MODELING INDIVIDUAL DIFFERENCES IN PERCEPTUAL DECISION MAKING

EDITED BY : Joseph W. Houpt, Cheng-Ta Yang and James T. Townsend

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**

MODELING INDIVIDUAL DIFFERENCES IN PERCEPTUAL DECISION MAKING

Topic Editors:
Joseph W. Houpt, Wright State University, USA
Cheng-Ta Yang, National Cheng Kung University, Taiwan
James T. Townsend, Indiana University Bloomington, USA

To deal with the abundant amount of information in the environment in order to achieve our goals, human beings adopt a strategy to accumulate some information and filter out other information to ultimately make decisions. Since the development of cognitive science in the 1960s, researchers have been interested in understanding how human beings process and accumulate information for decision-making. Researchers have conducted extensive behavioral studies and applied a wide range of modeling tools to study human behavior in simple-detection tasks and two-choice decision tasks (e.g., discrimination, classification).

In general, researchers often assume that the manner in which information is processed for decision-making is invariant across individuals given a particular experimental context. Independent variables, including speed-accuracy instructions, stimulus properties (i.e., intensity), and characteristics of the participants (i.e., aging, cognitive ability) are assumed to affect the parameters in a model (i.e., speed of information accumulation, response bias) but not the way that participants process information (e.g., the order of information processing). Given these assumptions, much modeling has been accomplished based on the grouped data, rather than the individual data. However, a growing number of studies have demonstrated that there were individual differences in the perceptual decision process. In the same task context, different groups of the participants may process information in different manners. The capacity and architecture of the decision mechanism were found to vary across individuals, implying that humans' decision strategies can vary depending on the context to maximize their performance.

In this special issue, we focused on a particular subset of cognitive models, particularly accumulator models, multinomial processing trees and systems factorial technology (SFT) as applied to perceptual decision making. The motivation for the focus on perceptual decision-making is threefold. Empirical studies of perception have grown out of a history of making a large number of observations for each individual so as to achieve precise estimates of each individual's performance. This type of data, rather than a small number of observations per individual, is most amenable to achieving precision in individual-level and group-level cognitive modeling. Second, the interaction between the acquisition of perceptual information and the decisions based on that information (to the extent that those processes are distinguishable) offers rich data for scientific exploration. Finally, there is an increasing interest in the practical application of individual variation in perceptual ability, whether to inform perceptual training and expertise, or to guide personnel decisions. Although these practical applications are beyond the scope of this issue, we hope that the research presented herein may serve as the foundation for future endeavors in that domain.

The contributions of Fific, Chechile, and Zhang et al. each represents fundamental advances in individual difference modeling that will be useful in future research on perceptual decision-making. Chechile's contribution argues for the viability of multinomial processing tree as a more informative model of perceptual decision-making than the more traditional signal detection approach. In the article, he demonstrates the use of a hierarchical application of the model to account for both group-level and individual-level performance. Fific's article gives an overview of SFT, a framework that is applied in many of the articles in this special issue, and contributes new analyses and details on the application of SFT. This new contributions include demonstrations of SFT's advantages for studying individual differences and group-level analyses, a consequence of the fact that the diagnostic SFT statistics are estimated at the individual level rather than from data aggregated across subjects. Next, Zhang et al. demonstrate a new approach to fitting a particular type of accumulator model to individual subject data: diffusion models with flexible, time-varying decision boundaries.

Yang & Wu's contribution includes an example of the dangers of averaging data across individuals: important patterns of performance at the individual level can be obscured when averaging across participants. In their contribution, they argue that an empirical phenomenon known as the category variability effect, which is important for distinguishing among models of perceptual categorization, may be common but often overlooked due to averaging across participants. By applying individual-level modeling, they found clear evidence for the category variability effect in some, but not all, individuals.

Blunden et al. also contribute to knowledge on individual differences in perceptual categorization. They explore the effect of categorization training on perceptual discrimination among faces generated by combining four different base faces. By applying multiple quantitative approaches (General Recognition Theory, multidimensional scaling, SFT and the logical-rules framework) they are able to classify individual participants based on whether they use parallel self-terminating processes and what type interactions occur between the perception of each stimulus dimension.

Yu et al. and Endres et al. investigate the connections between processing capacity and working memory capacity. Yu et al. systematically explore the connection between SFT capacity measures in three different target detection tasks and an operations span task score (a commonly used measure of working memory capacity). Endres et al. develop a new task to examine the relative effects of loading either visual-spatial items or phonetic items into working memory on visual processing capacity as a function of operation span task scores.

Houpt et al. examine variation in visual processing capacity as a function of a different construct, reading ability and particularly dyslexia diagnoses. Building on earlier work measuring word-superiority type effects across words, pseudowords and non-words with SFT, they demonstrate how various subpopulations within those diagnosed with dyslexia might be identified.

Nunez et al. explore the connection between cognitive models and EEG measures of attention. They find that individual differences in task performance are explained by parametric variation in an evidence accumulation model. Furthermore, the parametric differences across individuals, particularly in the evidence accumulation rates, are highly correlated with the EEG measure of attentional control.

Chang & Yang's article examines the connection between cultural differences, particularly individual thinking style, and visual processing capacity. Using both accumulator models and SFT, they find that individuals that have higher "middle-way thinking" scores (roughly, the tendency to consider many alternative perspectives) had higher visual processing capacity as well.

Table of Contents

Editorial: Modeling Individual Differences in Perceptual Decision Making

Joseph W. Houpt[1], Cheng-Ta Yang[2] and James T. Townsend[3]*

[1] Department of Psychology, Wright State University, Dayton, OH, USA, [2] Department of Psychology, National Cheng Kung University, Tainan, Taiwan, [3] Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IL, USA

The Editorial on the Research Topic

Modeling Individual Differences in Perceptual Decision Making

Researchers have been interested in how human beings accumulate and process information for decision-making since the development of experimental psychology in the late nineteenth century and then its renaissance in cognitive science in the 1960s. Whereas psychometrics and test theory, which also got their start in the nineteenth century have made individual differences the foundation of their fields, the study of cognitive processes has traditionally, and over many decades, assumed that the manner in which information is processed for decision-making is invariant across individuals given a particular experimental context.

The typical approach in cognitive psychology has assumed that individual variation affects perceptual processing parametrically (e.g., rate of information accumulation, response bias), but not structurally (e.g., the order of information processing). For example, when using information in working memory, some individuals may be faster, but it is assumed that all individuals use the information in the same manner. With that assumption, the usual practice of developing models is based on grouped data, rather than the individual data.

However, a growing number of studies have demonstrated systematic individual differences in perceptual decision-making. These individual differences can be reflected in both parametric variation corresponding to characteristics of the participants (e.g., working memory span) and structural differences (i.e., in the same task context, different individuals search across visual-spatial information and phonetic information in sequence while others search in parallel). Hence, we as researchers need more complex modeling tools than traditional linear models with null-hypothesis testing to investigate the influences of task, context, and individual differences as well as the potential for interactions among these factors.

In this special issue, we focused on a particular subset of cognitive models that explicitly allow for both structural and parametric variation across individuals, particularly multinomial processing trees, and systems factorial technology (SFT) applied to perceptual decision-making. The motivation for the focus on perceptual decision-making is threefold. Empirical studies of perception have grown out of a history of making a large number of observations for each individual so as to achieve precise estimates of each individual's performance. This type of data, rather than a small number of observations per individual, is most amenable to achieving precision in individual-level and group-level cognitive modeling. Second, the interaction between the acquisition of perceptual information and the decisions based on that information (to the extent that those processes are distinguishable) offers rich data for scientific exploration.

Finally, there is an increasing interest in the practical application of individual variation in perceptual ability, whether to inform perceptual training and expertise, or to guide personnel

decisions. That is, some research trajectories seem to be in the process of synthesis of contemporary cognitive psychology with the above mentioned psychometrics tradition.

The contributions of Fific et al., Chechile et al. and Zhang et al. represent fundamental theoretical advances in individual difference modeling.

Chechile's contribution argues for the viability of multinomial processing trees as a more informative model of perceptual decision-making than the traditional signal detection approach. His contribution includes a hierarchical application of the multinomial processing tree model. As signal detection theory is a fundamental tool in perceptual decision-making research, the potential information gain from a multinomial processing tree model could be significant across the field. Furthermore, the hierarchical modeling approach accommodates group-level analysis of individual differences.

Fific's article gives an overview of SFT, a framework that is applied in many of the articles in this special issue, and contributes new analyses and details on the application of SFT. These new contributions include demonstrations of SFT's advantages for studying individual differences and group-level analyses, a consequence of the fact that the diagnostic SFT statistics are estimated at the individual level rather than from data aggregated across subjects. This allows for the empirical investigation of structural individual differences in perceptual decision-making.

Next, Zhang et al. demonstrate a new approach to fitting a particular type of accumulator model to individual subject data: Diffusion models with flexible, time-varying decision boundaries. This approach can reveal individual differences in accumulating evidence toward a decision bound.

Yang and Wu's contribution includes an example of the dangers of averaging data across individuals: Important patterns of performance at the individual level can be obscured when averaging across participants. In their contribution, they argue persuasively that an empirical phenomenon known as the "category variability effect," which is important for distinguishing among models of perceptual categorization, may be common but often overlooked due to averaging across participants. By applying individual-level modeling, they found clear evidence for the category variability effect in some, but not all, individuals.

Blunden et al. explore individual differences in the effect of categorization training on perceptual discrimination among faces. They use faces generated by combining four different base faces. By applying multiple quantitative approaches (general recognition theory, multi-dimensional scaling, SFT, and the logical-rules framework) they were able to classify individual participants based on whether they use parallel self-terminating processes and what types of interactions occur between the perceptions of each stimulus dimension. This approach leads to a better understanding of individual perceptual categorization training for faces and demonstrates an improved method for exploring individual differences in perceptual categorization in general.

