, although it did not change in the state of high vigilance ($F_{(1,19)} = 1.97$, $p = 0.18$, $\eta^2_{ap} = 0.05$).

Congruency

Regarding the congruency (congruent × incongruent), we found significant main effects of congruency on accuracy ($F_{(1,19)} = 18.07$, $p < 0.01$, $\eta^2_{ap} = 0.46$), RT ($F_{(1,19)} = 32.75$, $p < 0.01$, $\eta^2_{ap} = 0.61$), and omitted responses ($F_{(1,19)} = 9.65$, $p = 0.01$, $\eta^2_{ap} = 0.30$). The congruent condition showed a higher accuracy (congruent: mean = 0.94, $SD = 0.01$, incongruent: mean = 0.91,
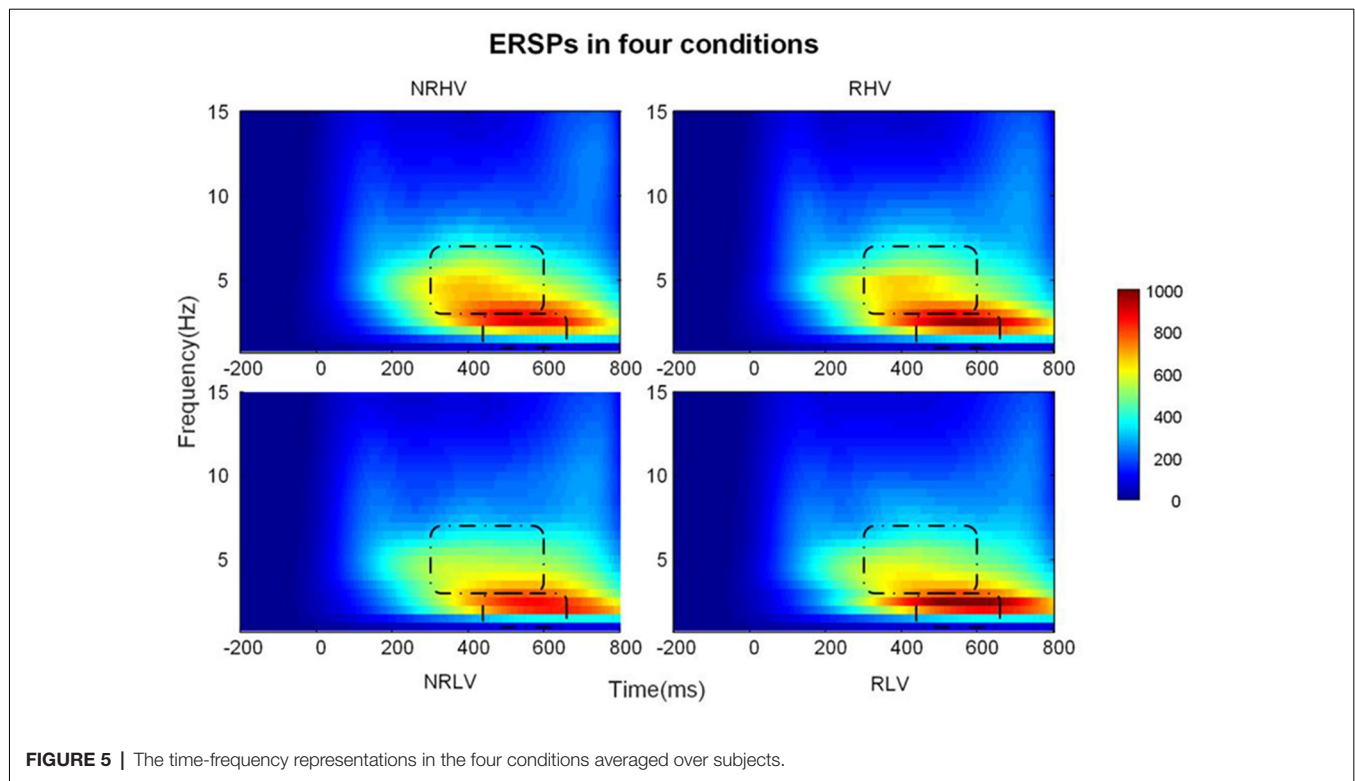
FIGURE 5 | The time-frequency representations in the four conditions averaged over subjects.

$SD = 0.02$), faster RTs (congruent: mean = 306.93, $SD = 10.08$, incongruent: mean = 315.28, $SD = 9.84$), and less omitted responses (congruent: mean = 2.98, $SD = 0.69$, incongruent: mean = 5.45, $SD = 1.31$) than the incongruent condition. However, no significant interaction between congruency and vigilance state or between congruency and reward state was observed in behavioral performance.

Event-Related Potentials

P300 Components

Amplitude

The left part of **Figure 3A** shows the averaged ERP amplitude waveforms with the time window of interest (P300 response at 440–660 ms after stimulus onset) depicted by a gray rectangle. The middle part of **Figure 3A** shows the corresponding topographies in the four experimental conditions, whereas **Figure 3B** illustrates the differences in P300 amplitude between the four conditions (left) and between the two main factors of vigilance state and reward state (right).

The repeated-measures ANOVA showed a significant main effect of the reward state on the P300 amplitude ($F_{(1,19)} = 7.08$, $p = 0.02$, $\eta_{ap}^2 = 0.23$) and a significant interaction between the vigilance state and reward state ($F_{(1,19)} = 6.78$, $p = 0.02$, $\eta_{ap}^2 = 0.22$). The follow-up ANOVA revealed that the P300 amplitude decreased with vigilance decrement (NRLV: mean = 4.17, $SD = 2.68$, NRHV: mean = 4.81, $SD = 2.74$, $F_{(1,19)} = 9.99$, $p = 0.01$, $\eta_{ap}^2 = 0.31$) in the no-reward condition, although no significant difference was found between the low and high vigilance states in the reward condition (RLV: mean = 4.96,

$SD = 3.01$, RHV: mean = 4.89, $SD = 2.98$, $F_{(1,19)} = 0.07$, $p = 0.80$, $\eta_{ap}^2 = -0.05$). When the effect of reward was tested in states of low and high vigilance separately, the reward improvement presented only in the low vigilance state ($F_{(1,19)} = 15.88$, $p < 0.01$, $\eta_{ap}^2 = 0.43$) and not in the high vigilance state ($F_{(1,19)} = 0.12$, $p = 0.74$, $\eta_{ap}^2 = -0.05$).

Regarding the congruency (congruent vs. incongruent), a significant main effect of congruency was found for the P300 amplitude ($F_{(1,19)} = 22.19$, $p < 0.01$, $\eta_{ap}^2 = 0.51$). The amplitude was higher in the congruent condition (mean = 5.07, $SD = 1.15$) than in the incongruent condition (mean = 4.35, $SD = 1.14$). No interaction from congruency × vigilance state or congruency × reward state was detected.

Latency

Figure 4A illustrates the ERP waveforms in high and low vigilance states (reward and nonreward blocks coalesced), and **Figure 4B** shows the differences in P300 latency in the four experimental conditions (left) and the two main factors of vigilance state and reward state (right).

There was a significant main effect of the vigilance state on the P300 latency ($F_{(1,19)} = 52.20$, $p < 0.01$, $\eta_{ap}^2 = 0.72$) and an interaction between the vigilance state and the reward state ($F_{(1,19)} = 6.55$, $p = 0.02$, $\eta_{ap}^2 = 0.22$). Separate ANOVAs revealed the clear effects of vigilance states regardless of rewards. The P300 latency increased in the low vigilance state compared with the high vigilance state in both the no-reward (NRLV: mean = 557.45, $SD = 84.71$, NRHV: mean = 506.85, $SD = 81.31$, $F_{(1,19)} = 45.52$, $p < 0.01$, $\eta_{ap}^2 = 0.69$) and reward conditions (RLV: mean = 554.65, $SD = 83.94$, RHV: mean = 519.55, $SD = 80.44$,

$F_{(1,19)} = 37.97$, $p < 0.01$, $\eta^2_{ap} = 0.65$). When the effect of reward was tested in the states of low and high vigilance, there was no difference between reward and no-reward conditions in the low vigilance state ($F_{(1,19)} = 0.37$, $p = 0.55$, $\eta^2_{ap} = -0.03$), but we did find a decrease in the high vigilance state ($F_{(1,19)} = 6.81$, $p = 0.02$, $\eta^2_{ap} = 0.23$) after manipulating motivation.

Regarding congruency (congruent vs. incongruent), a significant main effect was found on the P300 latency ($F_{(1,19)} = 8.91$, $p = 0.01$, $\eta^2_{ap} = 0.28$). The latency was shorter in the congruent (mean = 465.59, SD = 8.43) than in the incongruent conditions (mean = 471.48, SD = 9.41). For P300 latency, no interaction from congruency × vigilance state or from congruency × reward state was detected.

Correlations Between ERPs and Behavioral Performance

To investigate the associations between task performance and ERPs affected by motivation and vigilance states, the correlations between the behavioral measures (accuracy, RT, and number of omitted responses) and ERPs (the amplitude and latency of P300) were calculated (**Table 1**). Significant negative correlations between the accuracy and P300 latency and significant positive correlations between the accuracy and P300 amplitude were detected. Additionally, the number of omitted responses and the RT were negatively correlated with the P300 amplitude and positively correlated with the P300 latency. Scatter diagrams showing the relationships between the behavioral measures and P300 amplitude and latency can be found in **Supplementary Figure S1**.

ERSP Analysis

Figure 5 illustrates the time-frequency representations (averaged over electrodes F1, F2, Fz, FC1, FC2, FCz, C1, C2, Cz, CP1, and CP2) in the four experimental conditions. A clear modulation of frequencies of approximately 1–4 Hz is visible in the time window of 440–660 ms. Separable modulations of approximately 4–8 Hz (in the time window of 300–500 ms) appear visually earlier than 1–4 Hz over the four conditions. The corresponding frequency bands and time windows are indicated by the dotted-line boxes. We also calculated the induced time-frequency representations after removing the phase-locked evoked responses from the total power (**Supplementary**

Figure S2). **Figure 6A** illustrates the topographic distribution (right) and power waveforms (left) averaged across the electrodes (referred above) corresponding to the delta band (averaged over 1–4 Hz). **Figure 6B** draws the topographic distribution (right) and power waveforms (left) of the theta band (averaged over 4–8 Hz), with activations in the frontal electrodes (F1, F2, Fz, FC1, FC2, and FCz).

In the 2 (vigilance states) × 2 (reward states) ANOVAs, for the delta band power, we found a significant interaction between the vigilance state and reward state ($F_{(1,19)} = 7.28$, $p = 0.01$, $\eta^2_{ap} = 0.24$). When the effect of the vigilance state was tested in the no-reward and reward conditions, the delta band power decreased with vigilance decrement in the no-reward condition (NRLV: mean = 703.43, SD = 162.89, NRHV: mean = 768.16, SD = 153.49, $F_{(1,19)} = 8.72$, $p = 0.01$, $\eta^2_{ap} = 0.28$), but no significant difference was detected between low and high vigilance states in the reward condition (RLV: mean = 750.86, SD = 206.30, RHV: mean = 757.08, SD = 200.46, $F_{(1,19)} = 0.05$, $p = 0.82$, $\eta^2_{ap} = -0.05$). When the effect of reward was separately tested in the states of low and high vigilance, the effect of reward on delta band power was detected only in the low vigilance state ($F_{(1,19)} = 4.57$, $p = 0.04$, $\eta^2_{ap} = 0.19$) and not in the high vigilance state ($F_{(1,19)} = 0.23$, $p = 0.64$, $\eta^2_{ap} = 0.01$).

For the theta band, vigilance state had a significant main effect ($F_{(1,19)} = 18.56$, $p < 0.01$, $\eta^2_{ap} = 0.47$). Follow-up ANOVA revealed that theta power was weaker in low vigilance state than in high vigilance state in both no-reward condition (NRLV: mean = 432.43, SD = 115.29, NRHV: mean = 512.47, SD = 169.54, $F_{(1,19)} = 11.38$, $p < 0.01$, $\eta^2_{ap} = 0.34$) and reward condition (RLV: mean = 455.70, SD = 118.69, RHV: mean = 498.61, SD = 137.47, $F_{(1,19)} = 6.36$, $p = 0.02$, $\eta^2_{ap} = 0.21$). The separate ANOVAs revealed that reward did not play a role in low vigilance state ($F_{(1,19)} = 1.73$, $p = 0.20$, $\eta^2_{ap} = 0.03$) or in high vigilance state ($F_{(1,19)} = 0.21$, $p = 0.65$, $\eta^2_{ap} = -0.04$).

DISCUSSION

We examined the alterations in behavioral performance and brain electrophysiology produced by the vigilance level and reward during a prolonged period of selective visual attention tasks. Behavioral measures (accuracy, RT, and number of omitted responses), evoked responses (P300 amplitude and latency), and spectral power (delta and theta bands) were analyzed. A clear deterioration in behavioral performance was demonstrated over time (**Figure 2**). The monetary reward improved the performance in accuracy, RT, and number of omitted responses only in the low vigilance state. The P300 amplitude was smaller in low than in high vigilance state; however, in the low vigilance state, reward increased the P300 amplitude to its level in the high vigilance state. The P300 latency was sensitive to vigilance decrement but insensitive to rewards, with longer latency in low than in high vigilance states. Changes in spectral power at 4–8 Hz purely reflected the vigilance level, being stronger in the high vigilance state than in the low vigilance state. Similarly, the spectral responses at 1–4 Hz also decreased with vigilance decrement. However, the reward selectively increased the spectral power at 1–4 Hz

TABLE 1 | Correlations between behavioral performance and P300 measures in the four conditions.

	NRHV	RHV	NRLV	RLV
Model	r	r	r	r
ACC vs. AMP	0.40	0.57*	0.60*	0.59*
ACC vs. LAN	−0.52*	−0.53*	−0.53*	−0.41
RT vs. AMP	−0.65**	−0.68**	−0.82**	−0.83**
RT vs. LAN	0.54*	0.51*	0.53*	0.64*
OMIT vs. AMP	−0.39	−0.45*	−0.65**	−0.65**
OMIT vs. LAN	0.54*	0.46*	0.71**	0.59*

Note: The AMP and LAN represent the amplitude and latency of P300. ACC and OMIT represent the accuracy and number of omitted responses. Correlation cofficients in 2-tailed condition were marked by *p < 0.05 and **p < 0.01. The correlations were corrected by executing the Benjamini and Hochberg procedure for controlling the false discovery rate (FDR; Benjamini and Hochberg, 1995).

FIGURE 6 | The temporal waveforms of power modulation and topographies of the delta **(A)** and the theta band **(B)** in the four experimental conditions.
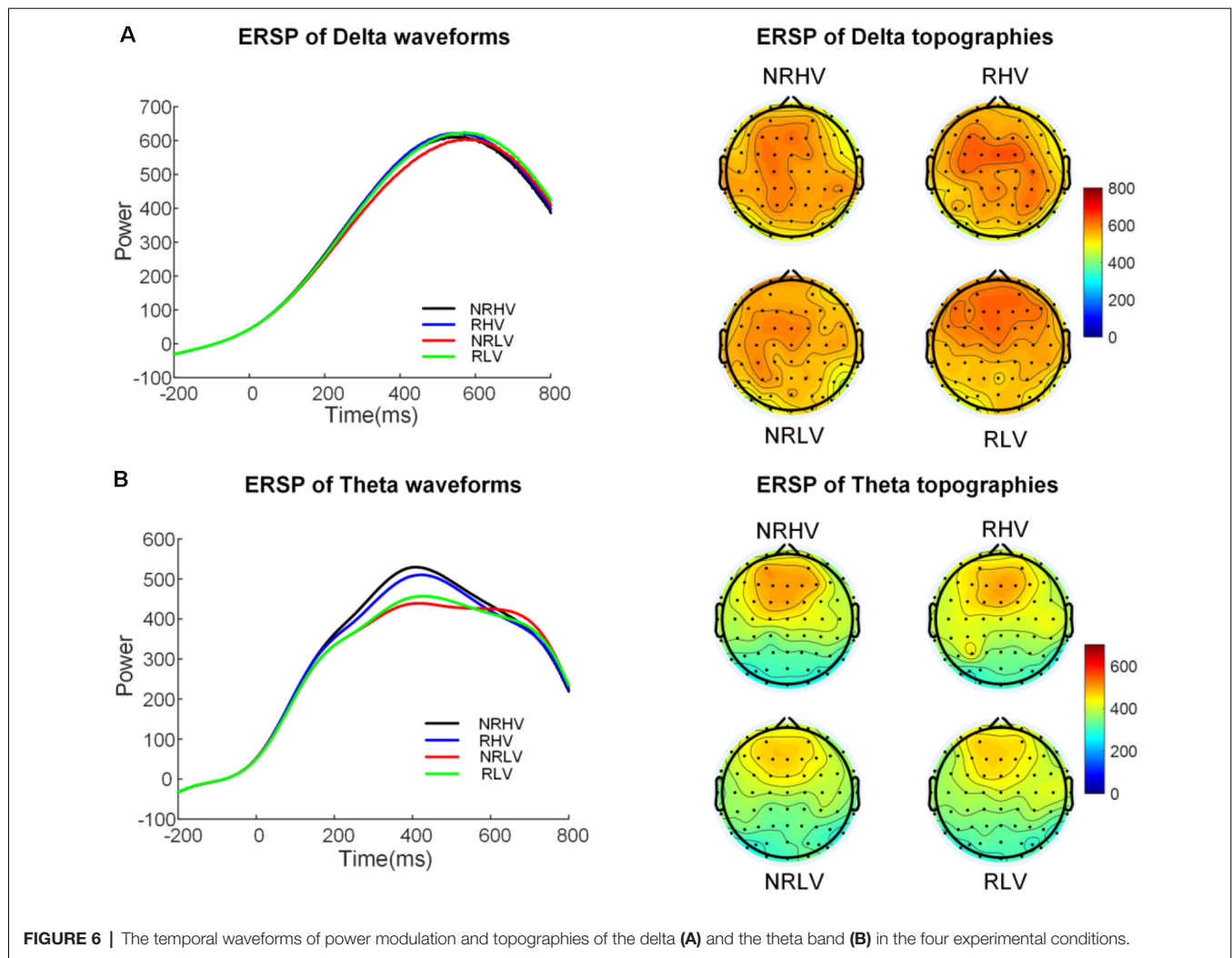
in the low vigilance state to the strength levels in the high vigilance state.

Although the time of emergence of fatigue during the prolonged performance of cognitive tasks has not been defined, earlier studies suggest cumulative effects in performance and neurophysiology with time-on-task (Boksem et al., 2005, 2006; Lorist et al., 2005; Faber et al., 2012; Möckel et al., 2015; Reteig et al., 2019). In line with these findings, our study found a significant decrease in accuracy and an increase in RT and the number of omitted responses with time-on-task (**Figure 2**). The decline of behavioral performance in prolonged attention tasks is in line with our assumption that time-on-task is associated with the decrement of vigilance levels. These results provided justification for testing the interactions between vigilance and reward states, where blocks 1 and 2 (first 40 min) were regarded as representing the high vigilance state, whereas blocks 5 (80–100 min) and 6 (100–120 min) were regarded as the low vigilance state. This selection was also supported by earlier findings of the effects of mental fatigue after 60- to 90-min tasks (Kastner and Ungerleider, 2000; Lorist et al., 2000; Marcora et al., 2009).