Yu et al. systematically explore the connection between individual variation in SFT capacity measures in three different redundant-target detection tasks and an operations span task score (a commonly used measure of working memory capacity).

They find that only the SFT capacity in an audiovisual detection task was positively correlated to the working memory capacity, suggesting that perceptual processing for audiovisual information and the executive function in working memory share similar cognitive resources. The contribution of this study is to demonstrate the use of individual-level modeling to further the understanding of the theoretical links between different levels of capacity measures.

Like Yu et al., Endres et al. focus on connections between SFT capacity measures and individual differences in working memory using parametric models. They develop a new task to examine the relative effects of loading either visual-spatial items or phonetic items into working memory on visual processing capacity as a function of operation span task scores. Standard analyses of response times and accuracy indicated clear differences between individuals with high working memory span and those with low working memory span. Despite this difference, there was no evidence of a difference across groups in the efficiency with which individuals were able to combine the two sources of information. By applying models to the study of individual differences in working memory, Endres et al. better isolate the behavioral locus of working memory deficiencies, which can in turn be used to better understand the mechanism by which working memory varies across individuals.

Houpt et al. examine variation in visual processing capacity as a function of a different construct, reading ability, and particularly dyslexia diagnoses. Building on earlier work measuring word-superiority-type effects across words, pseudowords, and non-words with SFT, they demonstrate how various subpopulations within those diagnosed with dyslexia might be identified. Even with clear differences between those with dyslexia and the control participants on standard diagnostic measures, some of the participants with dyslexia exhibited word superiority effects that were not distinguishable from control participants while others with dyslexia were clearly different. These data inform the current debate about the heterogeneity of dyslexia and indicate that the model-based measure of word superiority may offer additional diagnostic insights.

Nunez et al. explore the connection between cognitive models and EEG measures of attention. They find that individual differences in task performance are explained by parametric variation in an evidence accumulation model. Furthermore, the parametric differences across individuals, particularly in the evidence accumulation rates, are highly correlated with the EEG measure of attentional control.

Chang and Yang's article examines the connection between cultural differences, particularly individual thinking style, and visual processing capacity. Using both accumulator models and SFT, they find that individuals who have higher "middle-way thinking" scores (roughly, the tendency to consider many alternative perspectives) had higher visual processing capacity as well. These findings provide a reasonable cognitive mechanical account for the behavior of high middle-way thinkers. The contribution of this work is to demonstrate that the application of individual-level modeling to study the culture-sensitive behavior.

In sum, many of the "laws" of human thought and behavior garnered over the past one hundred thirty-seven years since Wilhelm Wundt epochally established his laboratory in

Leipzig, are based on grouped means and, given the increasing appearance of individual differences in even elementary perceptual, cognitive, and motor tasks, they will likely come under increased scrutiny. Together, the articles gathered in this special issue, demonstrate both the need for models of individual differences in perceptual decision-making and the strength of applying such models. We believe this imposing body of research offers a significant advance toward having the necessary tools for studying the joint influences of task, context, and individual differences on perception.

AUTHOR CONTRIBUTIONS

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

FUNDING

frontiers in PSYCHOLOGY

Double jeopardy in inferring cognitive processes

Mario Fific *

Department of Psychology, Grand Valley State University, Allendale, MI, USA

Inferences we make about underlying cognitive processes can be jeopardized in two ways due to problematic forms of aggregation. First, averaging across individuals is typically considered a very useful tool for removing random variability. The threat is that averaging across subjects leads to averaging across different cognitive strategies, thus harming our inferences. The second threat comes from the construction of inadequate research designs possessing a low diagnostic accuracy of cognitive processes. For that reason we introduced the systems factorial technology (SFT), which has primarily been designed to make inferences about underlying processing order (serial, parallel, coactive), stopping rule (terminating, exhaustive), and process dependency. SFT proposes that the minimal research design complexity to learn about n number of cognitive processes should be equal to 2^n. In addition, SFT proposes that (a) each cognitive process should be controlled by a separate experimental factor, and (b) The saliency levels of all factors should be combined in a full factorial design. In the current study, the author cross combined the levels of jeopardies in a 2×2 analysis, leading to four different analysis conditions. The results indicate a decline in the diagnostic accuracy of inferences made about cognitive processes due to the presence of each jeopardy in isolation and when combined. The results warrant the development of more individual subject analyses and the utilization of full-factorial (SFT) experimental designs.

Keywords: individual differences, averaging across subjects, factorial design, inferring cognitive processes, SFT

INTRODUCTION

The central goal of cognitive modeling is to learn the underlying structure of mental processes, which essentially take place in a black box. Learning about cognitive mechanisms inside the box is challenging, as many mental processes are not consciously accessible. Therefore, a reverse engineering procedure has been used to learn about these cognitive processes: an input in the form of stimuli variations is carefully selected and fed to a black box, and an output in the form of response behavior is observed. Knowing a device's blueprint, a good engineer can control input, examine output, and identify the organization of the device's subsystems.

Unlike engineers, cognitive psychologists have to infer a blueprint from the input-output relationship. Take for example two proposed models of short-term-memory (STM) search. In a serial system the memory items are scanned in a sequential fashion. In a parallel system items are scanned simultaneously. To differentiate between these two models scientists have used memory load (number of memorized items 1–6) as the input and the response time (RT) as the output. In theory, the serial and parallel systems would make different predictions for the relationship between memory load and RT. A serial system (of limited capacity) would predict linearly increasing RT as a function of memory load size. A parallel system (but of unlimited capacity) would predict a flat RT as a function of memory load size. Thus, to learn the blueprint of the STM black box a scientist would use an input consisting of a varying number of items to be memorized, then would record the output response times. Then she would compare the results with the predictions of the serial and parallel

systems and decide which is the most likely model supported by the results.

However, it is not quite that simple. One of the main obstacles to unveiling the content of a black box is noisy output. A novice scientist would be (unpleasantly) surprised to learn that hardly any two human response times are of a similar value, even when the exact same task is repeated. To illustrate, here are four recorded responses times belonging to a single subject who repeated the same STM task: 455, 245, 300, and 801 ms. The output response measures varied widely although the input to the black box had a fixed memory load size (one memorized item). The question is: Why would the same set of processes used to process one item show variability when repeated? One answer, is that RTs may vary so much because the cognitive processes, operating in a black box, are not deterministic and can naturally vary in their duration over time. Another source of measurement error can arise from individual subject differences. RT measures will vary across different subjects even when the same task is used. Although subjects might employ the same set of processes in a task, their responses will vary because the processes of interest may rely on cognitive components that process at different rates.

All of these random response fluctuations are known as measurement errors, in which each observation is considered a random departure of the response from the true value associated with the process of interest.

The question remains: Is it possible to remove the measurement error from the output variable? The most robust method for doing so is the averaging tool (data aggregation) on an increased

sample size. Scientists of all different disciplines have used the averaging tool to calculate precise distances between stellar bodies, plot brain activity, compare smokers with non-smokers or simply to determine the longevity of a 9-volt battery. Fueled by the central limit theorem and the law of large numbers, the sample's average value converges to the true (expected) value. The averaging tool would replace the aforementioned noisy data set with a single sample mean RT value. The simplicity and effectiveness of the averaging tool has justified its widespread use in research. However, this simplicity does not guarantee that the data averaging tool is free of conditional assumptions.

When using the averaging tool to make correct inferences about the organization of cognitive processes[1], researchers must be aware of an unfortunate double jeopardy.

DOUBLE JEOPARDY

The first way correct inferences can be jeopardized is when observed data is averaged across subjects. Free of random variability, the averaged data should show the true results pertaining to the underlying processes. But before choosing to average data a scientist should be aware of the necessary conditional assumption: that all subjects use an identical set of cognitive operations[2]. The validity of the data averaging tool depends heavily on this assumption. Take for example a group of subjects who are all serial STM processors but each subject scans an item with a different processing rate (that is constant across different memory loads). The individual results would show a set of linearly increasing response times (RTs) as a function of memory load size, each with a different slope value. Such a slope value would indicate a measure of processing rate per one item in a serial system (Sternberg, 1966). When the averaging tool is used across subjects, the resulting function would also be linearly increasing with a slope value that is the average of the individual slope values. Thus, that averaged result is an unbiased indicator of the underlying processes, presumably showing the true parameter value of an item's serial processing rate, and not a value of random individual variations.

Several major cognitive theories have advocated the idea that humans use identical cognitive operations. Such theories include the conventionally adopted ideal observer approach, or the concept of a rational decision maker. However, that hypothesis is not tenable, and it is likely false. Consider the following case in which researchers aim to explore the cognitive processes engaged in the multiplication of numbers. Suppose that they randomly sampled half of the subjects from a Western Caucasian population and another half from an East Asian population. Westerners are more likely to use their known method of long multiplication; one multiplies the multiplicand by each digit of the multiplier and then

adds up all the appropriately shifted results. Easterners may use the traditional Asian stick method (sometimes referred to as the Chinese or Japanese stick multiplication method), a more visual way of using drawn lines to find the result. The average of such data would describe a non-existing method for multiplication, as the average result placed the expectations between two very different cognitive strategies. Averaging across subjects could have a clearly detrimental effect on inferences about the processes of interest and would lead to false conclusions.

In the last decade many researchers have voiced concerns about the futility of the averaging tool in learning about the true values associated with specific cognitive operations (e.g., Estes, 1956; Maddox, 1999; Gallistel, 2009; Fific et al., 2010; Fitousi and Wenger, 2011; Koop and Johnson, 2011; Hills and Hertwig, 2012; Benjamin, 2013; Pachur et al., 2014). There is a rapidly increasing trend toward accounting for individual-specific cognitive operations in contrast to testing models based on universal cognitive operations. Accounting for individual differences is essential to assessing which model provides the best fit to experimental data (Broder and Schutz, 2009; Dube and Rotello, 2012; Kellen et al., 2013a,b; Turner et al., 2013). Evidence for individual differences has been reported in judgment strategies (e.g., Hilbig, 2008; Regenwetter et al., 2009), and the analyses of individual data have been called for repeatedly when investigating fast and frugal heuristics (Gigerenzer and Brighton, 2009; Marewski et al., 2010). On the other hand there are good reasons why aggregate data should be considered under some circumstances (Cohen et al., 2008; Chechile, 2009).

The second way correct inferences about underlying cognitive processes can be jeopardized occurs when researchers fail to create the appropriate input—that is—fail to create a minimally complex research design that is sufficient and necessary to obtain diagnostic response outputs. A non-diagnostic design does not permit differentiation between tested cognitive models as the models can mimic each other in the output. It logically follows then that the input (namely a research design), should be complex enough to allow for confident model differentiation in the output. But a more complex design is more expensive. Then the question becomes: What is the "price" one has to pay in the complexity of a design so that one can make correct inferences, and when do we start to see diminishing returns?

As in real life, the price of learning complex relations is sometimes underpaid. Take for example the above STM task research design used to make inferences about underlying serial/parallel STM processing. The design has only one independent variable of memory load and a dependent variable of response time. A researcher might believe that using say six memorized items in the input is the necessary and sufficient "price" to pay to learn about how six mental processes are organized. Here is the supposed bill: the sufficient and necessary price to pay to learn about the mental organization of a total of n cognitive processes (say six item comparisons) is a research design that has one independent variable with n number of levels. The price for one learned process is paid by one stimulus condition.

Unfortunately, using such a research design is likely to underestimate the true costs of diagnosing serial and parallel processing. This is because the serial and parallel cognitive models can

---

[1] In this study the terms cognitive strategies and cognitive operations are used interchangeably to refer to a set of mental processes organized in an identifiable mental network used in a specific task. In relevant literature these networks are also defined as mental architectures. In contrast, a cognitive process is subordinate term and indicates a single mental operation or a component of more complex cognitive system (mental architecture).

[2] Even this is not sufficient, e.g., every subject is exponential but their average appears to be from a different type of process (e.g., Brown and Heathcote, 2003).

easily mimic each other when only a memory load variable is used (Townsend, 1969, 1971, 1990; Townsend and Ashby, 1983).

Without a rigorous theory of how to define and measure the fundamental cognitive operations involved, minimal criteria for design complexity cannot be specified. In the absence of these criteria researchers will usually seek to increase the complexity of the research design. This is the case when cognitive models are tested by how well they can account for data across various tasks, that is, by seeking generalizability. In general it is advisable to challenge a cognitive model to account for as many possible findings when different inputs are manipulated. Only the model that can provide a good fit to as many different research conditions as possible is considered the most likely model, and those that fail to account for anything less than that are falsified[3] . So for example, the likely STM model should be able to account for all (various) observed effects (memory load, target serial position, stimulus modality, etc.) and should also be able to generalize easily to other conditions (e.g., Nosofsky et al., 2011). Although useful, generalizability doesn't precisely quantify the research design complexity value that is sufficient and necessary to diagnose the underlying cognitive structure of mental processes.

THE MINIMAL CRITERIA FOR THE COMPLEXITY OF A RESEARCH DESIGN

A recently proposed approach—the systems factorial technology (SFT)—sets the precise minimum required criteria for how complex a research design should be in order to be both sufficient and necessary to differentiate between several known properties of cognitive systems. The proposed SFT approach was designed to explore conditions under which the fundamental properties of mental processes, such as the order of processing (serial, parallel, coactive), stopping rule (terminating, exhaustive), process independence and capacity, could be inferred from data (e.g., Townsend and Ashby, 1983; Schweickert, 1985; Egeth and Dagenbach, 1991; Townsend and Nozawa, 1995; Schweickert et al., 2000). The SFT has been used in the context of various cognitive tasks: For perceptual processes (e.g., Townsend and Nozawa, 1995; Eidels et al., 2008; Fific et al., 2008a; Johnson et al., 2010; Yang, 2011; Yang et al., 2013), for visual and memory search tasks (e.g., Egeth and Dagenbach, 1991; Wenger and Townsend, 2001, 2006; Townsend and Fific, 2004; Fific et al., 2008b; Sung, 2008), for face perception tasks (Ingvalson and Wenger, 2005; Fific and Townsend, 2010), and for classification and categorization (e.g., Fific et al., 2010; Little et al., 2011, 2013).

---

[3]The current study doesn't evaluate model complexity as a quantitative criterion for model selection and falsification. The reasons are two-fold: (a) Current instantiation of SFT doesn't depend on model complexity to diagnose underlying cognitive models, it rather relies on recognition of qualitative patterns of RT and is completely non-parametric (for the parametric SFT approach see Fific et al., 2010). Nevertheless, one can argue that in the current paper the quantitative model comparison is possible as hypothesis testing is used to falsify certain classes of cognitive models. For example, in this paper the linear regression design is compared to the full factorial 2×2 ANOVA. (b) However, model selection is not necessary in this study: The linear regression model although a simpler model than the comparable full-factorial SFT design, makes logically incorrect inferences (as demonstrated in Supplementary Material). In such a case model complexity is a less important criterion to consider as one of the models is logically flawed.

To correctly diagnose an n number of cognitive processes, of an unknown cognitive system that is organized with respect to processing order, stopping rule and process dependency, SFT prescribes the following minimal criteria for a research design's complexity:

(a) The number independent variables used should be equal to the number of processes under examination, n.
(b) Each independent variable should vary between (at least) binary values of saliency. The saliency is operationally defined as a manipulation that selectively influences a single process of interest, such that the process is speeded up (H = high saliency) or slowed down (L = low saliency).
(c) The levels of all independent variables should be factorially combined, that is, orthogonally crossed with all other levels of the other variables. Thus, the total number of experimental conditions is equal to 2^n.

So, if a cognitive system under investigation consists of two processes that could be organized in either a serial or a parallel fashion, then the design should include two independent variables with two levels each, factorially combined, resulting in $2^2 = 4$ conditions. If a cognitive system consists of four processes, the design should include four factors, factorially combined with at least two levels of each factor, thus resulting in $2^4 = 32$ experimental conditions.

The required research design's complexity increases exponentially with research aspirations. In practice as the number of conditions increases this means that the SFT minimal criteria for differentiating between cognitive models could require lots of conditions and trials. So it is quite understandable that researchers usually use generalizability as criteria for model testing instead. The truth is that many of these research designs do not meet the minimal SFT criteria for testing different cognitive models, leading to conclusions that could be flawed.

In studies of the optimal research design, the SFT approach utilizes a so-called full-factorial design enabling a detailed processing structure analysis. If only a fraction of the full factorial design is used then this is broadly defined as a fractional-factorial design (FFD). In general FFD designs are useful as they can provide some important insights about the processes under consideration while saving on the complexity of a research design and thus saving time and effort. However, they may fail to identify important interactions between factors. As will be detailed in the next section, it is exactly the interaction information that provides the critical insights necessary to differentiate between cognitive processes. Although there is a great deal of published research about cognitive properties that can't be characterized as utilizing the FFD research design (e.g., Sternberg, 1966; Bradshaw and Wallace, 1971; Lachmann and van Leeuwen, 2004) this study will not analyze it in detail. For simplicity sake, this paper will refer to any incomplete SFT full-factorial design as an FFD design.

The second way correct inferences can be jeopardized is when using an FFD research design a researcher acts *as if* he/she has reduced the dimensionality of a full-factorial design. As such the important critical information about how to differentiate between cognitive systems is lost. So for example, the full-factorial

SFT design prescribes six variables and $2^6 = 64$ conditions to learn about six STM processes. Such could be a design in which each memory item's saliency (high-low) is factorially combined with all other memory items' saliencies (for $n = 2$ see Townsend and Fific, 2004; for up to $n = 4$ see Yang et al., 2014). If instead a researcher collapses the load variable across saliency, then the resulting design is a FFD design having only the memory load variable in the input. By collapsing across the input variables the critical test conditions are dropped out, and the minimal SFT diagnostic criteria have not been reached. Thus, the likelihood of making correct inferences about any underlying cognitive processes decreases dramatically.

The remainder of this paper will outline the basic SFT tools applied on cognitive systems with two processes. Then the author will proceed with the empirical evidence showing how SFT combined with individual subject analysis can be used to improve inferences rendered unreliable by the two jeopardies.

A GENERIC COGNITIVE TASK

Take for example a generic short-term memory/visual memory search task: the search set consists of two items ($n = 2$) and the task is to decide whether a target item was in the search set. For simplicity the author limits the analysis to target-absent trials only, in which a subject has to search an entire search set. This is the case of an exhaustive search. The question is whether processing is serial, parallel, coactive, or none of the above. In general, limiting the analysis only on target-absent responses potentially can harm diagnostic accuracy as it neglects a possible decision criteria trade-off between target-present and target absent responses. The analysis of target-absent responses only would still be sufficient for the current illustration purposes.

THE SFT FULL-FACTORIAL DESIGN

The adequate minimal SFT research design of the above task should include two factors with at least two levels, thus the total number of conditions should be $2^2 = 4$.

The first factor is operationally defined as the saliency of the first item in the search set, and the second factor is defined as the saliency of the second item in the search set. The saliency has binary values which allow for speeding up or slowing down of a particular process. (In what follows, H indicates a fast process, or high item-to target dissimilarity, and L a slow process, or low item-to-target dissimilarity). The idea here is that the memorized item with high saliency is processed faster than the item with low saliency, as the H item is more dissimilar to the target. In the generic task described above the cognitive operation of item scanning requires less processing time to determine that an H item is not a target, and can reject it quicker than an L item.

In each trial two items make a search set, and thus the factorial combination of items' saliencies will result in four experimental conditions: HH, HL, LH, and LL—the so-called double factorial design (2×2, as employed in an analysis of variance). For example, HLindicates a condition where the first factor (processing the first item) is of high saliency and the second factor (processing of the second item) is of low saliency (see **Figure 1A**).