The limited processing capacity biased toward goal-directed selection is the core of selective visual attention (Robert and Duncan, 1995; Polich, 2009). The P300 component is considered as an important indicator of attentional capacity in visual tasks (Polich, 2009). In line with earlier results showing a fatigue-related decrease in P300 amplitude during brain-computer interface performance (Käthner et al., 2014), our results demonstrated that the P300 amplitude decreased with vigilance decrement during a selective visual attention task, presumably reflecting a less efficient engagement or limited capacity of attentional resources. The insufficient attention resources allocation in the low vigilance state has also been reflected by the P300 latency, which is thought to provide a specific index for the timing of information processing and stimulus evaluation (Polich and Kok, 1995; Verleger, 1997; Käthner et al., 2014). In our study, the P300 latency was significantly prolonged in the state of low vigilance, in line with earlier studies (Kutas et al., 1977; Boksem et al., 2006; Kato et al., 2009). The result indicates that prolonged task performance accompanies longer evaluation time for processing information. Therefore, in agreement with existing studies, our

results demonstrated a decrease in P300 amplitude and an increase in latency along with vigilance decrement.

The close link between the modulations in behavioral performance and the changes in brain electrophysiology is demonstrated by strong correlations between the behavioral and P300 measures (**Table 1**). It is noteworthy that, although the RT correlates with both P300 amplitude and latency, the association between the decrease of P300 amplitude and the increase in RT in particular is clear (**Supplementary Figure S1**).

After reward manipulation, the P300 amplitude increased only in the low vigilance state, reaching the same level as in the high vigilance state. Our results are consistent with earlier studies demonstrating a monetary-reward based improvement of neural measures in mental fatigue (Boksem et al., 2006; Hopstaken et al., 2015). We further provide quantitative evidence for the recovery of attentional resources from low vigilance to high vigilance states. Our results verify that, when rewards are provided, the capacity of attentional resources in the low vigilance state can reach the level of capacity in the high vigilance state, at least within the limited time duration of 2 h 20 min for a visual attention task.

Interestingly, there was no significant reward-induced improvement in P300 latency either in high or low vigilance states despite the reward-related improvement in RT in the low vigilance state. These results are in line with a previous study (Boksem et al., 2006), suggesting that the P300 latency is an unstable electrophysiological marker of motivation compared with the P300 amplitude. These diverging results concerning the RT and P300 latency could be interpreted to indicate improvement in motor response generation (as reflected by the RT) but not in the preceding stage of information processing (as reflected by the P300 latency). However, these results might also reflect the complex composition (subcomponents) of the P300 responses. The P300 component has been shown to comprise two subcomponents—P3a and P3b—with different functional correlations (Demiralp et al., 2001). The P3a with frontal topography has been suggested to contribute to attention engagement in top-down task-relevant processing, whereas the P3b with centroparietal topography has been linked to the level of cognitive workload and memory encoding (Polich, 2009). They are activated in different time windows, and P3a usually emerges earlier than P3b (Polich, 2009). Although we fail to disentangle the two subcomponents in this study, it is possible that, in addition to changes in the amplitude, also changes in the emphasis of these neural subprocesses are associated with vigilance decrement. This complicates the interpretation of latency measures.

The topography of the P300 component in the present study is more anterior than that in some earlier studies (Demiralp et al., 2001; Käthner et al., 2014). This likely reflects the task requirements of the present study. To successfully perform a selective visual attention task, humans are able to filter out task-irrelevant stimuli and engage their limited capacity in task-relevant processing (Robert and Duncan, 1995). Our task is likely to harness—although it was not designed to differentiate—these two subprocesses. The modified Eriksen Flanker Task applied in our study required responses for every trial and was specifically adapted to make the target letter distinction visually hard, emphasizing the need for the active inhibition of the flankers. Some interpretations, especially in studies showing anteriorly located P300 generators, emphasize the role of the inhibitory control underlying the modulation of P300 responses, for example, as a result of aging (Kuba et al., 2012; van Dinteren et al., 2014). Importantly for the present findings, this interpretation is also sensible in the context of vigilance decrement, which is often accompanied by a reduction in the capacity for top-down inhibitory control (Guo et al., 2018).

P300 seems to provide a reliable measure of cognitive performance, but the analysis of phase-locked ERP offers only limited windows to explore the underlying neural processes in more detail. Previous studies have indicated that both delta (Keller et al., 2017) and frontal theta oscillations (Knyazev, 2012) are involved in visual attention tasks, although the influence of time-on-task was not studied in these studies. Focusing on the spectral patterns and the power modulations at different frequency bands may provide additional sensitivity to separate vigilance- and reward-related processes in the brain. The oscillatory activity at low-frequency bands [i.e., the delta (1–4 Hz) and theta (4–8 Hz) bands] has been shown to increase during the transition to the low vigilance state in spontaneous conditions (Lal and Craig, 2001). Although the modulations of the oscillatory activity triggered by cognitive tasks are different from those from spontaneous activity, the temporal variations in the power of these frequency bands triggered by a visual attention task may tap on the same underlying processes as reported based on more spontaneous conditions.

In the present study, the changes in the spectral power at 1–4 Hz and 4–8 Hz reflected different topographies, with the delta band distributed in the centroparietal electrodes and the theta band distributed more focally in the frontal electrodes. In addition, the temporal characteristics of the changes in power in these two frequency bands differed, with an earlier emergence of the modulation at 1–4 Hz (300–600 ms) than at 4–8 Hz (440–660 ms). Therefore, it is likely that the two separable changes in spectral power reflect two distinct cognitive functions involved in selective visual attention. The theta band has been shown to be an indicator of attention allocation to task-relevant stimuli (Keller et al., 2017), whereas the delta band has been linked with internal processing in an attention task (Harmony et al., 1996). Prolonged engagement in a selective visual attention task led to the reduced spectral power in both delta and theta bands. These results are in line with the analysis of the evoked P300 responses and suggest that vigilance decrement impairs both attention allocation and information processing. In the rewarded low vigilance condition, the spectral power at approximately 1–4 Hz increased to the same level as that recorded in the high vigilance state. However, the spectral power at approximately 4–8 Hz did not increase in the rewarded low vigilance state compared to the rewarded high vigilance state. Consequently, the power in these two frequency bands was thus differently modulated by motivation. It would be tempting to associate the current findings with the distinct roles suggested for the theta and delta bands, suggesting that intrinsically driven regulation of information processing can be influenced by reward

(as reflected by delta-band changes, Knyazev, 2012), but the top-down attentional control (reflected by theta-band changes, Cavanagh and Frank, 2014) is insensitive to reward. Interestingly, accumulating evidence exists linking theta oscillations with the attentional sampling of the environment, especially during higher task demands (Bastiaansen and Hagoort, 2003; Landau et al., 2015; Spyropoulos et al., 2018; Karamacoska et al., 2019). Interpreted in the context of the current results suggesting the insensitivity of the theta band to reward manipulation, theta power might subserve rather low-level attentional sampling, which is not directly linked with the reward system, at least in the context of nonprimary, extrinsic rewards such as money. On the contrary, delta band oscillations may reflect a separate compensatory mechanism (Knyazev, 2012), which supports the recovery of functions after manipulating motivation.

However, these interpretations must be treated with caution, especially regarding the role of fatigue-related oscillatory dynamics. The current experimental paradigm is not optimal for studying ongoing oscillations, and the changes in rhythmic activation are strongly linked with the visual trigger. It is important to distinguish between ongoing oscillations and stimulus-related changes in spectral power. This fact is highlighted by the detected decrease in the spectral power by the decrease in vigilance in the present study, while ongoing oscillations at low-frequency bands generally show a fatigue-related increase (Lal and Craig, 2001). When the influence of phase-locked evoked activation (**Supplementary Figure S2**) was removed from the spectral responses, the theta-band modulations strongly decreased. Rather than reflecting neural computation in the theta band, the time-frequency results might be at least partly driven by the phase-locked evoked responses. Our analysis can be seen as advancing the interpretability of evoked responses, and different experimental paradigms are needed to focus purely on the fatigue-related changes in oscillatory dynamics.

Based on our results, vigilance decrement changes the neural processes underlying selective visual attention, as demonstrated by the changes in the spectral power at 1–4 Hz and at 4–8 Hz, as well as evoked P300 response. Motivation plays a different role in the high and low vigilance states, with improvement of performance only in the low vigilance state. This appears inconsistent with the active fatigue framework (which states that the vigilance decrement is the result of the depletion of cognitive resources, and motivation cannot improve the performance impaired by vigilance decrement, Helton and Warm, 2008) because the impairment in the state of low vigilance is improved after motivation manipulation. On the other hand, our results seem to agree with the motivation control framework—that vigilance decrement is a subconscious balancing between the costs and benefits to expend or conserve energy (Kurzban et al., 2013a). When the cost of efforts to carry out a task outweighs the benefits, humans are unwilling to do so, leading to vigilance decrement. However, not all of the neural measures are improved after providing reward. The P300 latency in the low vigilance state was not modulated by reward. Furthermore, the spectral power in the delta but not in the theta band was modulated by motivation manipulation, which means that motivation partially

alleviates neural activity in the low vigilance state. In general, our results imply that motivation is not enough to completely restore the impairment induced by vigilance decrement and provide support for the mental fatigue framework, which integrates the evaluation of expected rewards and energetic costs (Boksem and Tops, 2008).

Further studies are inevitably needed to establish a more comprehensive picture of the underlying neural processes affected by motivation and vigilance states. We only analyze the changes in high vs. low vigilance states; nevertheless, focusing on the ongoing changes while performing the task can significantly advance the understanding of the dynamic emergence of mental fatigue. Our study did not consider the effects of monetary values during a long period of attention task engagement. It is also not possible to completely disengage the dimensions of vigilance and motivation, as it is likely that a decrease in vigilance is also accompanied by decreased motivation to perform a task. Furthermore, providing rewards is not the only method to motivate individuals. Further studies should further elaborate the particular differences in sensitivity to reward (positive) and punishment (negative).

CONCLUSION

Both the behavioral and electrophysiological measures were modulated by vigilance decrement. The neurocognitive processes were only partially recovered by manipulating rewards. In particular, increasing motivation using rewards differentially influenced brain activations in the high vs. low vigilance states, with more evident improvement in the low than in the high vigilance state. The fatigue-related decrease in latency of P300 responses did not recover with rewards, whereas the P300 amplitude increased to the same level as in the high vigilance state. The spectral power of the delta band was specifically increased by motivation, whereas the decrease of the theta band was not recovered by reward. These findings provide evidence for the dissociable effects of motivation in the states of low and high vigilance and might validate the mental fatigue framework integrating the evaluation of expected rewards and energetic costs.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed during the present study are available from the corresponding author on reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Liaoning Normal University. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JL, HS, and TP designed the experiment. JL, CZ, and YZ analyzed the data. JL and YL collected the data. JL and TP conducted the

statistics. JL, FC, TP, and TR wrote the manuscript. FC, TR, and TP provided the fundings and guidance for all the conduction of work.

REFERENCES

Bastiaansen, M., and Hagoort, P. (2003). Event-induced theta responses as a window on the dynamics of memory. *Cortex* 39, 967–992. doi: 10.1016/s0010-9452(08)70873-6

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate? A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Benoit, C.-E., Solopchuk, O., Borragán, G., Carbonnelle, A., Van Durme, S., and Zénon, A. (2019). Cognitive task avoidance correlates with fatigue-induced performance decrement but not with subjective fatigue. *Neuropsychologia* 123, 30'40. doi: 10.1016/j.neuropsychologia.2018.06.017

Boksem, M. A. S., Meijman, T. F., and Lorist, M. M. (2005). Effects of mental fatigue on attention: an ERP study. *Cogn. Brain Res.* 25, 107–116. doi: 10.1016/j.cogbrainres.2005.04.011

Boksem, M. A. S., Meijman, T. F., and Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biol. Psychol.* 72, 123–132. doi: 10.1016/j.biopsycho.2005.08.007

Boksem, M. A. S., and Tops, M. (2008). Mental fatigue: costs and benefits. *Brain Res. Rev.* 59, 125–139. doi: 10.1016/j.brainresrev.2008.07.001

Boksem, M. A. S., Tops, M., Kostermans, E., and De Cremer, D. (2008). Sensitivity to punishment and reward omission: evidence from error-related ERP components. *Biol. Psychol.* 79, 185–192. doi: 10.1016/j.biopsycho.2008.04.010

Cavanagh, J. F., and Frank, M. J. (2014). Frontal theta as a mechanism for affective and effective control. *Trends Cogn. Sci.* 18, 414–421. doi: 10.1016/j.tics.2014.04.012

Demiralp, T., Ademoglu, A., Comerchero, M., and Polich, J. (2001). Wavelet analysis of P3a and p3b. *Brain Topogr.* 13, 251–267. doi: 10.1023/a:1011102628306

Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16, 143–149. doi: 10.3758/bf03203267

Faber, L. G., Maurits, N. M., and Lorist, M. M. (2012). Mental fatigue affects visual selective attention. *PLoS One* 7:e48073. doi: 10.1371/journal.pone.0048073

Gergelyfi, M., Jacob, B., Olivier, E., and Zénon, A. (2015). Dissociation between mental fatigue and motivational state during prolonged mental activity. *Front. Behav. Neurosci.* 9:176. doi: 10.3389/fnbeh.2015.00176

Gulbinaite, R., Johnson, A., de Jong, R., Morey, C. C., and van Rijn, H. (2014). Dissociable mechanisms underlying individual differences in visual working memory capacity. *NeuroImage* 99, 197–206. doi: 10.1016/j.neuroimage.2014.05.060

Guo, Z., Chen, R., Liu, X., Zhao, G., Zheng, Y., Gong, M., et al. (2018). The impairing effects of mental fatigue on response inhibition: an ERP study. *PLoS One* 13:e0198206. doi: 10.1371/journal.pone.0198206

Harmony, T., Fernández, T., Silva, J., Bernal, J., Díaz-Comas, L., Reyes, A., et al. (1996). EEG delta activity: an indicator of attention to internal processing during performance of mental tasks. *Int. J. Psychophysiol.* 24, 161–171. doi: 10.1016/s0167-8760(96)00053-0

Helton, W. S., and Russell, P. N. (2015). Rest is best: the role of rest and task interruptions on vigilance. *Cognition* 134, 165–173. doi: 10.1016/j.cognition.2014.10.001

Helton, W. S., and Warm, J. S. (2008). Signal salience and the mindlessness theory of vigilance. *Acta Psychol.* 129, 18–25. doi: 10.1016/j.actpsy.2008.04.002

Himberg, J., and Hyvärinen, A. (2003). "ICASSO: software for investigating the reliability of ICA estimates by clustering and visualization," in *In Proceedings of the IEEE Workshop on Neural Networks for Signal Processing (NNSP'2003)* (Toulouse, France: IEEE), 259–268.

Hopstaken, J. F., van der Linden, D., Bakker, A. B., and Kompier, M. A. J. (2015). A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology* 52, 305–315. doi: 10.1111/psyp.12339

Karamacoska, D., Barry, R. J., De Blasio, F. M., and Steiner, G. Z. (2019). EEG-ERP dynamics in a visual continuous performance test. *Int. J. Psychophysiol.* 146, 249–260. doi: 10.1016/j.ijpsycho.2019.08.013

Kastner, S., and Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23, 315–341. doi: 10.1146/annurev.neuro.23.1.315

Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., Kübler, A., and Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface. *Biol. Psychol.* 102, 118–129. doi: 10.1016/j.biopsycho.2014.07.014

Kato, Y., Endo, H., and Kizuka, T. (2009). Mental fatigue and impaired response processes: event-related brain potentials in a Go/NoGo task. *Int. J. Psychophysiol.* 72, 204–211. doi: 10.1016/j.ijpsycho.2008.12.008

Keller, A. S., Payne, L., and Sekuler, R. (2017). Characterizing the roles of alpha and theta oscillations in multisensory attention. *Neuropsychologia* 99, 48–63. doi: 10.1016/j.neuropsychologia.2017.02.021

Knyazev, G. G. (2012). EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neurosci. Biobehav. Rev.* 36, 677–695. doi: 10.1016/j.neubiorev.2011.10.002

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology* 38, 557–577. doi: 10.1017/s0048577201990559

Kuba, M., Kremlácek, J., Langrová, J., Kubová, Z., Szanyi, J., and Vít, F. Š. (2012). Aging effect in pattern, motion and cognitive visual evoked potentials. *Vision Res.* 62, 9–16. doi: 10.1016/j.visres.2012.03.014

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00207/full#supplementary-material.

FIGURE S1 | Scatter diagrams showing the relationships between the behavioral measures (accuracy, reaction time, and number of omitted responses) and the P300 amplitude and latency in the four conditions. NRHV is block 1 in the no-reward high vigilance state, RHV is block 2 in the reward high vigilance state, NRLV is block 5 in the no-reward low vigilance state, and RLV is block 6 in the reward low vigilance state.

FIGURE S2 | Comparison of different ways of calculating time-frequency representation changes for the current data. **(A)** Calculation of the power with the continuous wavelet transform (CWT) in each trial, and then averaged (presented in the present study). **(B)** Calculation of the power with the CWT from epochs, from which the contribution of averaged evoked responses are removed from each trial (averaged ERP is subtracted from each trial). **(C)** Calculation of the power with the CWT from epochs averaged in the evoked responses (spectra power of averaged ERP).

Kurzban, R., Duckworth, A., Kable, J. W., and Myers, J. (2013a). An opportunity cost model of subjective effort and task performance. *Behav. Brain Sci.* 36, 661–679. doi: 10.1017/s0140525x12003196

Kurzban, R., Duckworth, A., Kable, J. W., and Myers, J. (2013b). Cost-benefit models as the next, best option for understanding subjective effort. *Behav. Brain Sci.* 36, 707–726. doi: 10.1017/s0140525x13001532

Kutas, M., McCarthy, G., and Donchin, E. (1977). Augmenting mental chronometry?: the P300 as a measure of stimulus evaluation time. *Science* 197, 792–795. doi: 10.1126/science.887923

Lal, S. K. L., and Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biol. Psychol.* 55, 173–194. doi: 10.1016/s0301-0511(00)00085-5

Landau, A. N., Schreyer, H. M., Van Pelt, S., and Fries, P. (2015). Distributed attention is implemented through theta-rhythmic gamma modulation. *Curr. Biol.* 25, 2332–2337. doi: 10.1016/j.cub.2015.07.048

Lim, J., and Kwok, K. (2016). The effects of varying break length on attention and time on task. *Hum. Factors* 58, 472–481. doi: 10.1177/0018720815617395

Lorist, M. M., Boksem, M. A. S., and Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cogn. Brain Res.* 24, 199–205. doi: 10.1016/j.cogbrainres.2005.01.018

Lorist, M. M., Klein, M., and Nieuwenhuis, S. (2000). Mental fatigue and task control? planning and preparation mental fatigue and task control?: planning and preparation. *Psychophysiology* 37, 614–625. doi: 10.1111/1469-8986.3750614

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique.* Cambridge, MA: MIT Press.

Mackworth, N. H., Kaplan, I. T., and Metlay, W. (1964). Eye movements during vigilance. *Percept. Mot. Skills* 18, 397–402. doi: 10.2466/pms.1964.18.2.397

Marcora, S. M., Staiano, W., and Manning, V. (2009). Mental fatigue impairs physical performance in humans. *J Appl. Physiol.* 106, 857–864. doi: 10.1152/japplphysiol.91324.2008

May, J. F., and Baldwin, C. L. (2009). Driver fatigue: the importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transp. Res. Part F Traffic Psychol. Behav.* 12, 218–224. doi: 10.1016/j.trf.2008.11.005

Möckel, T., Beste, C., and Wascher, E. (2015). The effects of time on task in response selection—an ERP study of mental fatigue. *Sci. Rep.* 5:10113. doi: 10.1038/srep10113

Moore, T., and Zirnsak, M. (2017). Neural mechanisms of selective visual attention. *Ann. Rev. Psychol.* 68, 47–72. doi: 10.1146/annurev-psych-122414-033400

Mordkoff, J. T. (2019). A simple method for removing bias from a popular measure of standardized effect size: adjusted partial eta squared. *Adv. Methods Pract. Psychol. Sci.* 2, 228–232. doi: 10.1177/2515245919855053

Oken, B. S., Salinsky, M. C., and S. Elsas, M. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clin. Neurophysiol.* 117, 1885–1901. doi: 10.1016/j.clinph.2006.01.017

Paschke, L. M., Walter, H., Steimke, R., Ludwig, V. U., Gaschler, R., Schubert, T., et al. (2015). Motivation by potential gains and losses affects control processes *via* different mechanisms in the attentional network. *NeuroImage* 111, 549–561. doi: 10.1016/j.neuroimage.2015.02.047

Polich, J. (2009). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Polich, J., and Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biol. Psychol.* 41, 103–146. doi: 10.1016/0301-0511(95)05130-9

Reteig, L. C., van den Brink, R. L., Prinssen, S., Cohen, M. X., and Slagter, H. A. (2019). Sustaining attention for a prolonged period of time increases temporal variability in cortical responses. *Cortex* 117, 16–32. doi: 10.1016/j.cortex.2019.02.016

Robert, D., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.

Seli, P., Cheyne, J. A., Xu, M., Purdon, C., and Smilek, D. (2015). Motivation, intentionality and mind wandering: implications for assessments of task-unrelated thought. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 1417–1425. doi: 10.1037/xlm0000116

Shen, J., Barbera, J., and Shapiro, C. M. (2006). Distinguishing sleepiness and fatigue: focus on definition and measurement. *Sleep Med. Rev.* 10, 63–76. doi: 10.1016/j.smrv.2005.05.004

Spyropoulos, G., Bosman, C. A., and Fries, P. (2018). A theta rhythm in macaque visual cortex and its attentional modulation. *Proc. Natl. Acad. Sci. U S A* 115, E5614–E5623. doi: 10.1073/pnas.1719433115

Taya, F., Dimitriadis, S. I., Dragomir, A., Lim, J., Sun, Y., Wong, K. F., et al. (2018). Fronto-parietal subnetworks flexibility compensates for cognitive decline due to mental fatigue. *Hum. Brain Mapp.* 39, 3528–3545. doi: 10.1002/hbm.24192

Thomson, D. R., Besner, D., and Smilek, D. (2015). A resource-control account of sustained attention: evidence from mind-wandering and vigilance paradigms. *Perspect. Psychol. Sci.* 10, 82–96. doi: 10.1177/1745691614556681

Tops, M., Boksem, M. A. S., Wester, A. E., Lorist, M. M., and Meijman, T. F. (2006). Task engagement and the relationships between the error-related negativity, agreeableness, behavioral shame proneness and cortisol. *Psychoneuroendocrinology* 31, 847–858. doi: 10.1016/j.psyneuen.2006.04.001

van Dinteren, R., Arns, M., Jongsma, M. L. A., and Kessels, R. P. C. (2014). P300 development across the lifespan: a systematic review and meta-analysis. *PLoS One* 9:e87347. doi: 10.1371/journal.pone.0087347

Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology* 34, 131–156. doi: 10.1111/j.1469-8986.1997.tb02125.x

Zhang, C., Cong, F., Kujala, T., Liu, W., Liu, J., Parviainen, T., et al. (2018). Network entropy for the sequence analysis of functional connectivity graphs of the brain. *Entropy* 20:311. doi: 10.3390/e20050311

Is the Deliberate Practice View Defensible? A Review of Evidence and Discussion of Issues

David Z. Hambrick[1]*, Brooke N. Macnamara[2] and Frederick L. Oswald[3]

[1] *Department of Psychology, Michigan State University, East Lansing, MI, United States,* [2] *Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH, United States,* [3] *Department of Psychological Sciences, Rice University, Houston, TX, United States*

The question of what explains individual differences in expertise within complex domains such as music, games, sports, science, and medicine is currently a major topic of interest in a diverse range of fields, including psychology, education, and sports science, to name just a few. Ericsson and colleagues' *deliberate practice view* is a highly influential perspective in the literature on expertise and expert performance—but is it viable as a testable scientific theory? Here, reviewing more than 25 years of Ericsson and colleagues' writings, we document critical inconsistencies in the definition of deliberate practice, along with apparent shifts in the standard for evidence concerning deliberate practice. We also consider the impact of these issues on progress in the field of expertise, focusing on the empirical testability and falsifiability of the deliberate practice view. We then discuss a multifactorial perspective on expertise, and how open science practices can accelerate progress in research guided by this perspective.

Keywords: deliberate practice, expertise, talent, skill, individual differences

IS THE DELIBERATE PRACTICE VIEW DEFENSIBLE? A REVIEW OF EVIDENCE AND DISCUSSION OF ISSUES

Not infrequently, a single theoretical perspective becomes extremely influential in an area of scientific inquiry, shaping the trajectory of research in the field for years or even decades. More than 25 years ago, K. Anders Ericsson and colleagues proposed what has arguably become the most influential theoretical perspective in the scientific literature on expertise and expert performance. In a pivotal *Psychological Review* article, Ericsson et al. (1993) theorized that expert performance reflects a long period of *deliberate practice,* which they stated "includes activities that have been specially designed to improve the current level of performance" (p. 368). In studies of violinists (Study 1) and pianists (Study 2), Ericsson et al. (1993) operationally defined deliberate practice as "practice alone" with the goal of improving performance. The most accomplished musicians reported having accumulated an average of around 10,000 h of practice alone by early adulthood, which was thousands of hours more than the averages for this measure of practice for less accomplished groups.

Applying their framework to several domains of expertise, Ericsson et al. (1993) concluded that "individual differences in ultimate performance can largely be accounted for by differential

amounts of past and current levels of practice" (p. 392). They further explained:

> [H]igh levels of deliberate practice are necessary to attain expert level performance. Our theoretical framework can also provide a sufficient account of the major facts about the nature and scarcity of exceptional performance. Our account does not depend on scarcity of innate ability (talent) (Ericsson et al., 1993, p. 392).

Reiterating this perspective, Ericsson et al. (2007) stated that "[i]t is possible to account for the development of elite performance among healthy children without recourse to unique talent (genetic endowment)—excepting the innate determinants of body size" (p. 4). And writing in the *Harvard Business Review*, Ericsson et al. (2007) explained, "Our research shows that even the most gifted performers need a minimum of 10 years (or 10,000 h) of intense training before they win international competitions" (p. 119).

Ericsson and colleagues' perspective, which we refer to as the *deliberate practice view*, has had a monumental impact on expertise research. As of this publication, the Ericsson et al. (1993) article has been cited nearly 11,000 times in a wide range of literatures, and there have been nearly 200 theses and dissertations on deliberate practice in universities around the world. As portrayed in popular press books such as Malcolm Gladwell's (2008) bestseller *Outliers: The Story of Success* and Daniel Coyle's (2009) *The Talent Code*, Ericsson and colleagues' research has also had a profound influence on the public's thinking about the origins of expertise. Taking his inspiration from Ericsson et al.'s (1993) findings, Gladwell wrote that "10,000 h is the magic number of true expertise" (p. 11). In his own popular press book, *Peak: Secrets from the New Science of Expertise*, Ericsson wrote, "Deliberate practice can open the door to a world of possibilities that you may have been convinced were out of reach. Open that door" (Ericsson and Pool, 2016, p. 179).

We credit and commend Ericsson and colleagues for their highly influential work. However, here we will discuss what we believe are serious concerns with whether the deliberate practice view is viable as a *scientific theory*—that is, whether it is empirically testable and falsifiable. [For a similar type of review, see Gottfredson's (2003) critique of Sternberg's practical intelligence theory; Sternberg et al. (1995)]. Before doing so, however, we note two uncontroversial claims about expertise, by which we simply mean a person's measurable (i.e., quantifiable) level of performance in a domain. First, as Ericsson and colleagues have emphasized (e.g., Ericsson, 2006), *expertise is acquired gradually*. In other words, people are not *literally* born as experts, innately endowed with the type of specialized knowledge that underpins high-level skill in activities like hitting a golf ball, solving math equations, playing an instrument, or choosing a move in a chess game. Domain-specific knowledge and skill can only be acquired gradually over time through some form of training.

The second uncontroversial claim is that training can lead to large, even massive, improvements in people's level of expertise (i.e., domain-relevant performance). This point was amply illustrated by some of Ericsson and colleagues' earliest research. For example, in a classic study, Ericsson et al. (1980) showed that

after more than 200 h of training, a college student improved his performance in a random digit memorization task from a typical 7 digits to 79 digits (the world record is currently an astounding 547 digits)[1]. In short, notwithstanding the issues raised in this article, Ericsson and colleagues' deliberate practice view has important value in society, serving as a useful reminder to the layperson that training of some form is necessary to achieve a high level of performance in a domain.

The controversial question in research on expertise is not whether some form of training is necessary to explain intraindividual (i.e., within-person) increases in expertise (it must be), or whether these increases can be massive (they can be). Rather, the controversial question is the extent to which interindividual (i.e., between-person) differences in accumulated amount of training explain interindividual differences in expertise (for a discussion of the distinctions between interindividual and intraindividual variability, see Molenaar et al., 2003). In statistical terms, what is the direction (and the magnitude) of the correlation between expertise and accumulated amount of training? Somewhat counterintuitively, as **Figure 1** illustrates, the necessity of training to explain intraindividual increases in expertise has no direct implication for the answer to this question. That is, taking as a given that the relationship between training and expertise is positive within individuals, the correlation between training and expertise *between* individuals could be positive (top panel), indicating higher levels of performance for individuals who have engaged in more training; negative (middle panel), indicating lower levels of performance for individuals who have engaged in more training; or zero (bottom panel).

What, then, is the correlation between deliberate practice and expertise *across* individuals? The first step in attempting to answer this question is to operationalize deliberate practice—that is, to develop measures of deliberate practice based on the definition of the construct. Unfortunately, as we document in this article, there remains a great deal of confusion about the definition of deliberate practice, despite more than 25 years of research on the topic. As Ericsson and colleagues themselves recently noted: "It has been common for scientists to be confused about the definition of DP (deliberate practice)" (Dearani et al., 2017, p. 1333). Here, we discuss possible sources of this continued confusion, the confusion around the measurement of deliberate practice that has resulted (e.g., inclusion criteria for meta-analyses), and the impact of this confusion on the science of expertise. We conclude this article with thoughts on how to advance the scientific study of expertise and expert performance. For a companion presentation to this article, visit https://osf.io/buqsk/.

WHAT IS DELIBERATE PRACTICE?

It is undoubtedly the case that different types of domain-relevant activities vary in their importance for developing expertise. For example, training under a qualified golf instructor is almost

---

[1] http://www.world-memory-statistics.co.uk/disciplines.php

FIGURE 1 | Schematic illustration of a positive (top panel), negative (middle panel), and zero (bottom panel) between-person correlation (as illustrated by the dashed ellipses) between training hours and expertise, assuming a positive relationship between the variables within individuals (as represented by the learning curves for four hypothetical individuals).

certainly more beneficial for improving golf skill than mindlessly hitting practice balls at a driving range. In their article on deliberate practice, Ericsson et al. (1993) distinguished among three forms of domain-specific experience. They described

work as engagement in activities for external rewards (e.g., music performances, sports competitions), *play* as participating in activities for pleasure (e.g., playing a sport with friends for recreation), and *deliberate practice* as a "highly structured activity, the explicit goal of which is to improve performance" (Ericsson et al., 1993, p. 368).

However, Ericsson and colleagues have been inconsistent on critical elements of the definition of deliberate practice, and consequently it has been unclear what activities do and do not qualify as deliberate practice. For example, Ericsson et al. (1993) stated that "the teacher designs practi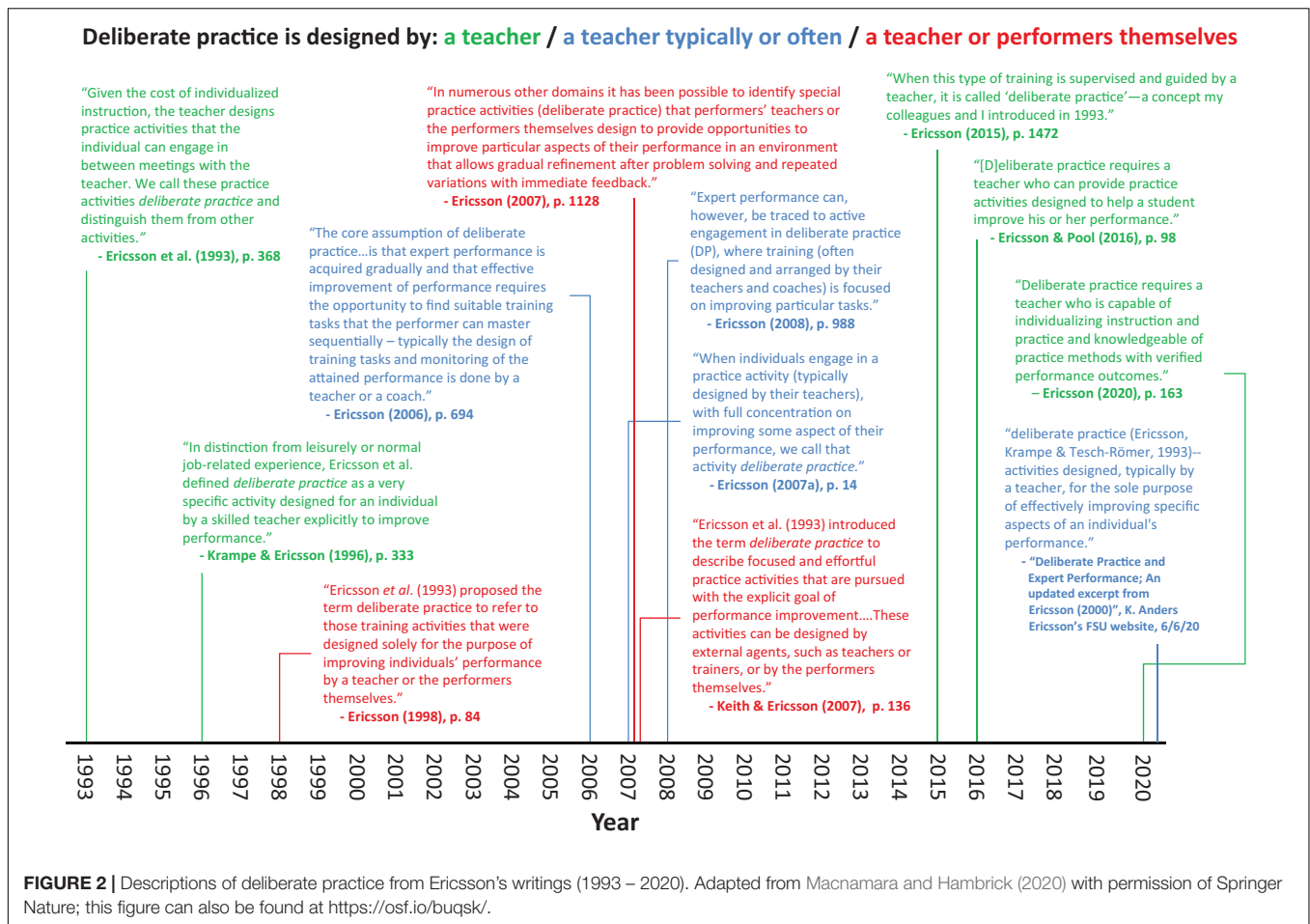ce activities that the individual can engage in between meetings with the teacher" (p. 368). A few years later, however, Ericsson (1998) stated that "Ericsson et al. (1993) proposed the term deliberate practice to refer to those training activities that were designed solely for the purpose of improving individuals' performance by a teacher *or the performers themselves*" (p. 84, emphasis added). This latter statement indicated that deliberate practice, as Ericsson et al. (1993) originally defined the term, encompasses a broader range of activities than just teacher-designed practice. Yet, as shown in **Figure 2**, in subsequent articles, Ericsson and colleagues were inconsistent on this critical point, sometimes indicating that deliberate practice must be designed by a teacher (e.g., Ericsson, 2015), but other times stating that it can be designed by teachers or the "performers themselves" (e.g., Keith and Ericsson, 2007). If deliberate practice *must* be designed by a teacher, then presumably it cannot also be designed by performers themselves.

As another example of definitional confusion, it is unclear from Ericsson and colleagues' writings whether deliberate practice must be a solitary activity, or whether it can also be a group/team activity. Citing research on team sports (Helsen et al., 1998), Ericsson (2006) observed that "the amount of time spent in team-related deliberate practice activities correlates reliably with skill level in team sports" (p. 695). It seems clear from this observation that there can be team deliberate practice. Recently, however, Ericsson and Harwell (2019) indicated that this is *not* the case, commenting that "it is important to point out that organized team training may be quite effective in improving performance, but it does not meet all the criteria for deliberate practice" (p. 6). This is another apparent shift in the definition of and criteria for deliberate practice, creating still more confusion about what the "correct" definition is. In short, it is unclear what activities do and do not qualify as deliberate practice.

This confusion surrounding the definition of deliberate practice is not a minor matter—it directly impacts how deliberate practice is measured in empirical studies and what evidence (i.e., effect sizes) should be included in a meta-analysis. These decisions, in turn, directly impact the evaluation of the deliberate practice view: whether evidence is concluded to support the view or not. Certainly, definitions of theoretical constructs can and do evolve over time as science progresses, but the shifts in the definition of deliberate practice reflected in **Figure 2** do not appear to reflect this sort of progression. There are meaningful changes even over short spans of time (e.g., compare Keith and Ericsson's, 2007, description of deliberate practice with Ericsson's, 2006). Confusion around the definition of deliberate practice persists in the expertise

FIGURE 2 | Descriptions of deliberate practice from Ericsson's writings (1993 – 2020). Adapted from Macnamara and Hambrick (2020) with permission of Springer Nature; this figure can also be found at https://osf.io/buqsk/.

literature, even as researchers are attempting to investigate the deliberate practice view.