It is important to note that using the double factorial design, the different cognitive processing orders will exhibit different data
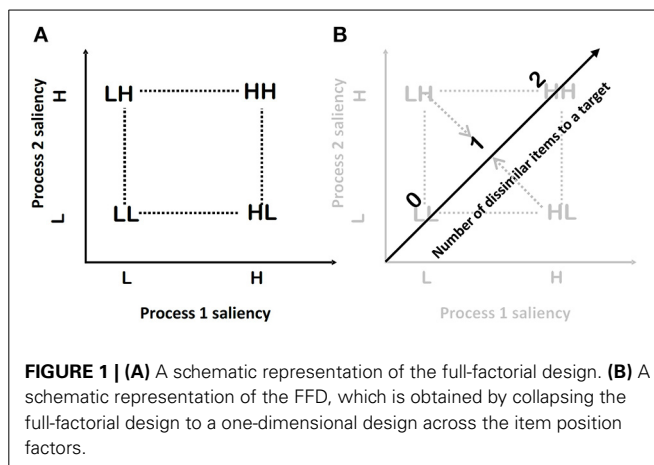
FIGURE 1 | (A) A schematic representation of the full-factorial design. **(B)** A schematic representation of the FFD, which is obtained by collapsing the full-factorial design to a one-dimensional design across the item position factors.

patterns of mean reaction times, which brings us to the main statistical tests used in SFT.

Mean Interaction Contrast (MIC): The MIC statistic calculates the interaction between the factors, similarly as in an interactive analysis of variance (ANOVA) (Sternberg, 1969; see also Schweickert, 1978; Schweickert and Townsend, 1989):

$$\text{MIC} = (\text{RT}_{LL} - \text{RT}_{LH}) - (\text{RT}_{HL} - \text{RT}_{HH}) = \text{RT}_{LL}$$
$$- \text{RT}_{LH} - \text{RT}_{HL} + \text{RT}_{HH} \qquad (1)$$

where RT is response time. This statistic is obtained by taking the double difference of mean RTs associated with each level of separate experimental factors (in this case, 2×2 factorial conditions). So, for example, mean RT_{HL} indicates mean response time for the condition where the first factor (processing the first item) is of high saliency and the second factor (processing the second item) is of low saliency. **Figure 2** shows typical patterns of MIC tests that are expected for different processing orders, for the fixed exhaustive stopping rule.

MIC is considered a valid test providing that the following conditional assumptions hold: (a) Processing rate for any position L is always slower than H, (b) The single factors selectively influence only single sub-processes (position one and two), and (c) The independence between processes hold. Violation of any or all assumptions leads to a violation of the mean RT orderings of the experimental situations $\text{RT}_{LL} > \text{RT}_{LH}$, $\text{RT}_{HL} > \text{RT}_{HH}$, which is considered a quick test of the conditional assumptions.

The pattern of "additivity" is reflected by an MIC value of 0 (**Figure 2**). In an ANOVA, additivity is indicated by an absence of interaction between factors, thus implying that the effects of individual factors simply "add" together. This finding supports serial processing, in which the total response time is the sum of individual times stemming from each factor. Likewise, "overadditivity" is reflected by an MIC > 0 (a positive MIC), and "underadditivity" is reflected by an MIC < 0 (a negative MIC). Formal proofs of the results expressed below are provided by Townsend (1984), Townsend and Nozawa (1995) for parallel and serial systems, and for a wide variety of stochastic mental networks by Schweickert and Townsend (1989). Townsend and Thomas (1994, also see Dzhafarov et al., 2004) showed the consequences of the failure

FIGURE 2 | Schematic illustration of three main patterns of mean RTs, mean interaction contrasts (MICs), and the corresponding underlying cognitive processes when the stopping rule is fixed to be exhaustive.
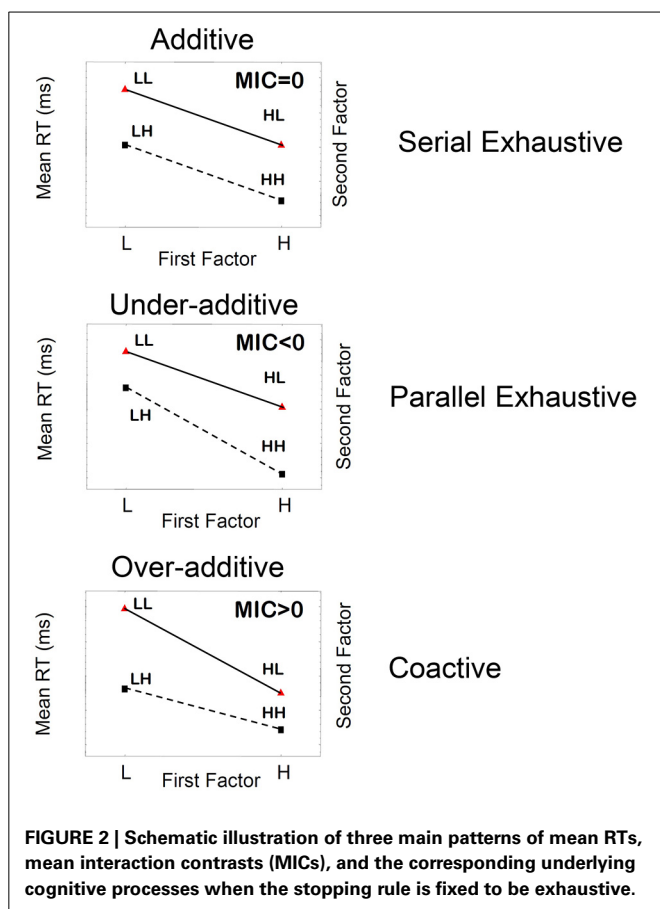
of selective influence when channels (items, features, etc.) are correlated.

If processing is strictly serial, then the MIC value will equal zero; that is, the pattern of mean RTs will show additivity. For instance, if processing is serial exhaustive, then the increase in mean RTs for LL trials relative to HH trials will simply be the result of the two individual processes slowing down, giving us the pattern of additivity illustrated in **Figure 2**, top panel. Parallel exhaustive processing results in a mean RT pattern of under-additivity (MIC < 0) (**Figure 2**, middle panel). Finally, coactive processing will lead to a pattern of overadditivity of the mean RTs (MIC > 0), as illustrated in **Figure 2** bottom panel. Coactive processing is a form of parallel processing in which information from parallel processing units are pooled together into one unit, by the virtue of summation of signals from the two units. Coactivation gives rise to perceptual unitization, forming perceptual objects whose features are not analytically separable.

The SFT provides strong grounds for model comparison and model falsification, in both the non-parametric and parametric treatments of the theoretical processes. Useful statistical tools are described in several publications and are available online (Townsend et al., 2007; Houpt et al., 2014).

THE FRACTIONAL-FACTORIAL DESIGN (FFD)
To get an FFD the author reduces the dimensionality of the above full-factorial design (**Figures 1A,B**). The resulting FFD design

uses only 3 conditions from the original full-factorial design. The collapse of the full factorial design across the item position factors could be visualized as a projection of the conditions to a new single dimension (**Figure 1B**). I define this dimension as the number of items in a search set that are dissimilar to the target. In the HH condition, both items are dissimilar. Thus, the value is two. In the HL and LH conditions, only one item is dissimilar thus the value is one; and in the LL condition both items are similar, and thus the number of dissimilar items is zero. The observed mean RT can be plotted as a function of the number of dissimilar items, defining the RT-dissimilarity function.

Surprisingly this particular FFD design has been used in several studies to explore cognitive processes. The RT-dissimilarity function has been employed previously in the same-different judgment task (Nickerson, 1965, 1969; Egeth, 1966; Miller, 1978; Proctor, 1981; Farell, 1985; see Sternberg, 1998 for review).The general finding was that RT decreased as a number of differing dimensions between the items (Goldstone and Medin, 1994), number of dissimilar items in search set, or as a function of the structural complexity (Checkosky and Whitlock, 1973; Schmidt and Ackermann, 1990; Lachmann and van Leeuwen, 2004).

The important diagnostic feature here is the shape of the RT-dissimilarity function: if the function is strictly linear it indicates serial processing (Egeth, 1966; Posner and Mitchell, 1967; Lachmann and van Leeuwen, 2004), and if the function is non-linear it indicates parallel processing (Posner, 1978). The property of linearity can be assessed by conducting a linear regression analysis and would be shown in the coefficient of determination R^2-value (e.g., Lachmann and Geissler, 2002; Lachmann and van Leeuwen, 2004, p. 11, inferred serial processing by showing linear functions, $0.98 \leq R^2 \leq 0.99$).

Indeed different cognitive models predict the characteristic change in RT-dissimilarity function shape. Serial exhaustive models predict that the mean RT would linearly decline as a function of item-to-target dissimilarity. Provided that a low-dissimilar item is processed slower than a high-dissimilar item, and that processing is conducted in the item-to-item fashion, the mean RT should decline with the same rate as the number of dissimilar items increases in the search set. Parallel exhaustive models predict a convex non-linear RT-dissimilarity function. In contrast, the coactive model predicts a concave non-linear RT as a function of target-to-item dissimilarity (see Supplementary Material for the derivations).

It is important to note that even though the mean RT-dissimilarity function is FFD, some diagnostic cues enable differentiation between cognitive processing strategies.

The robustness of the SFT and FFD designs to the first jeopardy: Averaging across subjects' mixed cognitive strategies and predictions of the two designs.

Neither of the two approaches is immune to the first jeopardy. When we average results of subjects who used different cognitive strategies, the resulting MIC signature and RT-dissimilarity function could reveal the most dominant cognitive system or could indicate a ghost cognitive system—a non-existing one.