Challenges to the Deliberate Practice View

Notwithstanding this confusion over the definition of deliberate practice, there have been numerous attempts to test the deliberate practice view. One of the first noteworthy tests came from a study of chess expertise by Gobet and Campitelli (2007), who administered a questionnaire to 90 members of a Buenos Aries chess club to assess lifetime engagement in deliberate practice and tournament chess rating. The self-reported amount of deliberate practice (hours of studying alone plus hours of group practice) correlated positively and moderately with chess rating ($r = 0.42$). This is a sizeable correlation by psychological standards—a "medium" effect size in Cohen's (1992) widely used classification scheme (i.e., $r = 0.10$, small; $r = 0.30$, medium; $r = 0.50$, large). However, this finding also challenges the deliberate practice view because it means that deliberate practice left a large amount of the total variance in chess ratings *un*explained. To be exact, a correlation of $r = 0.42$ between two variables indicates that one variable explains 18% of the variance in the other variable (i.e., $r = 0.42 \times 100 = 18\%$). In the present case, it must be

assumed that some of this unexplained variance reflects random measurement error, because neither a measure of deliberate practice nor a measure of performance can be assumed to be perfectly reliable (we discuss this issue further below). However, the correlation was not large enough to suggest that, even after taking this psychometric artifact into account, participants at similar levels of chess skill would have reported similar amounts of deliberate practice. Instead, it suggests that the chess players varied substantially in the amount of deliberate practice they required to reach a given level of skill. Indeed, according to the data, they did: As Gobet and Campitelli (2007) described in their article, the self-reported estimate of number of hours of deliberate practice required to reach "master" status in their sample ranged from 3,016 to 23,608 h—a difference of nearly a factor of 8. The implication is that although deliberate practice clearly contributes to individual differences in chess expertise, other factors must contribute as well, as we discuss further in the final section of this article.

In our own first effort to test the deliberate practice view (Hambrick et al., 2014b), we reanalyzed results from expertise studies in the domains of chess and music. Our specific goal was to test Ericsson et al.'s (1993) aforementioned claim that "individual differences in ultimate performance can *largely be accounted for* by differential amounts of past and current levels

of practice" (p. 392, emphasis added). We identified six studies of chess and eight studies of music that reported a correlation between a measure of deliberate practice and performance. The average correlation was $r = 0.49$ for chess and $r = 0.43$ for music before applying the standard psychometric correction for measurement error variance (unreliability) of the constituent measures. For deliberate practice, we assumed a reliability coefficient of 0.80 based on information we could find about the reliability of this variable, as well as on Tuffiash et al.'s (2007) statement that "self-report practice estimates repeatedly from experts in sports and music have reported test-retest reliabilities at or above 0.80" (p. 129) and Ericsson's (2013) statement that "[t]he collected reliability of cumulated life-time practice at different test occasions in large samples has typically been found to range between 0.7 and 0.8" (p. 534). For music performance, we used reliability estimates from the studies, or if not reported, from studies that collected similar performance measures. The average amount of reliable variance in expertise explained by deliberate practice was 34% for chess and 29.9% for music. This is a substantial amount of variance, but it is not enough to support the claim that deliberate practice largely accounts for individual differences in expertise. This claim implies that deliberate practice should at least explain *most* of the variance in expertise, and evidence suggests it does not.

Subsequently, we set out to test the importance of deliberate practice as a predictor of individual differences in expertise, by way of a formal and comprehensive meta-analysis (Macnamara et al., 2014; see also Macnamara et al.'s, 2018, corrigendum for the article), ultimately reviewing over 11,000 articles. Ericsson et al. (1993) explained that deliberate practice "includes activities that have been specially designed to improve the current level of performance" (p. 368). Accordingly, we defined deliberate practice as structured activities designed to improve performance in a domain; and given Ericsson and colleagues' inconsistency on whether a teacher is required to design deliberate practice, we decided to include both teacher- and performer-designed activities. Identifying 88 studies, we found that deliberate practice explained 14% of the variance in performance overall, and 24% for games, 23% for music, 20% for sports, 5% for education, and 1% for professions. We also determined that deliberate practice left more of the variance in performance unexplained than it explained, across a range of possible values for measurement reliability. We concluded that the "amount of deliberate practice—although unquestionably important as a predictor of individual differences in performance from both a statistical and a practical perspective—is not as important as Ericsson and his colleagues have argued" (Macnamara et al., 2014, p. 1615).

In a later meta-analysis that focused on sports (Macnamara et al., 2016b), we found that the contribution of deliberate practice to sports performance varied by skill level: Among elite athletes (e.g., national-level and above), deliberate practice explained only 1% of the performance variance. Although it must be assumed that range restriction (another psychometric issue) would limit the deliberate practice-performance correlation when considering only elite performers, it is critical to note that it was Ericsson et al. (1993) themselves who stated that deliberate practice is *still* an important predictor of performance differences at the elite level. In their own words, "Individual differences, *even among elite performers,* are closely related to assessed amounts of deliberate practice" (Ericsson et al., 1993, p. 363, emphasis added). This finding from the Macnamara et al. (2016b) meta-analysis on sports is inconsistent with this claim. Furthermore, we found the relationship between deliberate practice and performance to be very similar whether the practice activities were solitary or in a group (see Macnamara et al., 2016b).

In a commentary on our meta-analysis (Macnamara et al., 2014), Ericsson (2014a) rejected 87 of the 88 studies that we included in our meta-analysis, claiming that we included studies that "violated [their] criteria for deliberate practice" (p. 2). However, in doing so, Ericsson (2014a) rejected numerous studies that he himself had previously used to explicitly argue for the importance of deliberate practice (see quotations from Ericsson's writings in **Table 1**). Thus, by any reasonable account, the standard for evidence concerning deliberate practice had shifted dramatically. Most perplexingly, in applying this new standard for evidence, Ericsson rejected several of his own studies of deliberate practice (e.g., Duffy et al., 2004; Tuffiash et al., 2007; Duckworth et al., 2011), seeming to undermine the case he had for decades been attempting to make for the importance of deliberate practice. Ericsson did not acknowledge that he had once used these studies he was now rejecting to argue for the importance of deliberate practice. This evaluation of evidence challenging the deliberate practice view seems indefensible.

Ericsson's (2014a) rejection of his own study of darts (Duffy et al., 2004) and his rejection of Charness and colleagues' studies of chess (Charness et al., 2005) were especially noteworthy (see the **Appendix** for Ericsson's varying characterizations of the Charness studies). Ericsson's (2014a) stated reason for rejecting these studies was that they provided "no record of a teacher/coach supervising all or most of practice" (see Ericsson, 2014b, **Table 2**). However, in a chapter published in the very same year, Ericsson (2014c) used both these studies to argue for the importance of deliberate practice, stating:

> [I]n a study of district, national, and professional dart players Duffy, Baluch, and Ericsson (2004) found that solitary deliberate practice was closely related to performance, whereas the amount of social dart activities did not predict performance (Ericsson, 2014c, p. 191).

> In chess, Charness and his colleagues (Charness, Krampe, and Mayr, 1996; Charness, Tuffiash, Krampe, Reingold, and Vasyukova, 2005) have found that the amount of solitary chess study was the best predictor of performance at chess tournaments, and when this type of deliberate practice was statistically controlled, there was no reliable benefit from playing chess games (Ericsson, 2014c, p. 191)[2].

Another inconsistency was Ericsson's (2014a) rejection of his own study of spelling bee contestants (Duckworth et al., 2011) for

---

[2]There were two studies in this project. Charness et al. (1996) is a chapter and provides an initial report of data from Study 1 ($N = 136$ of an eventual $N = 239$); Charness et al. (2005) is a journal article and reports the full results (Study 1 $N = 239$; Study 2 $N = 169$).

TABLE 1 | Examples of studies that Ericsson rejected for violating his criteria for deliberate practice but previously used to argue for the importance of deliberate practice.

Study rejected by Ericsson (2014a) for violating his criteria for deliberate practice	Previous use of the same study by Ericsson and colleagues to argue for the importance of deliberate practice
Hodges and Starkes (1996)[1]	"Several studies and reviews have since found a consistent relation between performance and amount and quality of **deliberate practice**…in sports (…Hodges and Starkes, 1996…)." - Ericsson (1998, p. 87)
Helsen et al. (1998)[1]	"Research conducted in several domains such as…sports (Helsen, Starkes, and Hodges, 1998…) suggests that the amount of accumulated **deliberate practice** is closely related to an individual's attained level of performance." - Keith and Ericsson's (2007, p. 136)
Duffy et al. (2004)[2]	"The engagement of the dart-related activities differed between groups for three types, namely playing in league darts, solitary practice and total **deliberate practice.** The latter two findings were in line with prior expectations namely; the more an individual engages in **deliberate practice** (particularly solitary practice) the more proficient their performance is likely to be. This finding supports one of the main tenets of Ericsson et al.'s (1993) theory whereby expertise is acquired through a vast number of hours spent engaging in activities purely designed to improve performance, i.e., **deliberate practice**." - Duffy et al. (2004, pp. 242–243).
	"[I]n a study of district, national, and professional dart players Duffy, Baluch, and Ericsson (2004) found that solitary **deliberate practice** was closely related to performance, whereas the amount of social dart activities did not predict performance." - Ericsson (2014c, p. 191)
Charness et al. (2005)[2]	"The paper by Charness, Tuffiash, Krampe, Reingold, and Vasyukova [2005]…extends an earlier classic chapter by Charness, Krampe, and Mayr (1996) and examines retrospective estimates by a large sample of chess players about their training during the development of their skill and expertise. This paper reports the most compelling and detailed evidence for how designed training (**deliberate practice**) is the crucial factor in developing expert chess performance." - Ericsson (2005, p. 237).
	"In chess, Charness and his colleagues (Charness, Krampe, and Mayr, 1996; Charness, Tuffiash, Krampe, Reingold, and Vasyukova, 2005) have found that the amount of solitary chess study was the best predictor of performance at chess tournaments, and when this type of deliberate practice was statistically controlled, there was no reliable benefit from playing chess games." - Ericsson (2014c, p. 191)
Tuffiash et al. (2007)[2]	"Several researchers have reported a consistent association between the amount and quality of solitary activities meeting the criteria of **deliberate practice** and performance in different domains of expertise, such as…SCRABBLE (Tuffiash et al., 2007)." - Ericsson et al. (2009, p. 9)
Duckworth et al. (2011)[2]	"Our major findings in this investigation are as follows: **Deliberate practice**—operationally defined in the current investigation as the solitary study of word spellings and origins—was a better predictor of National Spelling Bee performance than either being quizzed by others or engaging in leisure reading. With each year of additional preparation, spellers devoted an increasing proportion of their preparation time to **deliberate practice**…Grittier spellers engaged in **deliberate practice** more so than their less gritty counterparts, and hours of **deliberate practice** fully mediated the prospective association between grit and spelling performance." - Duckworth et al. (2011, p. 178)

In each quotation, the boldface emphasis on "deliberate practice" is added. [1]Rejected because article "do[es] not record assigned individualized practice tasks with immediate feedback and goals for practice" (see Ericsson, 2014b; Table 3). [2]Rejected because article "do[es] not record a teacher or coach supervising and guiding all or most of the practice" (see Ericsson, 2014b; Table 2). Table from Hambrick et al. (2018b); used with permission from John Wiley and Sons.

violating this same teacher/coach criterion. Just 2 years earlier, in criticizing a journalist for his description of the study, Ericsson (2012) emphatically stated that the study *had* collected data on deliberate practice:

> In that study we (as I was also one of the co-authors) collected data on 'deliberate practice.' We found that 'Grittier spellers engaged in deliberate practice more so than their less gritty counterparts, and hours of deliberate practice fully mediated the prospective association between grit and spelling performance' (p. 6).

In a commentary on a subsequent meta-analysis of deliberate practice in sports performance, Ericsson (2016) again insisted that our broad definition of deliberate practice was incorrect (for a reply, see Macnamara et al., 2016a). Yet he did not

resolve or acknowledge the material inconsistencies in his past descriptions of deliberate practice, especially those concerning the important question of who designs deliberate practice activities (see **Figure 2**). Furthermore, Ericsson again criticized our inclusion of studies that he had previously used to argue for the importance of deliberate practice (e.g., Helsen et al., 1998; Hodges and Starkes, 1996). It is difficult, if not impossible, for scientists to test a theory if the definition of and standard for evidence are changed repeatedly, with no acknowledgment of and no explanation for the changes (Ferguson and Heene, 2013).

New Types of Practice

Around this same time, in their aforementioned popular press book *Peak: Secrets from the New Science of Expertise,* Ericsson and Pool (2016) proposed a distinction between deliberate practice

TABLE 2 | Ericsson's evaluations of the 14 studies and associated effect sizes included in Ericsson and Harwell's (2019) meta-analysis, from prior to 2014 to the present.

Study	Evaluation by Ericsson prior to Macnamara et al.'s (2014) meta-analysis	Evaluation by Ericsson (2014a) in commentary on Macnamara et al. (2014)	Evaluation by Ericsson and Harwell (2019) in their own meta-analysis
Ericsson et al. (1993) – S1	Used to argue for the importance of DP	Rejected for violating a criterion for DP[1]	Coded as DP
Ericsson et al. (1993) – S2	Used to argue for the importance of DP	Accepted as DP	Coded as DP
Schultetus and Charness (1997)***	No cites to argue for importance of DP	Rejected for violating a criterion for DP[2]	Coded as DP
Baker et al. (2003) – DP2	No cites to argue for importance of DP	Rejected for violating a criterion for DP[2]	Coded as DP
Duffy et al. (2004)	Used to argue for the importance of DP	Rejected for violating a criterion for DP[3]	Coded as PP
Charness et al. (2005) – S1	Used to argue for the importance of DP	Rejected for violating a criterion for DP[3] Reinterpreted as PP*	Coded as DP
Charness et al. (2005) – S2	Used to argue for the importance of DP	Rejected for violating a criterion for DP[3] Reinterpreted as PP*	Coded as DP
De Bruin et al. (2007)**	Used to argue for the importance of DP	Rejected for violating a criterion for DP[3]	Coded as PP
Gobet and Campitelli (2007)	No cites to argue for importance of DP	Rejected for violating a criterion for DP[3]	Coded as DP
Tuffiash et al. (2007)	Used to argue for the importance of DP	Rejected for violating a criterion for DP[3] Reinterpreted as PP*	Coded as PP
Ruthsatz et al. (2008) – S2	No cites to argue for the importance of DP	Rejected for violating a criterion for DP[1]	Coded as DP
Harris (2008) – M1	Dissertation on DP from Ericsson's lab	Rejected for violating a criterion for DP[3]	Coded as PP
Duckworth et al. (2011)	Used to argue for the importance of DP	Rejected for violating a criterion for DP[3] Reinterpreted as PP*	Coded as PP
Maynard et al. (2014)	NA; unpublished study	Rejected because inadequate info. on study	Coded as PP

*DP, deliberate practice; PP, purposeful practice. Each study/effect size in red was rejected by Ericsson (2014a) for violating one of his criteria at that time for deliberate practice (see Supplementary Tables in Ericsson, 2014b): [1] "Restriction of range in attained performance and accumulated deliberate practice" (Ericsson, 2014c; Table 5). [2] "Articles do not record assigned individualized practice tasks with immediate feedback and goals for practice" (Ericsson, 2014c; Table 3). [3] "Articles do not record a teacher or coach supervising and guiding all or most of the practice" (Ericsson, 2014b; Table 2). *See Moxley et al. (2019), for reinterpretation of studies as focusing on purposeful practice rather than deliberate practice. **Ericsson and colleagues (see Ericsson and Towne, 2013) cite another report of this chess study by de Bruin and colleagues to argue for the importance of deliberate practice (De Bruin et al., 2008), but that article is based on the same data as reported in the De Bruin et al. (2007) article. ***As cited in Deakin and Cobley (2003).*

and two new forms of practice, perhaps in an attempt to address discrepancies and confusion surrounding the definition of deliberate practice that were being documented in the scientific literature (e.g., Hambrick et al., 2014a). They introduced and defined *naïve practice* as "essentially just doing something repeatedly, and expecting that the repetition alone will improve one's performance" (p. 14), and *purposeful practice* as an activity that has well-defined, specific goals and involves feedback, but which is self-directed rather than teacher-directed. Ericsson and Pool (2016) explained that "deliberate practice requires a teacher who can provide activities designed to help a student improve his or her performance....With this definition we are drawing a clear distinction between purposeful practice—in which a person tries very hard to push himself or herself to improve— and practice that is both purposeful *and* informed" (p. 98, emphasis added). They further explained that "some approaches to training are more effective than others" (p. 85) and that deliberate practice is "the most effective method of all. . ..the gold standard, the ideal to which anyone learning a skill should aspire" (Ericsson and Pool, 2016, p. 85).

Using this new framework, Ericsson and colleagues reinterpreted studies they had once used to argue for the importance of "deliberate practice" as studies of the less effective "purposeful practice," but without explicitly acknowledging and justifying the reinterpretation (see Macnamara and Hambrick, 2020). As an example, Ericsson (2005) described Charness et al.'s (2005) chess study (entitled *The Role of Deliberate*

Practice in Chess Expertise) as providing "the most compelling and detailed evidence for how designed training (deliberate practice) is the crucial factor in developing expert chess performance" (p. 237). Nevertheless, in their recent article, Moxley et al. (2019) explained that "Charness et al. (2005) found evidence for an independent effect of engagement in *purposeful practice* for chess skill" (p. 1163, emphasis added). As another example, Duckworth, Ericsson, and colleagues' spelling bee study (Duckworth et al., 2011) focused on deliberate practice: The article reporting the study was titled *Deliberate Practice Spells Success: Why Grittier Competitors Triumph at the National Spelling Bee* and the major conclusion of the study was that "[d]eliberate practice mediated the prediction of final performance by the personality trait of grit" (p. 174). Yet the recent Moxley et al. (2019) article stated that "[a]fter the questionnaire, we asked participants to fill out several additional personality measures that Duckworth et al. (2011) had found to be related to *purposeful practice* in preparation for competitions in spelling" (p. 1158, emphasis added).

In another instance, Ericsson and colleagues went from arguing that activities exist that meet the criteria for deliberate practice in the boardgame SCRABBLE, to arguing that it is not possible to engage in deliberate practice in SCRABBLE. Specifically, referring to Tuffiash et al.'s (2007) SCRABBLE study, Ericsson et al. (2009) stated that "[s]everal researchers have reported a consistent association between the amount and quality of solitary activities meeting the criteria of deliberate practice and

performance in different domains of expertise, such as. . .Scrabble (Tuffiash et al., 2007)" (p. 9). However, Moxley et al. (2019) wrote that because SCRABBLE lacks professional coaches "SCRABBLE players cannot engage in deliberate practice, but only purposeful practice and other types of practice" (p. 1150). Under this new framework, activities that once qualified as deliberate practice are now classified as less effective purposeful practice. Of course, it is appropriate for a theorist to reinterpret past evidence as a theory is refined and revised over time. But it is a serious problem, as in this case, when the reinterpretations of evidence are not *explicitly* acknowledged, explained, and justified. In the absence of such transparency, the reader will be led to think that the empirical support for the original theory was stronger than it is or ever was.

Recent Developments

Two reanalyses of our first meta-analysis (Macnamara et al., 2014) have been published in the past two years. The first was by Miller et al. (2018), who argued that some of the studies that we included in our meta-analysis did not capture deliberate practice. Reanalyzing our data, Miller et al. (2018) explained that they had raters code each study "using only the methods section" (p. 6) and explained that "a study was coded as DP (deliberate practice) if and only if it explicitly indicated it estimated the effects of deliberate practice" (p. 6). Their meta-analysis revealed an average correlation of $r = 0.40$ for "deliberate practice" (compared to our value of $r = 0.38$), and an average correlation of $r = 0.21$ for what they deemed "non-deliberate practice." However, as we pointed out in a reply (Hambrick and Macnamara, 2019a; see also Hambrick and Macnamara, in press), numerous studies Miller et al. (2018) coded as deliberate practice did not meet their own inclusion criteria. For example, Miller et al. (2018) coded Bilalić et al.'s (2007) chess study as a deliberate practice study. Yet Bilalić et al. (2007) made no mention of "deliberate practice" anywhere in their methods, and elsewhere in their article Bilalić et al. (see pp. 467–468) explicitly stated that they did *not* interpret their practice measures as deliberate practice. This study and numerous others fulfilled the broad definition of deliberate practice we used in our own meta-analysis, but they did not meet Miller et al.'s (2018) narrower definition. There are two possibilities here: Miller et al. (2018) made errors in coding their studies, or they used coding criteria different from those stated in their article.

In a recent reply, while not addressing the coding issue, Miller et al. (2019) highlighted that the average deliberate practice-performance correlation in their reanalysis and in our meta-analysis were very similar, which is an accurate observation ($r = 0.40$ vs. $r = 0.38$). Then they noted, "Still, something about our analyses [Miller et al.'s 2018, reanalysis] was displeasing to [Hambrick and Macnamara, 2019a]" (Miller et al., 2019, p. 289). We were actually clear about what the problem was: Miller et al.'s (2018) meta-analysis included studies that clearly did not meet their own stated inclusion criterion, rendering their results uninterpretable. Miller et al. (2019) added, "The central question of both studies was the role played by deliberate practice in the acquisition of expertize [*sic*]. One might think there would be a 'meeting of the minds' when the estimates from their analysis and ours returned such similar results" (p. 289). Miller et al. (2019) seem to suggest here that because their reanalysis yielded an

average deliberate practice-performance correlation very similar to the corresponding correlation in our meta-analysis, we should have found their results acceptable. However, findings from scientific research should be evaluated based on whether they are accurate and interpretable, not on whether they agree with findings from one's own research. On that note, we reiterate that until Miller and colleagues can clarify their methods the results of their reanalysis will only add to the confusion surrounding deliberate practice. As it stands, the results of their reanalysis remain uninterpretable.

The second reanalysis of our dataset was by Ericsson and Harwell (2019). Again, without resolving past inconsistencies in descriptions of deliberate practice in Ericsson and colleagues' writings, Ericsson and Harwell (2019) criticized our use of a general definition of deliberate practice, stating:

> There is no disagreement that the goal of improving performance is one characteristic of deliberate practice, and Ericsson et al. (1993) even wrote that "deliberate practice is a highly structured activity, the explicit goal of which is to improve performance" (p. 368). This sentence was, however, not a definition of deliberate practice any more than the true statement that "a dog is an animal" would imply the inference that "all animals are dogs." (p. 5).

According to this statement, Ericsson et al. (1993) never proposed the general definition of deliberate practice that we used for our meta-analysis. Yet, Lehmann and Ericsson (1997) stated that "Ericsson et al. (1993) *have defined deliberate practice* as a structured activity designed to improve performance" (p. 47, emphasis added). So, to be clear, Lehmann and Ericsson (1997) stated that Ericsson et al. (1993) defined deliberate practice as a structured activity designed to improve performance, but then Ericsson and Harwell (2019) stated that Ericsson et al. (1993) did *not* define deliberate practice as such. Contradictory statements such as these lead to confusion about what the definition of deliberate practice is—and even whether there is a fixed definition at all.

Compounding the issues already discussed, Ericsson and colleagues' standard for evidence concerning deliberate practice has apparently shifted yet again. For their main analysis, Ericsson and Harwell (2019) retained 14 of the 88 studies included in our meta-analysis (Macnamara et al., 2014), coding from these studies eight effect sizes as deliberate practice and six effect sizes as purposeful practice[3]. However, as shown in **Table 2**, seven of the eight effect sizes they coded as deliberate practice were from studies previously rejected by Ericsson (2014a) in his commentary on Macnamara et al.'s meta-analysis for not meeting the criteria for deliberate practice (recall that he rejected 87 of the 88 studies in that commentary). The major reason for this

---

[3]Ericsson and Harwell (2019) state they coded eight effect sizes as deliberate practice and six effect sizes as purposeful practice. They list the deliberate practice effect sizes being from: Baker et al. (2003), Charness et al. (2005), both effect sizes; Ericsson et al. (1993), both effect sizes; Schultetus and Charness (1997), Gobet and Campitelli (2007), and Ruthsatz et al. (2008). And they list the purposeful practice effect sizes being from: Duffy et al. (2004), Tuffiash et al. (2007), Harris (2008), Duckworth et al. (2011), and Maynard et al. (2014). We assume that De Bruin et al. (2007) was the sixth purposeful practice study and that its omission from the latter list is a clerical error by Ericsson and Harwell (2019).

discrepancy is that Ericsson and Harwell (2019) used a more lenient teacher/coach criterion for deliberate practice in their own meta-analysis than Ericsson (2014a) used when evaluating our meta-analysis. More specifically, whereas Ericsson (2014a) required that a study "record a teacher or coach supervising and guiding all or most of the practice" (see Ericsson, 2014b, **Table 2**), Ericsson and Harwell (2019) required only that a study *describe* individualized sessions with a coach or teacher, with no specification of the lower limit on the amount of supervision and guidance. As they explained:

> For coding purposeful versus deliberate practice, we looked for explicit mentions within the original study methods of individualized sessions with coaches or teachers as being included as part of the estimate of solitary practice. If a study did describe individualized instruction sessions as being part of the practice estimate it was considered to be deliberate practice (Ericsson and Harwell, 2019, Supplementary Material, p. 5).

Ericsson and Harwell (2019) do not explain why they made this shift from the stricter criterion that Ericsson (2014a) used when evaluating our meta-analysis to a more lenient teacher/coach criterion for themselves when they reanalyzed our meta-analytic data. Interestingly, they do not even cite Ericsson's (2014a) commentary, although Ericsson (2018) very recently used it as a basis for rejecting the conclusions of the Macnamara et al. (2014) meta-analysis (see also Ericsson and Pool, 2016).

Additional examples of apparent shifts in the standard for evidence concerning deliberate practice can be found in Ericsson and Harwell (2019). For example, in his earlier commentary on Macnamara et al. (2014), Ericsson (2014a) rejected Ruthsatz et al.'s (2008) studies of musicians as a valid study of deliberate practice because there was a restriction of range in the measures (see Ericsson, 2014b, Table 5). However, Ericsson and Harwell (2019) apparently no longer saw this as a problem, because they coded one of these previously rejected studies (Study 2) as deliberate practice and included it in their own meta-analysis.

Setting aside these issues, what would the results of Ericsson and Harwell's (2019) meta-analysis indicate if they were taken at face value? After their correction for measurement error variance (unreliability), the average correlation for the 14 studies that measured either purposeful practice or deliberate practice with performance increased from $r = 0.54$ to $r = 0.78$, indicating that purposeful/deliberate practice[4] explained 61% of the reliable variance in the performance. Although psychometric corrections come with larger standard errors and confidence intervals (Oswald et al., 2015), this point estimate of 61% might be interpreted to support Ericsson et al.'s (1993) central claim that individual differences in domain-relevant performance can largely be accounted for by accumulated amount of practice—at least a weak version of this claim (see Hambrick et al., 2018a).

However, it is critical to examine Ericsson and Harwell's (2019) assumptions about the reliability of the variables, because those

values determined the degree of correction to the correlation (i.e., the lower the assumed reliability, the greater the correction). As previously noted, Ericsson and colleagues previously stated that "self-report practice estimates repeatedly from experts in sports and music have reported test-retest reliabilities at or above 0.80" (Tuffiash et al., 2007, p. 129) and that "[t]he collected reliability of cumulated life-time practice at different test occasions in large samples has typically been found to range between 0.7 and 0.8" (Ericsson, 2012, p. 534). However, for reasons that are unclear, Ericsson and Harwell (2019) used a lower reliability estimate of 0.60 for purposeful/deliberate practice, apparently drawing on different information about the reliability of practice. This raises the important question of whether, by using a lower reliability value, they overcorrected the correlation between purposeful/deliberate practice and performance, and thus overestimated the variance shared between the two variables.

As is the case for most measured variables in psychological research, the reality is that the reliability of practice and performance variables must be estimated and can never be known with certainty; the accuracy of any reliability estimate will vary with the number of participants, the number of items/questions on the instrument, and other factors. Furthermore, the reliability of practice and performance variables may vary depending on factors such as the domain, the skill level and age of the participants, and so on. Thus, an approach we have adopted in our own research is to correct correlations between these variables, based on whatever point estimates of reliability are available, but also to report a sensitivity analysis in which the correlation is corrected under both lower and higher levels of reliability (e.g., Macnamara et al., 2014, Table S1).

Table 3 presents such a sensitivity analysis for the meta-analytic correlations that Ericsson and Harwell (2019) report between purposeful/deliberate practice and performance ($r = 0.54$, $k = 14$), between purposeful practice and performance ($r = 0.51$, $k = 6$), and between deliberate practice and performance ($r = 0.56$, $k = 8$). As can be seen, the strength of the correlations varies depending on the reliability estimates, with larger corrections for lower reliability estimates. For example, if reliability is assumed to be $r_{xx} = 0.80$ for deliberate practice (Ericsson, 2013) and $r_{yy} = 0.80$ for performance, then the corrected correlation of performance with deliberate practice is $r_c = 0.70$, indicating that deliberate practice explains 49% of the reliable between-person variance in performance (rather than 61%). Thus, conclusions will vary considerably depending on what values are used as the reliability estimates in the psychometric correction.

Across the board, the correlations between practice and domain-relevant performance (expertise) in **Table 3** are meaningfully large, from theoretical, statistical, and practical perspectives. At the same time, the correlations vary considerably in magnitude and may lead to different conclusions about the importance of the practice variables. We illustrate this point with reference to the deliberate practice correlations (the bottom set of results in **Table 3**). If deliberate practice explained, for example, 87% of the variance in performance (assuming relatively low reliabilities of $r_{xx} = 0.60$ and $r_{yy} = 0.60$), then a strong version of Ericsson et al.'s (1993) claim that individual differences in

---

[4]Following Ericsson and Harwell's (2019) terminology, when referring to all 14 effect sizes (six coded as purposeful practice and eight coded as deliberate practice) in the meta-analysis, we use the label "purposeful/deliberate practice."

TABLE 3 | Sensitivity analyses for Ericsson and Harwell (2019) meta-analytic correlations.

	Performance r_{yy}			
Purposeful/Deliberate r_{xx}	**0.60**	**0.70**	**0.80**	**0.90**
0.60	0.90 (81.0%)	0.83 (69.4%)	**0.78 (60.8%)**	0.73 (54.0%)
0.70	0.83 (69.4%)	0.77 (59.5%)	0.72 (52.1%)	0.68 (46.3%)
0.80	0.78 (60.8%)	0.72 (52.1%)	0.68 (45.6%)	0.64 (40.5%)
0.90	0.73 (54.0%)	0.68 (46.3%)	0.64 (40.5%)	0.60 (36.0%)

	Performance r_{yy}			
Purposeful r_{xx}	**0.60**	**0.70**	**0.80**	**0.90**
0.60	0.85 (72.3%)	0.79 (61.9%)	**0.74 (54.2%)**	0.69 (48.2%)
0.70	0.79 (61.9%)	0.73 (53.1%)	0.68 (46.4%)	0.64 (41.3%)
0.80	0.74 (54.2%)	0.68 (46.4%)	0.64 (40.6%)	0.60 (36.1%)
0.90	0.69 (48.2%)	0.64 (41.3%)	0.60 (36.1%)	0.57 (32.1%)

	Performance r_{yy}			
Deliberate r_{xx}	**0.60**	**0.70**	**0.80**	**0.90**
0.60	0.93 (87.1%)	0.86 (74.7%)	**0.81 (65.3%)**	0.76 (58.1%)
0.70	0.86 (74.7%)	0.80 (64.0%)	0.75 (56.0%)	0.71 (49.8%)
0.80	0.81 (65.3%)	0.75 (56.0%)	0.70 (49.0%)	0.66 (43.6%)
0.90	0.76 (58.1%)	0.71 (49.8%)	0.66 (43.6%)	0.62 (38.7%)

The table provides corrected correlation coefficients (r_cs) based on Ericsson and Harwell's (2019) reported meta-analytic correlations of $r_{purposeful/deliberate} = 0.54$, $r_{purposeful} = 0.51$, and $r_{deliberate} = 0.56$. Values in parentheses are variance estimates (i.e., $r_c^2 \times 100$). r_{xx}, reliability estimate for practice; r_{yy}, reliability estimate for performance; DP, deliberate practice. The bolded values are the corrected correlations using Ericsson and Harwell's (2019) reliability values. The formula for computing a correlation corrected for unreliability (r_c) is the observed correlation (r_{obs}) divided by the square root of the product of the variables' reliabilities (r_{xx} and r_{yy}): $r_c = r_{obs}/\sqrt{(r_{xx} \times r_{yy})}$ (see Schmidt and Hunter, 1999).

performance can largely be accounted for by deliberate practice would be supported. However, if deliberate practice explained just over half (56%) of the variance (assuming higher reliabilities of $r_{xx} = 0.80$ and $r_{yy} = 0.70$), then a weaker version of the claim would be supported. If deliberate practice explained an even smaller amount of the variance—for example, 39% (assuming even higher reliabilities of $r_{xx} = 0.90$ and $r_{yy} = 0.90$)—then an implication would be that factors other than deliberate practice explain more of the variance in performance than deliberate practice does. These different variance estimates might then lead to different priorities in expertise research. For example, the first scenario (87% of variance explained) might prompt an exclusive focus on training history, whereas the third scenario (39% of variance explained) might prompt a broader focus on multiple determinants of performance differences (e.g., training history, basic abilities). Elsewhere, we have argued that based on extant evidence, research should indeed investigate the role of a wide range of factors in explaining individual differences in expertise (Ullén et al., 2016; Hambrick et al., 2018b).

We further note that, if they were taken at face value, Ericsson and Harwell's (2019) findings would fail to support a central prediction of the latest version of the deliberate practice view. As mentioned earlier, Ericsson and Pool (2016) differentiated deliberate practice from naïve practice and purposeful practice, describing deliberate practice as "the most effective method of all…the gold standard" (p. 85). A straightforward prediction, following from these claims, is that the positive correlation

between deliberate practice and domain-relevant performance should be significantly greater than the positive correlation between purposeful practice and domain-relevant performance (i.e., $r_{deliberate} > r_{purposeful}$). Ericsson and Harwell's (2019) findings do not support this prediction: the average correlation was $r = 0.56$ for deliberate practice and $r = 0.51$ for purposeful practice, a non-significant difference ($p = 0.64$). Ericsson and Harwell (2019) report this finding, but they note only that "practice was positively associated with performance whether it was conducted under the guidance of a coach or teacher" (p. 11). From the standpoint of the distinction between the two types of practice (Ericsson and Pool, 2016), the equally important conclusion would seem to be that there is no evidence from the meta-analysis that deliberate practice has higher validity than purposeful practice in predicting individual differences in domain-relevant performance, as is predicted by Ericsson and colleagues' new framework.

HOW IMPORTANT IS DELIBERATE PRACTICE?

Again, how important is deliberate practice as a predictor of individual differences in expertise? It is somewhat difficult to say given the ambiguity over the definition of deliberate practice, but we can at least summarize evidence from meta-analyses by different groups of researchers that have attempted to

answer this question. Along with the aforementioned reanalysis of chess and music studies (Hambrick et al., 2014b), there have been five formal meta-analyses of the deliberate practice-performance relationship. **Table 4** summarizes the overall result from these meta-analyses (i.e., the overall correlation between deliberate practice and performance), and presents a sensitivity analysis showing variance estimates under different reliability assumptions, from unacceptable to excellent. As shown, across meta-analyses, deliberate practice explains a sizeable amount of the between-person variance in performance. However, we conclude that it is unlikely to be as important as Ericsson and colleagues have hypothesized it is. In nearly all cases, deliberate practice leaves a large amount of reliable variance unexplained, and in most cases, the unexplained variance exceeds the explained variance. Our conclusion, as in the past, is that deliberate practice is an important predictor of individual differences in expertise. However, deliberate practice is it is unlikely to be as important as Ericsson and colleagues have proposed it is.

IS THE DELIBERATE PRACTICE VIEW DEFENSIBLE?

In view of the issues discussed in the preceding pages, it seems reasonable to ask whether the deliberate practice view is scientifically defensible—that is, whether deliberate practice can be conceptualized and tested empirically in a consistent manner by the research community. In his book *The Logic of Scientific Discovery,* the philosopher of science Karl Popper (1959) argued, "In so far as a scientific statement speaks about reality, it must be falsifiable; and in so far as it is not falsifiable, it does not speak about reality" (p. 314). By definition, a theory is unfalsifiable when it cannot be rejected under any circumstances, because it can accommodate any finding. This happens when multiple, contradictory definitions of theoretical concepts are proposed by a theorist, and when the theoretical and operational criteria are kept in a fluid state. Under these conditions, evidence can be rejected or accepted depending on whether it supports the theory.

When a theory becomes unfalsifiable, it ceases to be a *scientific* theory, at least in the Popperian sense. The theory is always "right" and cannot be evaluated against competing theories. Ferguson and Heene (2013) described a theory that has entered this state as being "undead," like a zombie that is technically dead but remains animate. The undead theory "continues in use, having resisted attempts at falsification, ignored disconfirmatory data, negated failed replications through the dubious use of meta-analysis or having simply maintained itself in a fluid state with shifting implicit assumptions such that falsification is not possible" (Ferguson and Heene, 2013, p. 559).

Is the deliberate practice view falsifiable? We will leave it to readers of this article to draw their own conclusions. At the very least, it seems difficult to deny that there are serious problems with the deliberate practice view, as Ericsson and colleagues have presented it over the past 25 years. As we have documented here, Ericsson and colleagues have described deliberate practice in contradictory ways, creating major confusion about the definition of deliberate practice. Furthermore, Ericsson and colleagues'

standard for evidence—the specific criteria that need to be satisfied to use a study to argue for the importance of deliberate practice and even the criteria themselves—has appeared to shift multiple times.

To be sure, it is not only normal, but expected, for a scientist to revise a theory as evidence relevant to that theory accumulates through research. Theory revision is a fundamental part of what the philosopher of science Imre Lakatos (1978) called a "progressive" program of research. In this iterative process, revisions are explicitly acknowledged and clearly explained and justified so that they can be understood and critically evaluated by other scientists. When theory revisions are *not* made in a transparent manner, then in Lakatos' terminology a theory can be endlessly adjusted and readjusted to keep it "alive." The program of research then shifts from "progressive" to "degenerative" (see also Musgrave and Pigden, 2016).