Consider the generic task in which the stopping rule was set to be exhaustive. In order to make a correct decision all memorized items in the search set have to be processed. Each cognitive

strategy (serial, parallel, and coactive) could be used to search the search set, but some strategies may be more preferable under certain conditions. Serial processing could be employed when it is advantageous to invest all attention to one unit at a time with a possibility for early termination. Parallel processing may be employed when all information is available and the cognitive system does not see possible limitations due to capacity sharing between multiple concurrently processed items. Coactive processing may be involved with processes that have historically occurred together and thus built a joint path in the cognitive system (perhaps a neural unit). More importantly, what is unknown to researchers is whether or not each of these cognitive processing strategies may be individual subject specific. It could be expected that some human subjects have developed more reliance on some of these strategies than on the others.

In the SFT design the following three MIC signatures could be observed. Subjects could either exhibit a parallel search, showing the underadditive MIC pattern (**Figure 2** middle), a serial search showing the additive MIC (**Figure 2**, top), or a coactive search (parallel but not independent processes) showing the overadditive MIC pattern (**Figure 2** bottom). Provided that the base rate for each processing strategy is the same, the results of averaging across subjects would predict convergence to the MIC additive signature.

Similarly in the FFD design, the subjects would show all three types of curving in the RT-dissimilarity function, concave, convex and linear. The average outcome RT-dissimilarity function would tend to converge to the linear function.

A surprising result will occur when sampled subjects are only parallel and coactive processors: a ghost cognitive strategy will be inferred. Both the averaged MIC and the RT-dissimilarity would indicate serial processing (additive MIC and linear RT function), despite that not a single subject could be characterized as such.

THE COMPARISON TEST

The main goal of the current paper is to explore how effective the mean RT analysis methods are in inferring the organization of cognitive processes when both jeopardies are in place. Thus, this study cross combined the two jeopardies and compared the four resulting conditions (**Table 1**).

As a reference point the author will analyze the data from Condition 0 which both adheres to the SFT minimal criteria for the correct diagnosing of cognitive processes, and is based on individual subjects analyses (**Table 1**). Condition 0 uses the previously published MIC results of individual subject data on a large number of trials possessing lots of statistical power (Townsend and Fific, 2004; Fific et al., 2008b).

In Condition 1, the author tests the effect of the across-subject averaging on MIC test accuracy in identifying cognitive processes. In Condition 2 the author tests the effect of using an FFD design on making inferences regarding the individual subjects' data, using a regression analysis of the RT-dissimilarity function. Finally, in Condition 3 the data will be exposed to both jeopardies: the averaging across subjects and the design marginalization using FFD. In this condition the author analyzes the group mean RT-dissimilarity functions using linear regression analysis.

The expectation is that when compared to Condition 0 the three conditions will show deterioration in their ability to correctly diagnose cognitive processes. Most of the misdiagnoses should be observed in Condition 3. Although the current expectations could be logically derived from earlier works, such systematic evidence is sparse. The author hopes that the current study will illuminate both the role of individual subject analysis and the application of SFT in learning about cognitive processes.

METHODS

The results reported in this section are based on the reanalysis of data collected in previous studies (Townsend and Fific, 2004; Fific et al., 2008b). Specific details about the participants and stimuli are presented in the original papers. Here I outline the details which are pertinent to the current investigation.

PARTICIPANTS

Five participants, 2 females and 3 males participated in a short-term memory search study (Townsend and Fific, 2004). Four participants, two females, and two males participated in a visual search study (Fific et al., 2008b); four participants, three females, and one male participated in the visual search study on patterns (Fific et al., 2008b). All participants were paid for their participation.

STIMULI

Short-term memory study (Townsend and Fific, 2004)

Stimuli were pseudo-words in consonant-vowel-consonant (CVC) form. Two items made a search set, presented on different search-set positions (first, second). To produce the saliency effect, we manipulated phonemic dissimilarity of a search set-item to the target item. The items were drawn from two sets of phonologically confusable Serbian language consonants: fricatives (F, S, V) and semi-vocals (L, M, N). We generated different dissimilarity of search-set items to the target item by constructing the target and test items from letters drawn either from the same group or from different groups.

Visual search on pseudowords (Experiment 1, Fific et al., 2008b)

Stimuli were Cyrillic letter-strings constructed from letters of the Serbian alphabet. The visual complexity of the letter-string stimuli was manipulated by varying the number of letters that made up a single item (1, 2, or 3 consonants). The saliency effect was produced by manipulating the degree of visual dissimilarity

Table 1 | Cross combination of the levels of the two jeopardies in a 2 × 2 analysis, leading to four different analysis conditions.

	Analysis level	
	Individual	**Group**
RESEARCH DESIGN		
Full factorial (MIC)	0	1
Fractional factorial (regression)	2	3

The first jeopardy is defined as the difference between the individual and group subject analyses with regard to inferring the details associated with the cognitive processes of interest. The second jeopardy is defined as the difference between the full- and fractional-research designs with regard to inferring those same details.

between the item and the target items. We employed two sets of letters: letters with curved features and letters with straight-line features. We generated different dissimilarity of search-set items to the target item using the same principles as in the above study.

Visual search on visual patterns (Experiment 2: Fific et al., 2008b)

As stimuli, we used meaningless visual patterns taken from Microsoft's Windows standard fonts.

DESIGN AND PROCEDURE

Short-term memory search (Townsend and Fific, 2004)

Each trial consisted of a fixation point and warning low-pitch tone for 1 s, successive presentation of two items in the search set for 1200 ms, an inter-stimuli interval (ISI), and a target. The ISI was defined as the interval between the offset of a search set and the onset of the target. The ISI period started with a fixation point and a second warning high-pitch tone which lasted for 700 ms. Onset of this second warning signal was activated so that its end coincided with the end of the ISI period.

The task was to decide whether a target was presented in a search set. The target was randomly chosen to be present in one-half of the memory set trials and absent in the other half Participants signified their answer, "yes" with one index finger and "no" with the other. Only target-absent trials were analyzed.

The analyzed research design consisted of the three within-subject factors: Inter-stimulus interval (ISI, 700 and 2000 ms) × Dissimilarity of item in position one (H,L) × Dissimilarity of item in position two (H, L). The last two factors constituted the full factorial SFT design permitting the assessment of processing order.

Participants ran around 44 blocks of 128 trials each. Each block was divided into 6 sub-blocks of 20 trials (except the last one which had 28 trials). The participants were requested to achieve very high accuracy, and usually only one block was completed on a particular test day. Thus, each mean RT in a specific ISI condition and particular factorial combination possessed between 300 and 400 trials per participant (depending on duration of participation). Brief rest periods were allowed every 24 trials.

The ISI was manipulated between blocks, whereas factorial combinations (HH, HL, LH, LL) varied within blocks.

Visual search on pseudowords (Experiment 1, Fific et al., 2008b)

Each trial started with a fixation point that appeared for 700 ms and a low-pitch warning tone of 1000 ms, followed by the presentation of the target item for 400 ms. Then, a mask was presented for 130 ms, followed by two crosshairs that indicated the positions of the two upcoming test items that made the search set. A high pitch warning tone was then played for 700 ms, followed by the presentation of the two items in the search set.

The task was to decide whether or not the target was presented in the search set. Half of the trials were target present and half were target absent. On each trial, the participant had to indicate whether or not the target item appeared on the search set by pressing either the left or the right mouse key with his or her corresponding index finger. RTs were recorded from the onset of the test display, up to the time of the response. Participants were

asked to respond both quickly and accurately. Only target-absent trials were analyzed.

The analyzed research design consisted of three within-subject factors: Stimulus complexity (C = 1, 2, or 3) × Dissimilarity of item in the left position (H, L) × Dissimilarity of item in right position (H, L). The stimulus complexity was operationally defined as the number of letters used to form the stimulus items. The last two factors constituted the full factorial SFT design permitting the assessment of processing order.

The two test items in the most complex condition (C = 3, with the widest stimuli) spanned 5 cm horizontally. At a viewing distance of 1.7 m from the computer screen, this width corresponds to a visual angle of 1.86 degrees, well within the fovea.

Each participant performed on 30 blocks of 128 trials each. The order of trials was randomized within blocks. The complexity of the presented items (i.e., the number of letters: C = 1, 2, or 3) was manipulated between blocks, whereas factorial combinations (HH, HL, LH, LL) varied within blocks. For each participant, the mean RT for each conjunction of item complexity and factorial combination was calculated from approximately 200 trials.

Visual search on visual patterns (Experiment 2, Fific et al., 2008b)

This condition was identical to the C = 1 condition of the previous study, except that it employed visual patterns as stimuli instead of letters. Each participant performed in 10 blocks of 128 trials.

RESULTS

CONDITION 0: INDIVIDUAL SUBJECT DATA, MIC ANALYSIS

The results of the MIC tests are published elsewhere (Townsend and Fific, 2004; Fific et al., 2008b). The author summarizes the findings in **Table 2**.

All subjects' results satisfied the ordering of mean RTs ($RT_{LL} > RT_{LH}$, $RT_{HL} > RT_{HH}$), except for the first subject in the C = 1 condition of the visual search task (**Table 2**). In addition, all subjects showed significant main effects of the single factors, that is, the effect of high and low dissimilarity for each item position. Highly dissimilar items always showed on average faster processing rates than the low dissimilar items, for both item positions (1 and 2). These findings indicated that the basic manipulation of item-to-target dissimilarity produced the expected cognitive effect and furthermore that the processing of an item in each particular position occurred. Being uniform for all subjects, these results were not reported in the table.

The critical MIC test results were based on the inspection of the significance of an interactive component and the sign value of the MIC score. As reported in **Table 2** the individual-subject analyses showed individual subject variability in MIC values. All MIC values were interpretable (except the first subject in the C = 1 condition), and the signatures each fell into one of the expected categories.

Conclusion

The subjects' MIC values showed large variability across the three experiments. In the two visual studies subjects showed primarily over-additive results (9 subjects) and some additive results (6 subjects), thus implying coactive and serial processing. One subject's

Table 2 | Summarized ANOVA results for the MIC tests at different levels of subject analysis.

	MIC test ANOVA			Full-factorial design conditions					
	df 2	*F*	η²	LL (ms)	LH (ms)	HL (ms)	HH (ms)	*MIC* (ms)	Inference
GRAND MEAN									
	23992	15.4	0.001	772	662	662	572	20	Coactive
VISUAL SEARCH: PSEUDOWORDS									
Mean subjects	7458	50.2**	0.007	984	730	762	587	78	Coactive
Complexity									
C = 1	595	3.1†	0.005	619	564	623	530	−38	–
C = 2	633	34.9**	0.052	1106	711	802	581	175	Coactive
C = 3	631	4.4*	0.007	1302	885	934	579	63	Coactive
C = 1	591	1.2	0.002	557	506	509	470	12	Serial
C = 2	630	41.0**	0.061	908	649	681	554	132	Coactive
C = 3	626	3.1†	0.005	1149	799	848	549	51	Serial
C = 1	590	0.9	0.001	622	577	561	534	19	Serial
C = 2	632	59.4**	0.086	963	671	643	527	176	Coactive
C = 3	628	14.0**	0.022	1191	856	808	578	106	Coactive
C = 1	595	2.0	0.003	678	639	631	609	17	Serial
C = 2	633	33.8**	0.051	1194	869	949	766	142	Coactive
C = 3	630	11.7**	0.018	1446	995	1113	753	91	Coactive
VISUAL SEARCH: PATTERNS									
Mean subjects	2346	5.4*	0.002	668	577	587	532	36	Coactive
Complexity									
C = 1	587	2.1	0.004	863	699	746	630	49	Serial
C = 1	576	4.7*	0.008	750	642	655	617	70	Coactive
C = 1	584	2.0	0.003	520	469	465	432	17	Serial
C = 1	587	4.4*	0.007	545	494	482	452	22	Coactive
SHORT-TERM MEMORY SEARCH									
Mean subjects	14180	11.0**	0.001	676	641	622	571	−15	Parallel
Interstimulus interval									
ISI = 700	1375	0.8	0.001	606	565	559	507	−12	Serial
ISI = 700	1645	0.2	0.000	632	607	595	565	−5	Serial
ISI = 700	1202	11.4**	0.009	598	590	562	518	−37	Parallel
ISI = 700	1394	2.2	0.002	747	706	690	664	15	Serial
ISI = 700	1439	0.3	0.000	786	703	666	593	10	Serial
ISI = 2000	1379	0.3	0.000	628	567	561	507	7	Serial
ISI = 2000	1710	3.7†	0.002	640	628	600	567	−21	Serial
ISI = 2000	1201	14.7**	0.012	613	592	577	512	−43	Parallel
ISI = 2000	1387	5.6*	0.004	748	730	717	672	−27	Parallel
ISI = 2000	1412	4.2*	0.003	761	708	680	591	−36	Parallel

**p < 0.01, *p < 0.05, †p < 0.08. The df1s were 1.

results were inconclusive, violating the conditional assumptions of selective influence and or process independence. The subject could also exhibit an unknown type of cognitive strategy. In contrast, the subjects in the memory study showed either additivity (6 subjects) or under-additivity (4 subjects), thus implying the presence of both serial and parallel processing across subjects. See **Table 3** for summary.

CONDITION 1: AVERAGED SUBJECTS DATA, MIC ANALYSIS

First I analyzed the MIC results averaged across subjects and then across all experimental conditions (the visual and memory search conditions) to and obtained the grand mean MIC data (**Figure 3A**). Then, using ANOVA I tested the significance

of the interaction between two factors. Each factor is defined as the item's item-to-target-dissimilarity (high, low), for one of the two positions in the search set position. The interaction test is used to provide a statistical significance finding for the MIC test. The interaction between the two factors was found to be significant $F_{(1,\ 23992)} = 15.37$, $p < 0.01$, $\eta^2 = 0.001$. The observed MIC = 20 ms, indicating overadditivity (**Figure 3**, top left panel).

Conclusion: all subjects processing (26) was based on the coactive processing model

Next, I conducted the MIC test conditioned on the type of cognitive task used. I break down the overall mean RT results into three different experimental studies: the visual search task using

pseudowords, visual search task using visual patterns, and short-term memory task. The results of MIC tests are presented in **Table 2** (the rows "*Mean subjects,*" and also in **Figures 3B–D**).

Conclusion

The results indicated that when the MIC is calculated by averaging across all subjects the MIC test showed overadditivity (MIC > 0) in both of the visual search tasks, thus implying coactive processing (for $12 + 4$ subjects). In a sharp contrast, the MIC indicated underadditivity (MIC < 0) in the short-term memory experiment, thus implying parallel processing, for all 10 subjects.

CONDITION 2: INDIVIDUAL SUBJECT DATA, REGRESSION ANALYSIS

The individual mean RT-dissimilarity functions are analyzed. The author conducted the linear regression analysis between mean RT and the number of item-to-target dissimilar items in a search set (0, 1, 2 items in a search set dissimilar to the target) for

each individual subject across different experimental conditions (**Table 4**, left hand side).

Using linear regression, the linear relationship accounts for a large percent of mean RT variability for most of the subjects (it ranged from 94 to 100% across all subjects, with the mean $R^2 = 98\%$ and $SD = 0.0282$).

Conclusion 1

Extremely high R^2-values of linear function fits among subjects implied a strict serial exhaustive process.

It is questionable whether the results would indicate significant curving of the mean data points, either of the convex or concave type. The standard way to test whether the data could be better explained by the linear or non-linear (polynomial of a second degree) model, is to conduct the regression analysis using the second-order polynomial regression function (quadratic). But in this study the use of quadratic regression is precluded as there are only three data points to be fitted. That is, there would be the same number of free parameters as the number of points, so the test for the significant R^2 change from a linear to non-linear model would not be valid.

To provide the alternative test for curvature of the mean RT dissimilarity data the author conducted another regression analysis on the individual subject RT data this time by using all RTs not averaged across the dissimilarity conditions (0, 1, 2). Now the author compared whether the adding of a second order polynomial component could be used to significantly improve the goodness of fit (R^2-value) (**Table 4**, right hand side).

Table 3 | Summary of the inferences across different comparison conditions from Table 1.

	Serial	Parallel	Coactive	Unaccounted
Condition 0 (full)	12	4	9	1
Condition 1 (jeop 1)	–	10	16	–
Condition 2 (jeop 2)	13	4	9	–
Condition 3 (jeop 1 and 2)	26	–	–	–
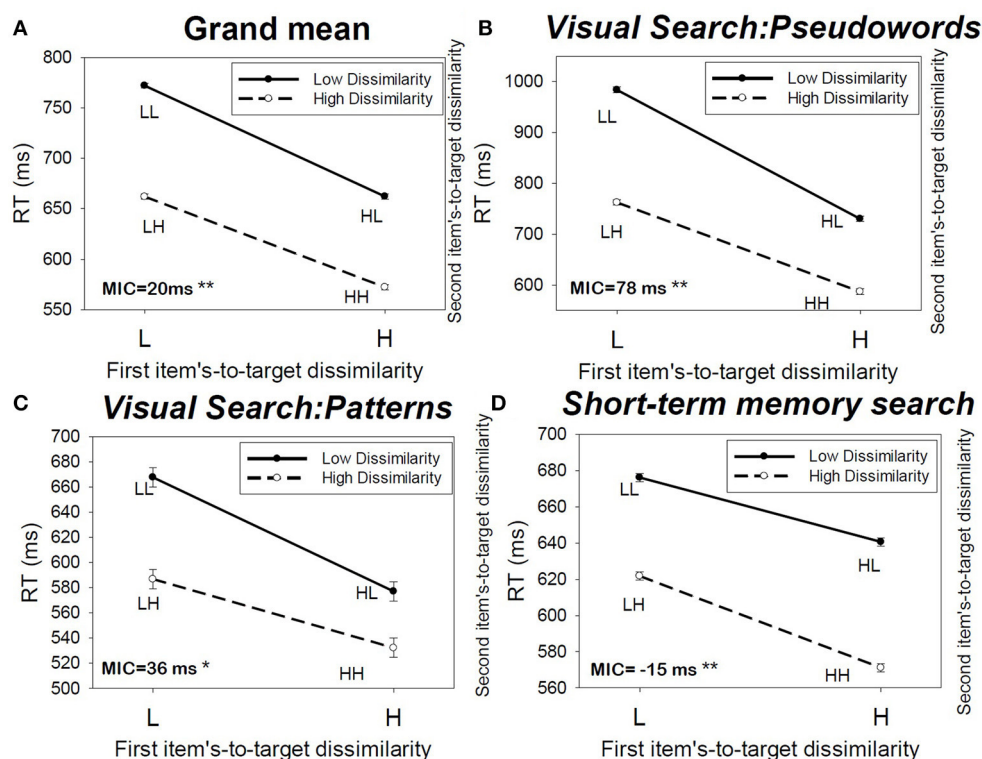
FIGURE 3 | (A) Mean RT averaged across the subjects, and **(C,D)** the MIC test results for different experimental conditions.

Table 4 | Summarized linear regression results for different levels of subject analysis.