On a related note, the concept of deliberate practice is arguably underspecified in ways that leave open the opportunity for numerous *post hoc* explanations of results. For example, Ericsson and colleagues have stated that deliberate practice requires that the performer have "full concentration" (e.g., Ericsson and Harwell, 2019). However, this is a psychological state that may be impossible to achieve. Would, for example, a person's awareness of the environment (e.g., the temperature) or a fleeting thought (e.g., about an event earlier in the day) mean that they were not fully concentrating on the training task? If yes, then it seems unlikely that a person could ever fulfill this criterion of deliberate practice. Furthermore, we are not aware of a method for objectively determining whether a person has *full* concentration on something. Research approaches from cognitive psychology (e.g., primary-secondary task paradigms) permit no more than relative statements about the degree to which a person is attending to one task (or stimulus) versus some other task(s). Except for subjective self-reports, we are also unaware of attempts by Ericsson and colleagues themselves to measure concentration level during practice. If a criterion for a theoretical construct (e.g., achieving "full concentration") either cannot be achieved or cannot be empirically verified, then imposing that criterion makes the theory unfalsifiable. This sort of flexibility in the deliberate practice view, along with the definitional confusion we have discussed, presents an additional problem for its scientific viability.

Why is it important for researchers to comment publicly when a theory in their research area appears to be degenerating? It is important because a degenerating theory—especially if it is influential—impedes actual progress in an area of research. The value of research to test the theory becomes questionable, because evidence is accepted as valid only if it supports the theory and rejected if it fails to support the theory. In turn, practical recommendations and applications based on the theory will lack a scientific foundation, because even conflicting recommendations and applications can be supported. If conclusions from this research do not have a solid empirical foundation, then recommendations based on the theory may be wasteful, counterproductive, or even harmful.

TABLE 4 | Results of meta-analyses of deliberate practice–performance relationship.

Study	Domain	K	r (95% CI)	% variance	<Acceptable (r_{xx}, r_{yy} = 0.60)	Acceptable (r_{xx}, r_{yy} = 0.70)	Good (r_{xx}, r_{yy} = 0.80)	Excellent (r_{xx}, r_{yy} = 0.90)
Macnamara et al. (2014)	Multiple	157	0.38 (0.33, 0.42)	14.4%	0.63 (40.1%)	0.54 (29.5%)	0.48 (22.6%)	0.42 (17.8%)
Platz et al. (2014)	Music	14	0.61 (0.54, 0.67)	37.2%	1.00 (100%)	0.87 (75.9%)	0.76 (58.1%)	0.68 (45.9%)
Macnamara et al. (2016b)	Sports	63	0.45 (0.36, 0.53)	20.3%	0.75 (56.3%)	0.64 (41.3%)	0.56 (31.6%)	0.50 (25%)
Miller et al. (2019)	Multiple	70	0.40 (0.34, 0.46)	16.0%	0.67 (44.4%)	0.57 (32.7%)	0.50 (25%)	0.44 (19.8%)
Ericsson and Harwell (2019)	Multiple	8	0.56 NR	31.4%	0.93 (87.1%)	0.80 (64.0%)	0.70 (49%)	0.62 (38.7%)

k, number of effect sizes. r, meta-analytic correlation between deliberate practice and performance.% variance, percentage of variance in performance explained by deliberate practice (i.e., r^2). For the corrected correlations, values in parentheses adjacent to correlations are variance estimates ($r_c^2 \times 100$). NR, not reported.

To be certain, we do not think that the concept of deliberate practice should be abandoned. Deliberate practice is a vital area of research in psychological science and other fields. However, given the issues we have discussed at length in this article, we do believe that it would be highly beneficial to expertise researchers (and scientists from other research areas interested in expertise) for proponents of the deliberate practice view to fully address and resolve the apparent inconsistencies and shifts in the definition of and standard of evidence for deliberate practice that we have documented here—and which raise serious concerns about the viability of the deliberate practice view as a scientific theory. This would allow other researchers to empirically evaluate the importance of deliberate practice as a predictor of individual differences in expertise, both in individual studies and meta-analyses, and to compare its predictive validity to that of other factors (e.g., general aptitudes, basic capacities) and the conditions of practice (e.g., spacing of practice sessions, type of feedback).

To this point, in the following sections, we summarize key evidence for the role of a diverse range of factors in explaining individual differences in expertise, and then discuss in broad terms what we believe are fruitful directions for future research to develop comprehensive models of expertise.

TOWARD A MULTIFACTORIAL MODEL OF EXPERTISE

What might explain individual differences in expertise, beyond any contribution of deliberate practice? We direct readers to recent theoretical/review articles in which we discuss this issue at length (Hambrick et al., 2016; Ullén et al., 2016; Macnamara and Hambrick, 2020). Here, we briefly summarize evidence concerning three major classes of factors.

Developmental Factors

The question of when specialized training should commence in a person's life is the subject of a longstanding debate in the field of expertise. The *early specialization view* argues that the earlier the training can begin, the better. The logic of this view is straightforward: Because it is both physically and psychologically taxing, a person can engage in only a few hours of deliberate practice a day (around 4 h on average; Ericsson et al., 1993) without burnout and/or injury. Therefore, the individual who begins training at a relatively late age (e.g.,

age 12) can never catch up to the individual who begins training earlier (e.g., age 6). However, in a meta-analysis of sports studies with samples representing a wide range of skill, we found no evidence for an earlier average starting age for high-skill athletes relative to lower-skill athletes. Furthermore, research suggests that the highest (elite) levels of sports performance are associated with a later starting age, combined with participation in a diverse range of sports in adolescence. For example, Güllich (2017) compared 83 international medalists (Olympic/World championship) to 83 non-medalists matched on sport, age, and gender. Up to age 18, the medalists had, on average, accumulated significantly *fewer* hours of organized training/practice in their main sport (by 948 h) than the non-medalists. Moreover, the average starting age was *later* by approximately a year-and-a-half for the medalists compared to the non-medalists. One possible explanation for this finding is that a starting age that is too early increases the risk for injury and/or burnout. Another possible explanation is that starting later allows for more early diverse experiences, increasing the likelihood that the individual will find a sport that is a good match to his or her profile of performance-relevant traits (Güllich, 2017).

Experiential Factors

A central tenet of the deliberate practice view is that deliberate practice is more predictive of individual differences in expertise than other forms of experience, such as work and play. As Boot and Ericsson (2013) explained, "Ericsson and colleagues. . .make a critical distinction between domain-related activities of work, play, and deliberate practice, and claim that the amount of accumulated time engaged in deliberate practice activities is the primary predictor of exceptional performance" (p. 146). The available evidence does not appear to support this claim. As already mentioned, if Ericsson and Harwell's (2019) findings are taken at face value, they reveal that deliberate practice, although claimed to be the "gold standard" for improving performance (Ericsson and Pool, 2016), is not a significantly stronger predictor of individual differences in expertise than mere purposeful practice (i.e., $r_{deliberate}$ = 0.56 vs. $r_{purposeful}$ = 0.51, difference in *r*s non-significant). Furthermore, measures of deliberate practice have not always been found to be stronger predictors of individual differences in expertise than measures seeming to meet the definition of "work" and "play." For example, in a study of insurance salespeople, Sonnentag and Kleine (2000) found

that correlations between sales performance and measures of deliberate practice ($r_{current} = 0.21$ and $r_{accumulated} = 0.13$) were not stronger than the correlation between sales performance and a measure fitting the description of work for this domain—the number of cases handled ($r = 0.37$; see Hambrick et al., 2016, for other examples).

Evidence further suggests that diverse forms of experience are important as well, especially in the early stages of training. For example, in Güllich's (2017) study comparing the 83 international medalists and 83 non-medalists, he not only found that the medalists had accumulated significantly less main-sport practice than their less-accomplished counterparts during childhood/adolescence, but also that the medalists had accumulated significantly more experience with other sports during this period (see also Güllich et al., 2017).

Ability Factors

Research has firmly established that cognitive ability explains a statistically and practically significant amount of the variability in people's acquisition of complex skills (Hambrick et al., 2019; also see Ackerman, 1987, for a review of early studies). That is, people higher in cognitive ability learn complex skills more readily and rapidly than people lower in cognitive ability. For example, in a study of music training, participants with little or no experience playing music completed tests of cognitive ability, music aptitude, and growth mindset, and then they were given instruction in playing a simple piece of music on the piano (Burgoyne et al., 2019). Higher-ability participants showed a greater rate of learning than lower-ability participants, with a general intelligence factor explaining approximately 30% of the individual differences in learning rate.

Ericsson (2014d) has theorized that general cognitive ability is important initially in acquiring complex skills, but its predictive power diminishes as domain-specific skills and knowledge are acquired, stating:

> For individuals who have acquired cognitive structures that support a high level of performance the expert performance framework predicts that these acquired cognitive structures will directly mediate superior performance and thus diminishing correlations between general cognitive ability and domain-specific performance (p. 84).

For complex tasks of interest to expertise researchers, evidence for this claim, which we termed the *circumvention-of-limits hypothesis* (Hambrick and Meinz, 2011), is weak and inconsistent. In a recent review (Hambrick et al., 2019), we searched through approximately 1,300 articles and identified 15 studies in the domains of games, music, sports, science, medicine/surgery, and aviation relevant to this hypothesis. Of the 15 studies, only three yielded any evidence supportive of the circumvention-of-limits hypothesis. Moreover, methodological limitations (e.g., small *N*s, measures with unknown or unreported reliability) precluded any strong conclusions from those few studies. Providing what might be considered the strongest evidence for the hypothesis, one of these three studies that seem to support the circumvention-of-limits hypothesis was a meta-analysis of chess studies (Burgoyne

et al., 2016; see also Burgoyne et al., 2018, corrigendum). As determined by a moderator test, fluid intelligence correlated significantly more strongly with chess rating in lower-skill chess players (avg. $r = 0.32$) than in higher-skill chess players (avg. $r = 0.14$). However, it is important to note that skill level was highly confounded with age (i.e., lower-ability samples were youth, whereas higher-ability samples were adults), limiting the strength of the evidence in support of the circumvention-of-limits hypothesis.

We also note that results that have sometimes been used to argue that the influence of general cognitive ability on expertise diminishes with increasing skill do not warrant this conclusion. For example, Ericsson (2014d) pointed to results by Ruthsatz et al. (2008) as support for this hypothesis. Ruthsatz et al. (2008) found that a measure of general cognitive ability (Raven's Progressive Matrices score) correlated positively and significantly with musical accomplishment in high school band members ($r = 0.25$, $p < 0.05$), but not in university music majors ($r = 0.24$) or conservatory students ($r = 0.12$). However, the critical question is not whether the lower-skill group correlation is statistically significant while the higher-skill group correlations are not. Rather, it is whether the former correlation and the latter correlations are significantly *different from each other*, as determined by the appropriate statistical test. As it happens, in the Ruthsatz et al. (2008) study, the correlations are not significantly different from each other (all *z* tests for differences in correlations are statistically non-significant). Thus, the results of Ruthsatz et al.'s (2008) study *fail* to support the hypothesis that ability-performance correlations diminish with increasing skill.

We also reviewed evidence relevant to the circumvention-of-limits hypothesis from the job performance literature, and here the evidence is more consistent and interpretable. General cognitive ability is regarded as the single best predictor of job training performance, and of subsequent job performance (Schmidt and Hunter, 2004; Schmidt, 2014). Higher ability people tend to learn job skills more rapidly and to a higher level than lower ability people, and in turn have greater success on the job. Furthermore, although the validity of general cognitive ability for job performance may drop somewhat as a function of job experience initially, it appears to remain a statistically significant predictor even at high levels of job experience. For example, in a study of 10,088 military personnel across 31 jobs, the correlation between a measure of general cognitive ability (the AFQT score) and hands-on job performance decreased as a function of job experience from 1 to 2 years of job experience ($r = 0.34$ to $r = 0.21$; *z* test for difference = 3.60, $p < 0.001$), but then stabilized and remained statistically and practically significant (see **Figure 3**). In their own review of the job performance literature, Reeve and Bonaccio (2011) concluded that "although validities might degrade somewhat over long intervals, we found no evidence to suggest that they degrade appreciably, thereby retaining practically useful levels of validity over very long intervals" (p. 269).

Genetic and Environmental Influences

Research in the field of behavioral genetics has demonstrated that both genetic and environmental variance across individuals

contribute to the total variance in a wide range of behavioral outcomes (Turkheimer, 2000), including ability factors that have been found to correlate with measures of expertise. The extent of the genetic contribution is captured by the heritability statistic (h^2), an estimate of the proportion (0 to 1) of the total variance in a trait that can be attributed to genetic (non-environmental) variance within a sample of individuals (Plomin et al., 2008). Because some of these factors correlate with expertise, it stands to reason that both genetic and environmental variance may also contribute to the total variance in expertise. Furthermore, basic abilities and characteristics that may predict individual differences in expertise have also been observed to be substantially heritable, including drawing ability (Arden et al., 2014), music aptitude (Ullén et al., 2014; see Mosing et al., 2018, for a review), and maximal oxygen consumption in athletic performance (VO_{2max}; Schutte et al., 2016).
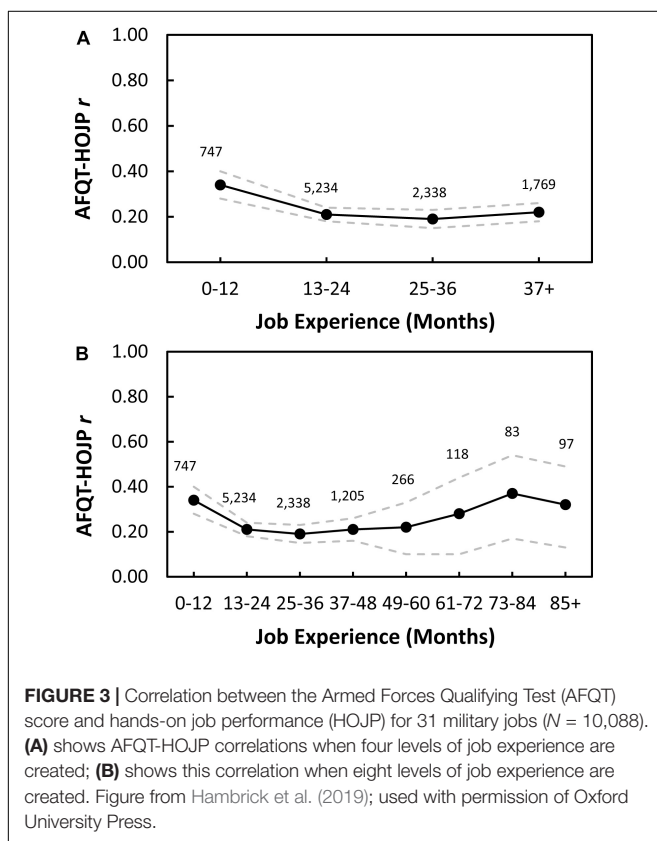
At the same time, no psychological trait is 100% heritable (Turkheimer, 2000), and even the most heritable psychological trait will have a sizeable environmental component. For example, heritability estimates for measures of general cognitive ability are typically in the 50 to 70% range in samples drawn from developed countries (e.g., Tucker-Drob and Bates, 2016), with the remaining variance (as much as 50%) explained by shared and/or non-shared environmental factors. This means that correlations between a measure of some trait (e.g., general cognitive ability) and a measure of expertise could be driven by the genetic variance or the environmental variance in the trait measure, or by both

types of variances. In other words, the finding that a measure of a heritable trait correlates with expertise is only *consistent with the possibility* that genetic variance is a component of individual differences in expertise.

It is also critical to note that genes and environments cannot generally be assumed to be uncorrelated across people. Rather, across people, genetically influenced factors may contribute to variance in the environments which people seek out and are exposed to. This is the idea of *gene-environment correlation,* or *r*GE (Plomin et al., 1977). For example, just as children who are tall might be more interested in playing basketball and more likely to be selected to play on basketball teams than children who are shorter, those with a high level of music aptitude may be more likely to take up, be selected for, and persist in music than those with a lower level of this aptitude. Consistent with this sort of speculation, there is now evidence to indicate that the propensity to practice in a domain is substantially heritable. In a large twin study, Mosing, Ullén, and colleagues found an average heritability of around 50% for accumulated amount of music practice (Mosing et al., 2014; see also Hambrick and Tucker-Drob, 2015). A possible explanation for this finding is that music aptitude, as well as more general ability and non-ability factors, differentially predispose people to engaging in music practice.

Genetic factors and environmental factors may not only correlate with one another; they may interact in influencing behavioral outcomes—what is known as *gene-environment interaction,* or G × E. G × E occurs when a genetically influenced factor moderates (increases or decreases) the effect of an environmental factor on an outcome. As one example of G × E, analyzing data from the National Merit Twin Study, Hambrick and Tucker-Drob (2015) found that heritability of a music accomplishment variable was 0.43 for individuals who reported engaging in music practice, versus 0.01 for those who did not. (This result was not due to range restriction, as there was still variability in music accomplishment among participants who reported not practicing). This finding suggests that music practice may activate genetic factors that vary across people.

Four additional points concerning the potential contribution of genetic factors to individual differences in expertise are important to note here. First, even if a measure of expertise is found to be heritable, this in *no way* implies that training is unnecessary to develop a high level of expertise, or that training is beneficial to only some people. Training is necessary and essential for developing a high level of expertise in a domain, and except when a condition rules out some type of training (e.g., a visual training regimen for a person who is blind), anyone would be expected to benefit from proper training. Second, heritability does not imply immutability. For example, in adults, height is highly heritable and relatively fixed, whereas weight is similarly heritable but can be modified through an environmental intervention—namely, dieting. Third, environmental interventions that change individual differences will also change heritability. For example, if an environmental intervention were introduced that allowed nearly everyone to reach about the same level of skill in some task, heritability would be expected to decrease. In a similar manner, heritability can differ across populations (e.g., in a developing country

FIGURE 3 | Correlation between the Armed Forces Qualifying Test (AFQT) score and hands-on job performance (HOJP) for 31 military jobs (*N* = 10,088). **(A)** shows AFQT-HOJP correlations when four levels of job experience are created; **(B)** shows this correlation when eight levels of job experience are created. Figure from Hambrick et al. (2019); used with permission of Oxford University Press.

vs. a non-developing country; Tucker-Drob and Bates, 2016). Fourth, and finally, it is safe to assume that to the degree that expertise is heritable, this would reflect variation in a great many genetic variants (i.e., single nucleotide polymorphisms, or SNPs), meaning there is no "expertise gene" in any domain. As Chabris and colleagues noted, "A typical human behavioral trait is associated with very many genetic variants, each of which accounts for a very small percentage of the behavioral variability" (Chabris et al., 2015, p. 305). Research is uncovering genetic variants that may contribute to individual differences in expertise, but it is highly unrealistic to expect that any one of these factors will account for a large amount of the variance in expertise.

Putting It All Together

Taken together, evidence suggests that individual differences in expertise arise from influences of multiple factors. This includes training and other forms of domain-relevant experience, as well as developmental factors (e.g., age of starting training), ability factors (e.g., aptitudes), non-ability factors (e.g., personality traits), and background factors (e.g., opportunity to engage in training). In recent articles, we have proposed the *multifactorial gene-environment interaction model* (MGIM) to describe how these factors relate to one another and to serve as a guide for future research on expertise (Ullén et al., 2016). As illustrated in **Figure 4**, the MGIM assumes that expertise arises from influences of both domain-general and domain-specific factors, which are assumed to be influenced by both genetic and environmental factors. The model further assumes that task/situational factors may moderate the influence of these factors on expertise (i.e., domain-relevant performance).