	Linear regression				Concavity/convexity test			
	R^2	F	Intercept (ms)	Slope (ms)	R^2 change	df2	F change	Inference
GRAND MEAN								
	1.00	299.4	769	−100	0.001	23993	15.38**	Coactive
VISUAL SEARCH PSEUDOWORDS								
Mean subjects	0.99	74.7	965	−196	0.005	7459	49.97**	Coactive
Complexity								
C = 1	0.94	16.4	625	−44	0.005	596	2.97†	Serial
C = 2	0.97	27.3	1077	−263	0.026	634	33.86**	Coactive
C = 3	1.00	399.2	1292	−361	0.002	632	4.33*	Coactive
C = 1	0.99	169.5	555	−44	0.002	592	1.20	Serial
C = 2	0.96	21.5	886	−177	0.032	631	40.64**	Coactive
C = 3	1.00	419.0	1140	−300	0.002	627	3.10†	Serial
C = 1	0.99	68.0	619	−44	0.001	591	0.86	Serial
C = 2	0.95	18.4	934	−218	0.042	633	59.21**	Coactive
C = 3	0.99	100.4	1173	−306	0.009	629	13.90**	Coactive
C = 1	0.98	49.2	675	−35	0.003	596	2.05	Serial
C = 2	0.97	27.6	1170	−214	0.026	634	32.57**	Coactive
C = 3	0.99	174.8	1431	−347	0.006	631	10.91*	Coactive
VISUAL SEARCH PATTERNS								
Mean subjects	0.97	36.0	663	−69	0.002	2347	5.36*	Coactive
Complexity								
C = 1	0.99	69.4	855	−116	0.003	588	2.08	Serial
C = 1	0.92	10.8	739	−67	0.008	577	4.69*	Coactive
C = 1	0.99	76.4	517	−44	0.003	585	2.05	Serial
C = 1	0.98	54.1	542	−47	0.006	588	4.41*	Coactive
MEMORY SEARCH								
Mean Subjects	0.99	152.8	679	−53	0.001	14181	10.93**	Parallel
Interstimulus interval								
SI = 700	1.00	222.6	608	−50	0.001	1376	0.81	Serial
ISI = 700	1.00	465.7	633	−33	0	1646	0.21	Serial
ISI = 700	0.94	14.4	605	−40	0.009	1203	11.27**	Parallel
ISI = 700	0.99	97.6	745	−42	0.001	1395	2.16	Serial
ISI = 700	1.00	1132.0	785	−97	0	1440	0.35	Serial
ISI = 2000	1.00	926.6	627	−60	0	1380	0.32	Serial
ISI = 2000	0.97	35.2	644	−37	0.002	1711	3.77†	Serial
ISI = 2000	0.94	16.6	620	−50	0.011	1202	14.64**	Parallel
ISI = 2000	0.96	23.7	753	−38	0.004	1388	5.57*	Parallel
ISI = 2000	0.99	68.7	767	−85	0.003	1413	4.07*	Parallel

**p < 0.01, *p < 0.05, †p < 0.08. Each linear regression was conducted with 1 degree of freedom for the concavity/convexity test. The first dfs were 1 as stated, and the df2s are reported in the table.

Conclusions 2

The results of the regression analysis showed a significant curving of the individual subject data (**Table 4**, under Concavity/convexity test). The inferences about cognitive processes paralleled those of the MIC tests conducted on individual subjects' data (**Table 2**).

The only exception was the first subject whom was categorized now as a serial processor unlike in the MIC test in which this subject couldn't be classified in one of the three processing strategies.

CONDITION 3: AVERAGED SUBJECTS' DATA, REGRESSION ANALYSIS

First, I analyzed the data when averaged across subjects (individual data combined from the three experimental conditions). I conducted the linear regression analysis between mean RT and the number of item-to-target dissimilar items in a search set (0, 1, 2 items in a search set dissimilar to the target).

The significant proportion of explained variability indicates that the mean RT linearly decreases with increasing the number of items that are dissimilar to the target (see **Figure 4**, and **Table 4**

FIGURE 4 | Linear regression analyses between Mean RT averaged across the subjects and the number of item-to-target dissimilar items in a search set, for different experimental conditions.
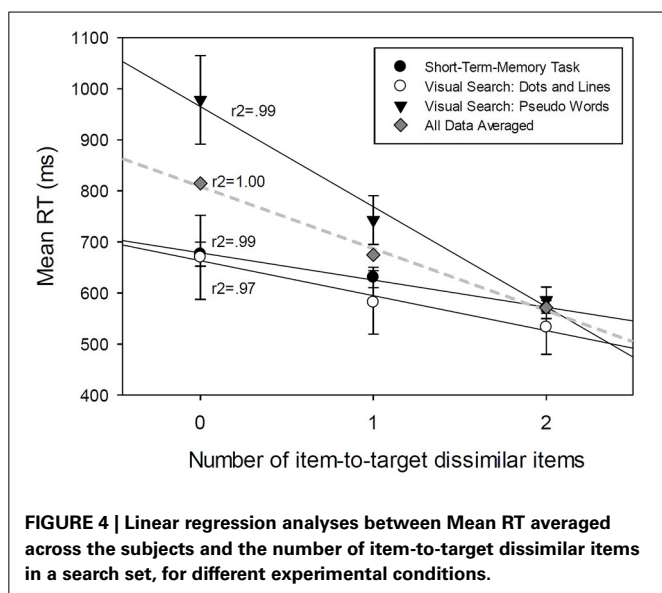
the first row *Grand Mean*). This relationship accounts for 100% of mean RT variability, $R^2 = 1$ (**Figure 4**).

Conclusion 1

All subjects (26) processed the stimuli using the serial processing strategy. The rate of sequential processing per item is defined by the value of the regression function slope which was estimated from data to be 100 ms per item.

Second, I conducted the regression analysis on RT averaged across subjects but sorted by type of experimental condition. I break down the overall mean RTs into three different experimental studies: the visual search task using pseudowords, the visual search task using visual patterns, and the short-term memory task.

The results of the linear regression analysis between the mean RT and the number of dissimilar items are presented in **Table 4** (the rows *Mean subjects*) and **Figure 4**. All three relationships accounted for between 97 and 99% of mean RT variability ($0.97 \leq r^2 \leq 0.99$). The explained variability indicated that the mean RT linearly decreases with the number of items that are dissimilar to the target (**Figure 4**).

Conclusion 2

All subjects (26) used the serial cognitive processing strategy across different conditions. The rate of sequential processing per item was different for different experimental studies (see **Figure 4**) and varied between 196 ms per unit for pseudowords to 68 and 53 ms per unit for simple visual stimuli and STM search.

GENERAL DISCUSSION

The main goal of the current paper was to explore the diagnostic accuracy of identifying the true underlying organization of cognitive processes in different experimental situations. The author discussed and analyzed two major concerns that could negatively impact the chances of achieving the main goals present in modern cognitive modeling trends.

The first concern deals with analyzing aggregated subjects data to infer the details associated with cognitive processes. Data aggregation across subjects has a long history of practice in the field. The main rationale is to use this powerful averaging tool to reduce random noise from observations and increase the power of diagnostic tests. The averaging tool rests on the conditional hypothesis that different subjects use the same cognitive operations. However, this hypothesis is rarely stated and substantiated. This is unfortunate, because when a researcher relaxes the conditional hypothesis that subjects use the same cognitive operations, surprising outcomes of averaging across subjects can occur. One of the most dramatic outcomes is inferring ghost cognitive processes. This error occurs when we average across two very different cognitive strategies. The resulting averaged data would support a strategy that may not exist and/or may not be theoretically feasible.

Research in the cognitive domain has over the years reached a critical view of the issue of individual differences in cognitive operations. It has become a pressing matter to address the issue of individual subject analysis. Scanning the current literature, the author found several such publications in the Journal of Psychological Review (Fific et al., 2010; Hills and Hertwig, 2012; Benjamin, 2013; Kellen et al., 2013b; Turner et al., 2013), the leading edge in theoretical advances relevant to the problem of averaging data across subjects.

The second concern deals with selection of the most appropriate research design to provide the best diagnostic performance in detecting cognitive processing details. A major trend in the cognitive domain relies on the principle that more complex designs make for better inferences. This is common practice in all areas of psychological research, which follows up on a recommendation for external generalizability. In that sense validation of a cognitive model should be based on the model's ability to generalize to as many as possible results and conditions as possible. In principle this is the right way to make scientific advances, especially in an area where it is not possible to precisely specify the minimal criteria for a research design complexity. For that reason the author introduced the SFT, which has been primarily designed to make inferences about underlying processing order (serial, parallel, coactive), stopping rule (terminating, exhaustive), and process dependency. The SFT approach proposes criteria for minimal research design complexity that can be used to construct the most effective diagnostic tools.

In this study the author reported the analysis of the effects of two possible ways inferences about cognitive processes can be jeopardized. The effect of the first jeopardy was measured by comparing the analysis of data averaged over the subjects to the analysis of individual subjects' data. The effect of the second jeopardy was measured by comparing the results of the analysis of the full factorial design (MIC) to the comparable FFD (linear regression on RT-difference function). More importantly the author cross combined the levels of jeopardies in a 2×2 analysis, leading to four different analysis conditions (**Table 1**). Condition zero served as a reference condition as it was the least influenced by both jeopardies. **Table 3** shows the summary of inferences about the cognitive processes across the conditions.

Aggregating the data across subjects (Jeopardy 1) reduced the diagnostic accuracy of our inferences about cognitive process to about half (accuracy = 13/26). The analyses of the effect of subjects' data aggregation (Condition 1 and 0), showed not only omissions in detecting of some cognitive strategies, such as missing to detect 12 cases of serial processing, but also showed a number of false recognitions of parallel or coactive processing. Comparing the diversity of individual strategies revealed by the MIC test in Condition zero to the strategies inferred after the data aggregation shows an interesting finding. The resulting aggregated inferences are not necessarily affected by the most inferred individual cognitive processes. As shown in the memory search experiment, the individual MIC analyses indicated 6 serial and 4 parallel subjects (**Table 2** bottom part—short term memory search). However, the inferences based on the aggregated values indicated parallel processing for all subjects (**Table 2** the line "*mean subjects*" for short-term memory). This could happen as the aggregated MIC score accumulated the size of effects from the individual subjects' data. The individual MIC scores showed 7 negative values, of which only 4 reached significance and were inferred to occur in parallel (**Table 2**, bottom).

Collapsing across the full-factorial research design to create a less complex design (Jeopardy 2) showed very good diagnostic accuracy of cognitive processes. Using the FFD as an alternative to the full-factorial design led to 25/26 correct inferences (see summary in **Table 3**, Condition 2, the individual results in **Table 4**). The study can conclude that the shape of the RT-difference function can be used as a complement to the MIC test.