THE PATH AHEAD IN EXPERTISE RESEARCH

Ericsson and colleagues' deliberate practice view has had a monumental impact on the field of expertise, and is important in the history of psychology, more generally. However, in our assessment, it is not clear that the deliberate practice view is defensible as a scientific theory. As described here in some detail, Ericsson and colleagues have defined deliberate practice in inconsistent ways (see **Figure 2**) and the standard for evidence concerning deliberate practice has appeared to shift multiple times (see **Tables 1**, **2**). These issues present problems for the empirical testability (i.e., falsifiability) of the deliberate practice view.

Embracing tenets of the *open science movement,* we believe that the path ahead is for expertise researchers to work together to develop testable theories that take into account a much wider range of potentially relevant causal constructs than have often been considered in previous research, and to use rigorous empirical methods to evaluate these theories. The open science movement promotes normative values aimed at increasing accuracy, openness, and fairness in scientific research and scholarship (see a special issue of the *Journal of Expertise* devoted to open science in expertise research; McAbee and Macnamara, 2020).

Nearly 80 years ago, the sociologist of science Robert Merton (1942) described four scientific "imperatives" that capture the values of the open science movement. First, *universalism*: scientific validity is independent of the status of the people conducting the research; evidence should be evaluated based on its own merits rather than the status or prominence of the person reporting the evidence. Second, *communism:* a theory does not belong to the theorist, it belongs to the field. The theorist has no greater right to the theory once it is made public than any other scientist. Third, *disinterestedness*: scientists should perform research to increase understanding of some phenomenon rather to advance self-interests, whatever they may be. Finally, *organized skepticism*: a field should scrutinize claims based on empirical evidence.

In the wake of the replication crisis in the social sciences, many measures have been proposed to increase the reproducibility of research findings in psychological science and to accelerate progress in research (see Munafò et al., 2017). Preregistration of study designs, primary outcomes, and data analysis plans can help safeguard against *post hoc* interpretation of data. Improved methodological training can help researchers avoid pitfalls in designing studies (e.g., omitting critical control conditions) and in data analysis (e.g., misinterpreting *p*-values). Collaboration can facilitate collection of large samples and help to ensure that multiple theoretical perspectives are considered in study design. The time is also ripe for a preregistered "adversarial collaboration"—a study in which researchers with differing views agree on an empirical test to resolve a theoretical dispute that is designed to provide a fair test of both views (see Kahneman and Klein, 2009; for a recent example, see Doherty et al., 2019). No less so than in any area of psychological research, we believe that open science will accelerate progress toward greater understanding of the nature and origins of expertise and expert performance.

Recommendations for Expertise Research

We make four general recommendations for conducting expertise research, which are based on best research practices in differential psychology (see Ackerman and Hambrick, 2020). First, after selecting a domain for the research, the researchers should seek to assess a wide range of potentially relevant causal factors (Simonton, 1999). Whatever the domain, it will be critical to measure key environmental factors, including various types of training and factors related to the opportunity to engage in these activities. However, drawing on the vast literature in differential psychology, it will be equally critical to include measures of basic abilities, capacities, dispositions, and other psychological traits that may affect performance directly, or indirectly through training. Only then can the relative and joint contributions of these factors to individual differences in expertise be evaluated.

The second recommendation is that multiple measures be used to index each of the hypothesized constructs. It is axiomatic in the psychological methods literature that virtually no observed measure (or indicator) is "construct pure." That is, a score collected by an instrument (test, questionnaire, etc.) designed
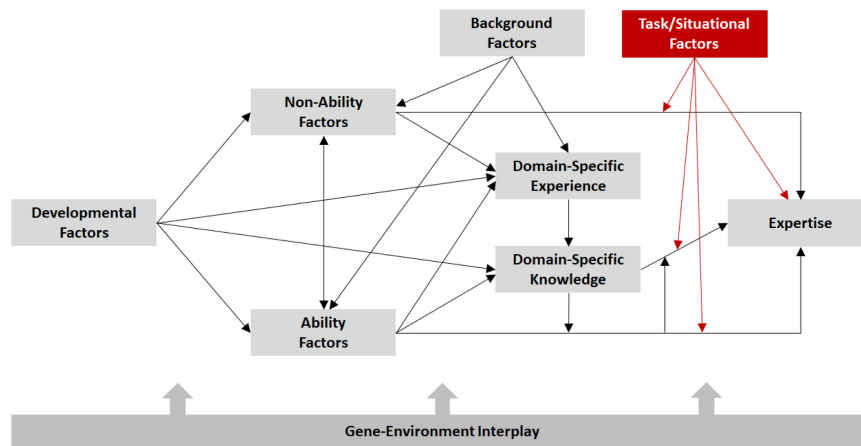
FIGURE 4 | Multifactorial model of expertise. From Hambrick et al. (2018c). Used with permission of Routledge.

to measure a given hypothetical construct may reflect that construct to some degree, but it will certainly reflect other, construct-irrelevant factors, such as participants' familiarity with a particular method of assessment (e.g., test format) and psychological states that may affect their responding (e.g., sleep deprivation and motivation). There is no perfect way to deal with this problem, but when multiple measures of a construct are obtained, it becomes possible to use data-analytic techniques (viz., structural equation modeling) that are explicitly designed to deal with this issue by allowing researchers to model latent variables that are closer to theoretical constructs of interest than observed variables are.

The third recommendation is that the sample of participants from the targeted domain should ideally represent a wide range of performance rather than extreme groups. As we have noted elsewhere (Hambrick et al., 2019), categories such as "novice" and "expert" are not naturally occurring—they are groups of performers created based on ultimately arbitrary cuts on performance scores. Accordingly, scientific research on expertise should endeavor to explain the full range of performance differences within different domains rather than differences between artificial groups of performers, and also continuities and discontinuities across this range (see Bliese and Lang, 2016).

The final recommendation is for expertise researchers to begin large-scale longitudinal studies. Longitudinal studies are expensive and, by their very nature, time-consuming. At the same time, they are common in psychology. For example, there are longitudinal studies in the area of cognitive aging that have been running for many decades, such as the Seattle Longitudinal Study (Schaie, 2005). There is also precedent for longitudinal studies of expertise, including Schneider and colleagues' important longitudinal study of tennis skill, which included two future World No. 1 players (Schneider et al., 1993). Although expensive, labor-intensive, and time-consuming, multi-site longitudinal studies of expertise will provide for much stronger conclusions concerning the underpinnings of expertise.

We also offer one more specific recommendation for future research on expertise. The goal of Friedlander and Fine's (2016)

grounded expertise components approach (GECA) is to identify predictors of individual differences in expertise in a theoretically neutral manner to minimize bias in findings from research concerning the relative importance of one class of factor versus another (e.g., training vs. basic abilities). The approach begins with exploration: administering a survey to a large sample of performers in domains, with questions about their level of engagement in training activities, as well as education, interests, hobbies, careers, accomplishments, and other characteristics that may be informative about potential predictors of expertise (i.e., performance differences). The survey data are then analyzed to identify a candidate set of potential predictors of expertise. Finally, a study is conducted to estimate the relative contributions of the factors to the prediction of individual differences in expertise. Programmatic research can then proceed on this basis.

A major goal of the GECA is to identify activities that consistently correlate to a practically and statistically significant degree with measures of expertise. These activities may differ across domains. More specifically, for some domains, the activities may meet the ultimate criteria for deliberate practice, whereas for other domains, they may include unstructured activities that do not fit the definition of deliberate practice, purposeful practice, or even naïve practice. Furthermore, within a domain, there may be multiple "routes" to developing expertise. That is, one performer may achieve a given level of expertise by engaging in one set of activities, whereas another performer may achieve the same level of expertise by engaging in a different set of activities. As an illustration, Berliner (1994) found that jazz musicians emphasized the importance of intentionally unstructured "jam" sessions for developing improvisational skill, and noted that "[s]trongly motivated students commonly learned musical instruments without formal instruction by synthesizing bits of knowledge from commercial method books, other young performers, and their own experimentation" (Location 738).

In short, the search for activities using Friedlander and Fine's (2016) GECA holds great potential to shed light on the role of different types of training in explaining expertise across a wide range of domains, including not only domains

traditionally studied in expertise research (e.g., chess, sports, and classical music) but also those that have received relatively little attention in research.

CONCLUSION

For decades, the field of expertise has focused on environmental factors as the major determinants of individual differences in expertise, whereas genetically influenced factors are assumed to play a relatively unimportant role, if any role at all (Ericsson et al., 1993; Ericsson, 2007a). Environmental factors certainly are important to consider in investigating the origins of individual differences in expertise, but a comprehensive scientific theory of expertise must take into account genetically influenced factors as well, including basic abilities and capacities ("talent"). At a more general level, we argue that it is time—*past* time—for the nature vs. nurture debate to be over in the field of expertise, as it has been in most areas of psychological research for decades (Turkheimer, 2000). Embracing the idea that expertise can be best understood as a product of gene-environment interplay (nature *and* nurture) will, as Plomin (2018) recently observed, move the field ahead and integrate it with the life sciences. At a practical level, findings from this research will provide a scientific foundation for principles and procedures designed and implemented to accelerate people's acquisition of complex skills across a wide range of domains and elevate the performance of individuals, organizations, and societies.

REFLECTION

As we were editing the page proofs for this article, we received the sad news of Anders Ericsson's passing. It is difficult to imagine the field of expertise without Anders—he was a pioneer. But his ideas will live on and continue to inspire scholarship and debate that will lead to greater understanding of the subject about which he was so passionate. We hope that Anders found our work as stimulating as we found his. He forced us to think critically about our most basic assumptions concerning expertise, and to try to put our best case for our perspective forward. We are in his debt, and extend our sincere condolences to his family, friends, students, and colleagues.

AUTHOR CONTRIBUTIONS

DH was the primary author and conceived the article. BM and FO provided extensive comments and edits on multiple drafts. All authors contributed to the article and approved the submitted version.

REFERENCES

Ackerman, P. L. (1987). Individual differences in skill learning: an integration of psychometric and information processing perspectives. *Psychol. Bullet.* 102, 3–27. doi: 10.1037/0033-2909.102.1.3

Ackerman, P. L., and Hambrick, D. Z. (2020). A primer on assessing intelligence in laboratory studies. *Intelligence.* 80:101440. doi: 10.1016/j.intell.2020.101440

Arden, R., Traskowski, M., Garfield, V., and Plomin, R. (2014). Genes influence young children's human figure drawings and their association with intelligence a decade later. *Psychol. Sci.* 25, 1843–1850.

Baker, J., Côté, J., and Abernethy, B. (2003). Learning from the experts: practice activities of expert decision makers in sport. *Res. Q. Exerc. Sport* 74, 342–347. doi: 10.1080/02701367.2003.10609101

Berliner, P. (1994). *Thinking in Jazz: The Infinite Art of Improvisation*. Chicago, IL: University of Chicago Press.

Bilalić, M., McLeod, P., and Gobet, F. (2007). Does chess need intelligence? A study with young chess players. *Intelligence* 35, 457–470. doi: 10.1016/j.intell.2006.09.005

Bliese, P. D., and Lang, J. W. B. (2016). Understanding relative and absolute change in discontinuous growth models: Coding alternatives and implications for hypothesis testing. *Organ. Res. Methods* 19, 562–592. doi: 10.1177/1094428116633502

Boot, W. R., and Ericsson, K. A. (2013). "Expertise," in *The Oxford Handbook of Cognitive Engineering*, eds J. D. Lee and A. Kirlik (Oxford: Oxford University Press), 143–158.

Burgoyne, A. P., Harris, L. J., and Hambrick, D. Z. (2019). Predicting skill acquisition in music: the role of general intelligence, music aptitude, and mindset. *Intelligence* 76:101383. doi: 10.1016/j.intell.2019.101383

Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., and Hambrick, D. Z. (2016). The relationship between cognitive ability and chess skill: a comprehensive meta-analysis. *Intelligence* 59, 72–83. doi: 10.1016/j.intell.2016.08.002

Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., and Hambrick, D. Z. (2018). Corrigendum: the relationship between cognitive ability and chess skill: a comprehensive meta-analysis. *Intelligence* 71, 92–96. doi: 10.1016/j.intell.2018.08.004

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., and Laibson, D. I. (2015). The fourth law of behavior genetics. *Curr. Direct. Psychol. Sci.* 24, 304–312. doi: 10.1177/0963721415580430

Charness, N., Krampe, R. Th, and Mayr, U. (1996). "The role of practice and coaching in entrepreneurial skill domains: an international comparison of life-span chess skill acquisition," in *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games*, ed. K. A. Ericsson (Mahwah, NJ: Erlbaum), 51–80.

Charness, N., Tuffiash, M., Krampe, R., Reingold, E., and Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Appl. Cogn. Psychol.* 19, 151–165. doi: 10.1002/acp.1106

Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159.

Coyle, T. (2009). *The Talent Code. Greatness Isn't Born. It's Grown. Here's How*. New York, NY: Bantam.

Deakin, J. M., and Cobley, S. (2003). "A search for deliberate practice: Expert performance in sports: Advances in research on sport expertise," in *Expert Performance in Sports: Advances in Research on Sport Expertise*, eds J. L. Starkes and K. A. Ericsson (Champaign, IL: Human Kinetics), 115–136.

De Bruin, A. B., Smits, H., Rikers, R. M. J. P., and Schmidt, H. G. (2008). Deliberate practice predicts performance over time in adolescent chess players and drop-outs: a linear mixed models analysis. *Br. J. Psychol.* 99, 473–497. doi: 10.1348/000712608x295631

De Bruin, A. B. H., Rikers, R. M. J. P., and Schmidt, H. G. (2007). The influence of achievement motivation and chess-specific motivation on deliberate practice. *J. Sport Exerc. Psychol.* 29, 561–583. doi: 10.1123/jsep.29.5.561

Dearani, J. A., Gold, M., Leibovich, B. C., Ericsson, K. A., et al. (2017). The role of imaging, deliberate practice, structure, and improvisation in approaching surgical perfection. *J. Thora. Card. Surgery* 154, 1329–1336. doi: 10.1016/j.jtcvs.2017.04.045

Doherty, J. M., Belletier, C., Rhodes, S., Jaroslawska, A., Barrouillet, P., Camos, V., et al. (2019). Dual-task costs in working memory: an adversarial collaboration. *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 1529–1551. doi: 10.1037/xlm0000668

Duckworth, A. L., Kirby, T. A., Tsukayama, E., Berstein, H., and Ericsson, K. A. (2011). Deliberate practice spells success why grittier competitors triumph at the national spelling bee. *Soc. Psychol. Pers. Sci.* 2, 174–181. doi: 10.1177/1948550610385872

Duffy, L. J., Baluch, B., and Ericsson, K. A. (2004). Dart performance as a function of facets of practice amongst professional and amateur men and women players. *Int. J. Sport Psychol.* 35, 232–245.

Ericsson, K. A. (1998). The scientific study of expert levels of performance: general implications for optimal learning and creativity. *High Ability Stud.* 9, 75–100. doi: 10.1080/1359813980090106

Ericsson, K. A. (2005). Recent advances in expertise research: a commentary on the contributions to the special issue. *Appl. Cogn. Psychol.* 19, 233–241. doi: 10.1002/acp.1111

Ericsson, K. A. (2006). "The influence of experience and deliberate practice on the development of superior expert performance," in *The Cambridge Handbook of Expertise and Expert Performance*, eds K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman (New York, NY: Cambridge University Press), 683–703.

Ericsson, K. A. (2007a). Deliberate practice and the modifiability of body and mind: toward a science of the structure and acquisition of expert and elite performance. *Int. J. Sport Psychol.* 38, 4–34.

Ericsson, K. A. (2007b). An expert-performance perspective of research on medical expertise: the study of clinical performance. *Med. Educ.* 41, 1124–1130. doi: 10.1111/j.1365-2923.2007.02946.x

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: a general overview. *Acad. Emerg. Med.* 15, 988–994. doi: 10.1111/j.1553-2712.2008.00227.x

Ericsson, K. A. (2012). The Danger of Delegating Education to Journalists: Why the APS Observer Needs Peer Review When Summarizing New Scientific Developments. Available online at: https://psy.fsu.edu/faculty/ericsson/ericsson.hp.html (accessed June 6, 2020).

Ericsson, K. A. (2013). Training history, deliberate practice and elite sports performance: an analysis in response to Tucker and Collins review—what makes champions? *Br. J. Sports Med.* 47, 533–535. doi: 10.1136/bjsports-2012-091767

Ericsson, K. A. (2014a). *Challenges for the Estimation of an Upper-Bound on Relations Between Accumulated Deliberate Practice and the Associated Performance of Novices and Experts: Comments on Macnemara [sic], Hambrick, and Oswald's (2014) Published Meta Analysis. Unpublished Commentary.* Available online at: https://psy.fsu.edu/faculty/ericsson/ericsson.hp.html (accessed June 6, 2020).

Ericsson, K. A. (2014b). *Supplemental online materials for "A Challenge to Estimates of an Upper-Bound on Relations Between Accumulated Deliberate Practice and the Associated Performance in Domains of Expertise: Comments on Macnemara [sic], Hambrick, and Oswald's (2014) Published Meta-Analysis." Unpublished Supplemental Materials to Unpublished Commentary.* Available online at: https://psy.fsu.edu/faculty/ericsson/ericsson.hp.html (accessed June 6, 2020).

Ericsson, K. A. (2014c). "Adaptive expertise and cognitive readiness: a perspective from the expert-performance approach," in *Teaching and Measuring Cognitive Readiness*, eds H. F. O'Neil, R. S. Perez, and E. Baker (Boston, MA: Springer), 179–197. doi: 10.1007/978-1-4614-7579-8_10

Ericsson, K. A. (2014d). Why expert performance is special and cannot be extrapolated from studies of performance in the general population: a response to criticisms. *Intelligence* 45, 81–103. doi: 10.1016/j.intell.2013.12.001

Ericsson, K. A. (2015). Acquisition and maintenance of medical expertise: a perspective from the expert-performance approach with deliberate practice. *Acad. Med.* 90, 1471–1486. doi: 10.1097/acm.0000000000000939

Ericsson, K. A. (2016). Summing up hours of any type of practice versus identifying optimal practice activities: Commentary on Macnamara, Moreau, & Hambrick (2016). *Perspect. Psychol. Sci.* 11, 351–354. doi: 10.1177/1745691616635600

Ericsson, K. A. (2018). "The differential influence of experience, practice, and deliberate practice on the development of superior individual performance of experts," in *The Cambridge Handbook of Expertise and Expert Performance*, 2nd Edn, eds K. A. Ericsson, R. R. Hoffmann, A. Kozbelt, and A. M. Williams (New York, NY: Cambridge University Press), 745–769. doi: 10.1017/9781316480748.038

Ericsson, K. A. (2020). Towards a science of the acquisition of expert performance in sports: Clarifying the differences between deliberate practice and other types of practice. *J. Sports Sci.* 38, 159–176. doi: 10.1080/02640414.2019.1688618

Ericsson, K. A., Chase, W. G., and Faloon, S. (1980). Acquisition of a memory skill. *Science* 208, 1181–1182.

Ericsson, K. A., and Harwell, K. W. (2019). Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: Why the original definition matters and recommendations for future research. *Front. Psychol.* 10:2396. doi: 10.3389/fpsyg.2019.02396

Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychol. Rev.* 100, 363–406. doi: 10.1037/0033-295x.100.3.363

Ericsson, K. A., Perez, R. S., Eccles, D. W., Lang, L., Baker, E. L., Bransford, J. D., et al. (2009). "The measurement and development of professional performance: An introduction to the topic and a background to the design and origin of this book," in *Development of Professional Expertise: Toward Measurement of Expert Performance and Design of Optimal Learning Environments*, ed. K. A. Ericsson (Cambridge University Press), 1–24. doi: 10.1017/CBO9780511609817.001

Ericsson, K. A., and Pool, R. (2016). *Peak: Secrets From the New Science of Expertise.* New York, NY: Eamon Dolan/Houghton Mifflin Harcourt.

Ericsson, K. A., Prietula, M. J., and Cokely, E. T. (2007). The making of an expert. *Harv. Bus. Rev.* 85, 114–121.

Ericsson, K. A., and Towne, T. J. (2013). "Experts and their superior performance," in *Oxford Library of Psychology. The Oxford Handbook of Cognitive Psychology*, ed. D. Reisberg (Oxford: Oxford University Press), 886–901.

Ferguson, C. J., and Heene, M. (2013). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspect. Psychol. Sci.* 7, 555–561. doi: 10.1177/1745691612459059

Feynman, R. (1963). *The Meaning of it All: Thoughts of a Citizen Scientist.* Available online at: http://movies2.nytimes.com/books/first/f/feynman-meaning.html (accessed June 6, 2020).

Friedlander, K. J., and Fine, P. A. (2016). The grounded expertise components approach in the novel area of cryptic crossword solving. *Front. Psychol.* 7:567. doi: 10.3389/fpsyg.2016.00567

Gladwell, M. (2008). *Outliers: The Story of Success.* New York, NY: Little, Brown.

Gobet, F., and Campitelli, G. (2007). The role of domain-specific practice, handedness, and starting age in chess. *Dev. Psychol.* 43, 159–172. doi: 10.1037/0012-1649.43.1.159

Gottfredson, L. S. (2003). Dissecting practical intelligence theory: Its claims and evidence. *Intelligence* 31, 343–397. doi: 10.1016/s0160-2896(02)00085-5

Güllich, A. (2017). International medallists' and nonmedallists' developmental sport activities—a matched-pairs analysis. *Motor Behav. Exp. Perform.* 35, 2281–2288. doi: 10.1080/02640414.2016.1265662

Güllich, A., Kovar, P., Zart, S., and Reimann, A. (2017). Sports activities differentiating match-play improvement in elite youth footballers – a 2-year longitudinal study. *J. Sports Sci.* 35, 207–215. doi: 10.1080/02640414.2016.1161206

Hambrick, D. Z., Altmann, E. M., Oswald, F. L., Gobet, F., and Campitelli, G. (2014a). Accounting for expert performance: the devil is in the details. *Intelligence* 45, 112–114. doi: 10.1016/j.intell.2014.01.007

Hambrick, D. Z., Burgoyne, A. P., Campitelli, G., and Macnamara, B. N. (2018a). "Working memory, thinking, and expertise," in *International Handbook of Thinking and Reasoning*, eds L. J. Ball and V. A. Thompson (New York, NY: Psychology Press), 268–288.

Hambrick, D. Z., Burgoyne, A. P., Macnamara, B. N., and Ullén, F. (2018b). Toward a multifactorial model of expertise: beyond born versus made. *Ann. N. Y. Acad. Sci.* 1423, 284–295. doi: 10.1111/nyas.13586

Hambrick, D. Z., Burgoyne, A. P., and Oswald, F. L. (2019). "Domain-general models of expertise: The role of cognitive ability," in *The Oxford Handbook of Expertise: Research and Application*, eds P. Ward, J. M. Schraagen, J. Gore, and E. Roth (Oxford: Oxford University Press), 56–84.

Hambrick, D. Z., Campitelli, G., and Macnamara, B. N. (2018C). "Introduction: A brief history on the science of expertise and overview of the book," in *The Science of Expertise: Behavioral, Neural, and Genetic Approaches to Complex Skill*, ed. D. Z. Hambrick, G. Campitelli, and B. N. Macnamara (New York, NY: Routledge).

Hambrick, D. Z., and Macnamara, B. N. (in press). "Toward a science of expertise: obstacles, implications, and applications. To appear," in *Conceptual Frameworks*

for Giftedness and Talent Development*, eds T. Cross and P. Olszewski-Kubilius (Dallas, TX: Prufrock Academic Press).

Hambrick, D. Z., and Macnamara, B. N. (2019a). More confusion about deliberate practice: Commentary on Miller et al. (2019). *High Ability Stud.* 30, 291–294. doi: 10.1080/13598139.2019.1607723

Hambrick, D. Z., Macnamara, B. N., Campitelli, G., Ullén, F., and Mosing, M. A. (2016). "Beyond born versus made: a new look at expertise," in *The Psychology of Learning and Motivation*, Vol. 64, ed. B. Ross (Amsterdam: Elsevier), 1–55. doi: 10.1016/bs.plm.2015.09.001

Hambrick, D. Z., and Meinz, E. J. (2011). Limits on the predictive power of domain-specific experience and knowledge in skilled performance. *Curr. Direct. Psychol. Sci.* 20, 275–279. doi: 10.1177/0963721411422061

Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., and Campitelli, G. (2014b). Deliberate practice: Is that all it takes to become an expert? *Intelligence* 45, 34–45. doi: 10.1016/j.intell.2013.04.001

Hambrick, D. Z., and Tucker-Drob, E. M. (2015). The genetics of music accomplishment: Evidence for gene-environment correlation and interaction. *Psychon. Bull. Rev.* 22, 112–120. doi: 10.3758/s13423-014-0671-9

Harris, K. R. (2008). *Deliberate Practice, Mental Representations, and Skilled Performance in Bowling. Unpublished Dissertation*. Tallahassee, FL: Florida State University.

Helsen, W., Starkes, J., and Hodges, N. (1998). Team sports and the theory of deliberate practice. *J. Sport Exerc. Psychol.* 20, 12–34. doi: 10.1123/jsep.20.1.12

Hodges, N. J., and Starkes, J. L. (1996). Wrestling with the nature of expertise: a sport specific test of Ericsson, Krampe, and Tesch-Römer's (1993) theory of deliberate practice. *Int. J. Sport Psychol.* 27, 400–424.

Kahneman, D., and Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *Am. Psychol.* 64, 515–526. doi: 10.1037/a0016755

Keith, N., and Ericsson, K. A. (2007). A deliberate practice account of typing proficiency in everyday typists. *J. Exp. Psychol.* 13, 135–145. doi: 10.1037/1076-898x.13.3.135

Krampe, R., and Ericsson, K. A. (1996). Maintaining excellence: deliberate practice and elite performance in young and older pianists. *J. Exp. Psychol.* 125, 331–359. doi: 10.1037/0096-3445.125.4.331

Lakatos, I. (1978). *The Methodology of Scientific Research Programmes: Volume 1: Philosophical Papers*. Cambridge: Cambridge University Press.

Lehmann, A. C., and Ericsson, K. A. (1997). Research on expert performance and deliberate practice: implications for the education of amateur musicians and music students. *Psychomusicology* 16, 40–58. doi: 10.1037/h0094068

Macnamara, B. N., and Hambrick, D. Z. (2020). Toward a cumulative science of expertise: commentary on Moxley, Ericsson, & Tuffiash (2017). *Psychol. Res.* doi: 10.1007/s00426-020-01307-2

Macnamara, B. N., Hambrick, D. Z., and Moreau, D. (2016a). How important is deliberate practice? Reply to Ericsson (2016). *Perspect. Psychol. Sci.* 11, 355–358. doi: 10.1177/1745691616635614

Macnamara, B. N., Hambrick, D. Z., and Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, professions, and education: a meta-analysis. *Psychol. Sci.* 25, 1608–1618. doi: 10.1177/0956797614535810

Macnamara, B. N., Hambrick, D. Z., and Oswald, F. L. (2018). Corrigendum: deliberate practice and performance in music, games, sports, education, and professions: a meta-analysis. *Psychol. Sci.* 29, 1202–1204. doi: 10.1177/0956797618769891

Macnamara, B. N., Moreau, D., and Hambrick, D. Z. (2016b). The relationship between deliberate practice and performance in sports: a meta-analysis. *Perspect. Psychol. Sci.* 11, 333–350. doi: 10.1177/1745691616635591

Maynard, D., Hambrick, D. Z., and Meinz, E. M. (2014). *How important is deliberate practice for explaining expertise in sports? The case of bowling.* Unpublished manuscript.

McAbee, S. T., and Macnamara, B. N. (2020). Introduction to the special issue on open research practices and reproducibility in expertise research. *J. Exp.* 2, 197–198.

Merton, R. (1942). A note on science and democracy. *J. Legal Polit. Sociol.* 1, 115–126.

Miller, S. D., Chow, D., Wampold, B., and Hubble, M. A. (2019). "You say tomatoe, I say tomawto": the importance of deliberate practice for improved performance. *High Ability Stud.* 30, 288–290. doi: 10.1080/13598139.2019.1652542

Miller, S. D., Chow, D., Wampold, B. E., Hubble, M. A., Del Re, A. C., Maeschalck, C., et al. (2018). To be or not to be (an expert)? Revisiting the role of deliberate practice in improving performance. *High Ability Stud.* doi: 10.1080/13598139.2018.1519410

Molenaar, P. C. M., Huizenga, H. M., and Nesselroade, J. R. (2003). "The relationship between the structure of interindividual and intraindividual variability: a theoretical and empirical vindication of developmental systems theory," in *Understanding Human Development: Dialogues With Lifespan Psychology*, eds U. M. Staudinger and U. Lindenberger (Norwell, MA: Kluwer), 339–360. doi: 10.1007/978-1-4615-0357-6_15

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., and Ullén, F. (2014). Practice does not make perfect: No causal effect of music practice on music ability. *Psychol. Sci.* 25, 1795–1803. doi: 10.1177/0956797614541990

Mosing, M. A., Peretz, I., and Ullén, F. (2018). "Genetic influences on music expertise," in *The Science of Expertise: Behavioral, Neural, and Genetic Approaches to Complex Skill*, eds D. Z. Hambrick, G. Campitelli, and B. N. Macnamara (New York, NY: Routledge), 272–282. doi: 10.4324/9781315113371-16

Moxley, J. H., Ericsson, K. A., and Tuffiash, M. (2019). Gender differences in SCRABBLE performance and associated engagement in purposeful practice activities. *Psychol. Res.* 83, 1147–1167. doi: 10.1007/s00426-017-0905-3

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie, et al. (2017). A manifesto for reproducible science. *Nat. Hum. Behav.* 1:21.

Musgrave, A., and Pigden, C. (2016). "Imre lakatos," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Stanford, CA: Metaphysics Research Lab, Stanford University).

Oswald, F. L., Ercan, S., McAbee, S. T., Ock, J., and Shaw, A. (2015). Imperfect corrections or correct imperfections? Psychometric corrections in meta-analysis. *Indust. Organ. Psychol.* 8, 1–4.

Platz, F., Kopiez, R., Lehmann, A. C., and Wolf, A. (2014). The influence of deliberate practice on musical achievement: a meta-analysis. *Front. Psychol.* 5:646.

Plomin (2018). "Foreward," in *The Science of Expertise: Behavioral, Neural, and Genetic Approaches to Complex Skill*, eds D. Z. Hambrick, G. Campitelli, and B. N. Macnamara (New York, NY: Routledge), xiv–xvii.

Plomin, R. J., DeFries, C., and Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol. Bull.* 84, 309. doi: 10.1037/0033-2909.84.2.309

Plomin, R. J., DeFries, C., McClearn, G. E., and McGuffin, P. (2008). *Behavioral Genetics*. New York, NY: Worth Publishers.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. London: Routledge Classics.

Reeve, C. L., and Bonaccio, S. (2011). On the myth and the reality of the temporal validity degradation of general mental ability test scores. *Intelligence* 39, 255–272. doi: 10.1016/j.intell.2011.06.009

Ruthsatz, J., Detterman, D. K., Griscom, W. S., and Cirullo, B. A. (2008). Becoming an expert in the musical domain: It takes more than just practice. *Intelligence* 36, 330–338. doi: 10.1016/j.intell.2007.08.003

Schaie, K. W. (2005). What can we learn from longitudinal studies of adult development? *Res. Hum. Dev.* 2, 133–158. doi: 10.1207/s15427617rhd0203_4

Schmidt, F. L. (2014). A general theoretical integrative model of individual differences in interests, abilities, personality traits, and academic and occupational achievement: a commentary on four recent articles. *Perspect. Psychol. Sci.* 9, 211–218. doi: 10.1177/1745691613518074

Schmidt, F. L., and Hunter, J. (1999). Theory testing and measurement error. *Intelligence* 27, 183–198. doi: 10.1016/s0160-2896(99)00024-0

Schmidt, F. L., and Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *J. Pers. Soc. Psychol.* 86, 162–173. doi: 10.1037/0022-3514.86.1.162

Schneider, K., Bös, K., and Rieder, H. (1993). "Performance prediction in young top athletes," in *Aufmerksamkeit und Energetisierung. Facetten con Konzentration und Leistung*, eds J. Beckmann, H. Strang, and E. Hahn (Göttingen: Hogrefe), 277–299.

Schultetus, S., and Charness, N. (1997). "Fencing data. unpublished data. cited in Deakin, J. M., and Cobley, S. (2003). A search for deliberate practice," in *Expert Performance in Sports: Advances in Research on Sport Expertise*, eds, J. L. Starkes and K. A. Ericsson (Champaign, IL: Human Kinetics), 115–136.

Schutte, N. M., Nederend, I., Hudziak, J. J., Bartels, M., and de Geus, E. J. C. (2016). Twin-sibling study and meta-analysis on the heritability of maximal oxygen consumption. *Physiol. Genom.* 48, 210–219. doi: 10.1152/physiolgenomics. 00117.2015

Simonton, D. K. (1999). Talent and its development: an emergenic and epigenetic model. *Psychol. Rev.* 106, 435–457. doi: 10.1037/0033-295x.106.3.435

Sonnentag, S., and Kleine, B. M. (2000). Deliberate practice at work: a study with insurance agents. *J. Occup. Organ. Psychol.* 73, 87–102. doi: 10.1348/096317900166895

Sternberg, R. J., Wagner, R. K., Williams, W. M., and Horvath, J. A. (1995). Testing common sense. *Am. Psychol.* 50, 912–927.

Tucker-Drob, E. M., and Bates, T. C. (2016). Large cross-national differences in gene x socioeconomic status interaction on intelligence. *Psychol. Sci.* 27, 138–149. doi: 10.1177/0956797615612727

Tuffiash, M., Roring, R., and Ericsson, K. A. (2007). Expert performance in Scrabble: Implications for the study of the structure and acquisition of complex skills. *J. Exp. Psychol. Appl.* 13, 124–134. doi: 10.1037/1076-898x.13.3.124

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Curr. Direct. Psychol. Sci.* 9, 160–164. doi: 10.1111/1467-8721.00084

Ullén, F., Hambrick, D. Z., and Mosing, M. A. (2016). Rethinking expertise: a multifactorial gene–environment interaction model of expert performance. *Psychol. Bull.* 142, 427–446. doi: 10.1037/bul0000033

Ullén, F., Mosing, M. A., Holm, L., and Madison, G. (2014). Psychometric properties and heritability of an online test for musicality, the swedish musical discrimination test. *Pers. Individ. Differ.* 63, 87–93. doi: 10.1016/j.paid.2014. 01.057
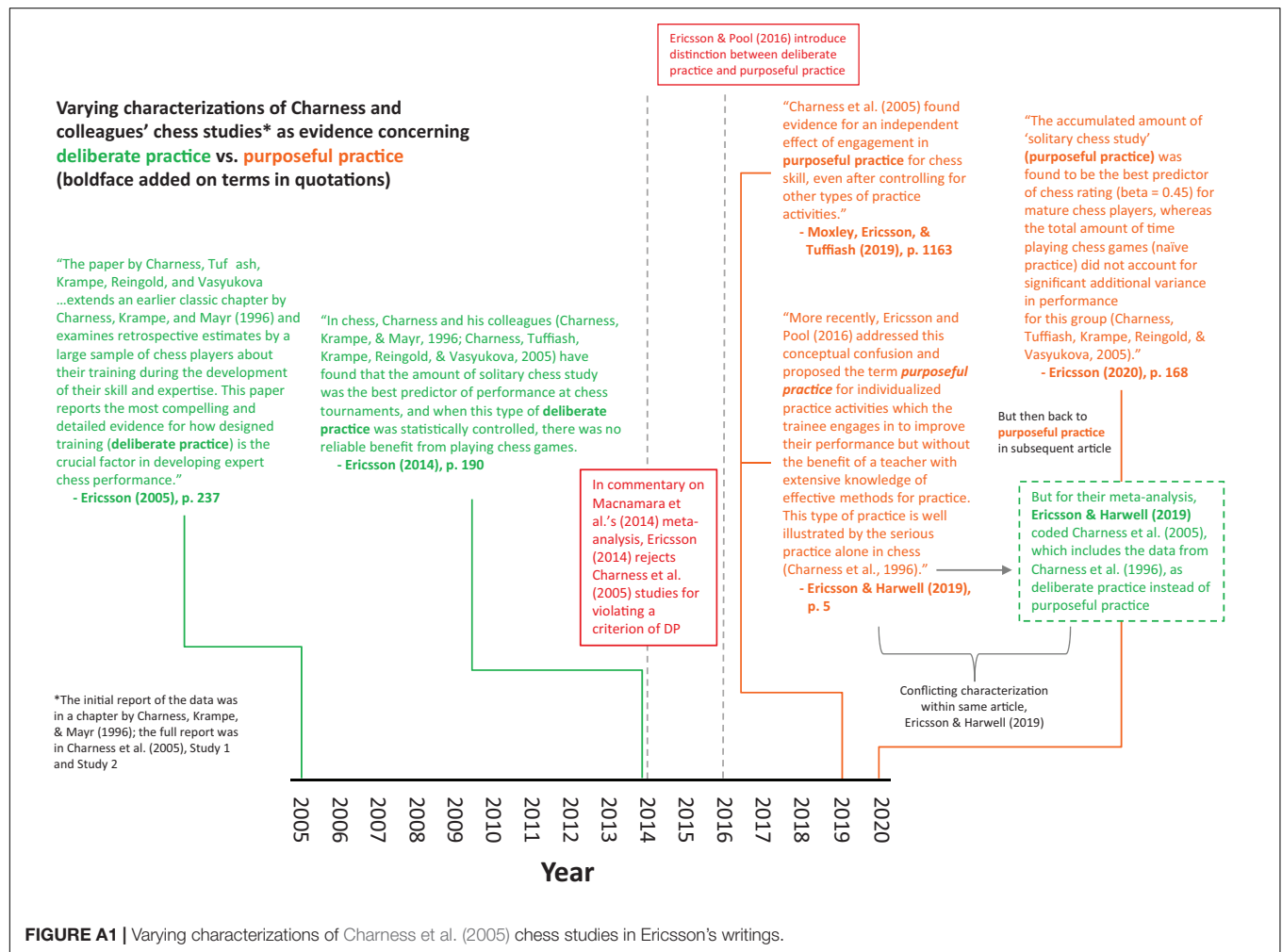
APPENDIX

FIGURE A1 | Varying characterizations of Charness et al. (2005) chess studies in Ericsson's writings.

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