However, this comes with three caveats. First, using the FFD would be very ineffective if the data was aggregated over the subjects (as presented in **Figure 4**). The results of regression analysis on the data aggregated over the subjects showed impressive fits to linear functions and showed very high R^2-values for each experiment. These results all point to the across subject uniform conclusion: serial processing (with low accuracy = 12/26). Second, even when the mean RT-difference functions are calculated for each separate subject (the Results Section, Condition 2, Conclusion 1) the curving of RT-difference functions may be difficult to detect using the conventional statistical test to reject the null hypothesis. To get the 25/26 correct detections, not only is the individual subjects analysis recommended but it is also recommended to use all data for each subject to test the curvature hypothesis (left-hand side **Table 4**, Concavity/convexity test). And the third and the most important caveat: using the FFD will very likely lead to increasing false alarm rates in detecting the known cognitive strategies, serial, parallel, or coactive. When scrutinized closely (Supplementary Material), the proposed FFD design shows good performance in inferring the correct cognitive strategy when all SFT conditional assumptions were met. However, if some of these assumption were not met, then FFD may not be able to detect that a violation occurred and will proceed to the incorrect inference. This is because FFD cannot test the mean RT ordering $RT_{LL} > RT_{LH}$, $RT_{HL} > RT_{HH}$, as the two LH and HL situations are aggregated. One such case is shown in **Table 3** and also in **Table 2**, the first row with C = 1. The subjects' MIC RT data showed a violation of the mean RT ordering $RT_{LL} >$

RT_{LH}, $RT_{HL} > RT_{HH}$ (**Table 2**, $RT_{LL} = 619$ ms, $RT_{LH} = 564$ ms, $RT_{HL} = 623$ ms, $RT_{HH} = 530$ ms) rendering the MIC test not valid for making inferences. The MIC test indicated that it is highly likely that some part of the conditional hypothesis was violated, thus preventing us from reaching a clear conclusion. However, when the FFD design is used the ordering of mean RTs allows for inferences ($RT_{LL} = 619$ ms, $RT_{LHandHL} = (564\,ms + 623\,ms)/2 = 593$ ms, $RT_{HH} = 530$ ms). The FFD design falsely inferred that this subject was a serial processor. In general the proposed FFD design is not an accurate test for the detection of "unknown" cognitive processes. The proof is shown in the corollary Supplementary Material.

Combining both jeopardies led to 12/26 correct inferences of serial processing (**Table 3**, Condition 3, see also **Figure 4**, "grand mean"). The linear regression analysis of RT-difference functions showed very high R^2-values of linear functions across different experiments, leaving practically no room for curving, and detection of either parallel or coactive processing. Thus, the results did not infer any parallel or coactive strategies which constitute almost half of the individual result's analyses. The disappointingly low level of 46% correct inferences clearly warrants the use of better methods. In the relevant published work so far the author was able to find several studies that may be characterized as using the Condition 3 methods (for example, Lachmann and Geissler, 2002; Lachmann and van Leeuwen, 2004) and thus could be challenged for the validity of their inferences about cognitive processes.

The results of the current study lead to the following recommendations. To improve the diagnostic accuracy of cognitive process, it is advisable to avoid the jeopardies by both adopting the minimal research design criteria as proposed by SFT, and also by conducting individual subject analysis, rather than conducting the analysis on aggregated subject data. Both jeopardies have been recognized in the scientific community as having detrimental effects on inferences but infrequently taken care of.

A review of current research trends reveals a number of researchers who are ready to switch from the subject aggregating procedures, and instead consider using individual subject analysis, if they are not already en route to developing and using such methods (e.g., Myung et al., 2000; Brown and Heathcote, 2003; Estes and Maddox, 2005; Soto et al., 2014). The main challenge in using individual subject data is to provide an integral assessment of such data that can enable clear communication between researchers. This is the case when one has to report a variety of individual differences in a large data set. Another issue is the question of what the best statistical methodology is for analyzing data while allowing for individual assessment. Some researchers have suggested using hierarchical Bayesian statistical inference as a principle tool for hypotheses testing, as it allows for natural incorporation of individual difference as a part of statistical tests (e.g., Rouder and Lu, 2005; Lee, 2008; Liu and Smith, 2009; Bartlema et al., 2014).

In this paper the author recommends that the research community pay attention to recent methodological advances that allow for specification of criteria for the minimal complexity of research designs. The SFT proposes that (a) each cognitive process should be controlled by a separate experimental factor over the manipulated process saliency, and (b) The saliency levels of

all factors should be combined in a full factorial design. The factor's saliency is a manipulation designed to *selectively influence* the speed of a certain cognitive process, so that the process is either speed up or slowed down (by provision of the selective influence). The minimal research design complexity is defined to be composed of 2^n experimental conditions. If your research design of exactly n number of processes has less than 2^n experimental conditions it is likely that the results of such a study will not be conclusive about the organization of the cognitive processes of interest. In that case, you may rather seek external generalizability, which will improve the likelihood of making correct inferences about the cognitive processes, though at an unknown rate.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014.01130/abstract

REFERENCES

Bartlema, A., Lee, M. D., Wetzels, R., and Vanpaemel, W. (2014). A Bayesian hierarchical mixture approach to individual differences: case studies in selective attention and representation in category learning. *J. Math. Psychol.* 59, 132–150. doi: 10.1016/j.jmp.2013.12.002

Benjamin, A. S. (2013). Where is the criterion noise in recognition? (Almost) everyplace you look: comment on kellen, klauer, and singmann (2012). *Psychol. Rev.* 120, 720–726. doi: 10.1037/a0031911

Bradshaw, J. L., and Wallace, G. (1971). Models for processing and identification of faces. *Percept. Psychophys.* 9, 443–448. doi: 10.3758/BF03210249

Broder, A., and Schutz, J. (2009). Recognition ROCs are curvilinear-or are they? on premature arguments against the too-high-threshold model of recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 587–606. doi: 10.1037/a0015279

Brown, S., and Heathcote, A. (2003). Averaging learning curves across and within participants. *Behav. Res. Methods Instrum. Comp.* 35, 11–21. doi: 10.3758/BF03195493

Chechile, R. A. (2009). Pooling data versus averaging model fits for some prototypical multinomial processing tree models. *J. Math. Psychol.* 53, 562–576. doi: 10.1016/j.jmp.2009.06.005

Checkosky, S. F., and Whitlock, D. (1973). Effects of pattern goodness on recognition time in a memory search task. *J. Exp. Psychol.* 100, 341–348. doi: 10.1037/h0035692

Cohen, A. L., Sanborn, A. N., and Shiffrin, R. N. (2008). Model evaluation using grouped or individual data. *Psychon. Bull. Rev.* 15, 692–712. doi: 10.3758/PBR.15.4.692

Dube, C., and Rotello, C. M. (2012). Binary ROCs in perception and recognition memory are curved. *J. Exp. Psychol. Mem. Learn. Cogn.* 38, 130–151. doi: 10.1037/a0024957

Dzhafarov, E. N., Schweickert, R., and Sung, K. (2004). Mental architectures with selectively influenced but stochastically interdependent components. *J. Math. Psychol.* 48, 51–64. doi: 10.1016/j.jmp.2003.12.003

Egeth, H., and Dagenbach, D. (1991). Parallel versus serial processing in visual search: further evidence from subadditive effects of visual quality. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 551–560. doi: 10.1037/0096-1523.17.2.551

Egeth, H. E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Percept. Psychophys.* 1, 245–252. doi: 10.3758/BF03207389

Eidels, A., Townsend, J. T., and Pomerantz, J. R. (2008). Where similarity beats redundancy: The importance of context, higher order similarity, and response assignment. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1441–1463. doi: 10.1037/a0012320

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychol. Bull.* 53, 134–140. doi: 10.1037/h0045156

Estes, W. K., and Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychon. Bull. Rev.* 12, 403–408. doi: 10.3758/BF03193784

Farell, B. (1985). "Same"-"different" judgments: A review of current controversies if perceptual comparisons. *Psychol. Bull.* 98, 419–456. doi: 10.1037/0033-2909.98.3.419

Fific, M., Little, D. R., and Nosofsky, R. M. (2010). Logical rule models of classification response times: a synthesis of mental architecture, random-walk, and decision-bound approaches. *Psychol. Rev.* 117, 309–348. doi: 10.1037/a0018526

Fific, M., Nosofsky, R. M., and Townsend, J. T. (2008a). Information-processing architectures in multidimensional classification: a validation test of the systems factorial technology. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 356–375. doi: 10.1037/0096-1523.34.2.356

Fific, M., and Townsend, J. T. (2010). Information-processing alternatives to holistic perception: identifying the mechanisms of secondary-level holism within a categorization paradigm. *J. Exp. Psychol.* 36, 1290–1313. doi: 10.1037/a0020123

Fific, M., Townsend, J. T., and Eidels, A. (2008b). Studying visual search using systems factorial methodology with target-distractor similarity as the factor. *Percept. Psychophys.* 70, 583–603. doi: 10.3758/PP.70.4.583

Fitousi, D., and Wenger, M. J. (2011). Processing capacity under perceptual and cognitive load: a closer look at load theory. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 781–798. doi: 10.1037/a0020675

Gallistel, C. R. (2009). The importance of proving the null. *Psychol. Rev.* 116, 439–453. doi: 10.1037/a0015251

Gigerenzer, G., and Brighton, H. (2009). Homo heuristics: why biased minds make better inferences. *Top. Cogn. Sci.* 1, 107–143. doi: 10.1111/j.1756-8765.2008.01006.x

Goldstone, R. L., and Medin, D. L. (1994). Time course of comparison. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 29–50. doi: 10.1037/0278-7393.20.1.29

Hilbig, B. E. (2008). Individual differences in fast-and-frugal decision making: Neuroticism and the recognition heuristic. *J. Res. Pers.* 42, 1641–1645. doi: 10.1016/j.jrp.2008.07.001

Hills, T. T., and Hertwig, R. (2012). Tow distinct exploratory behaviors in decisions from experience: comment on gonzalez and dutt (2011). *Psychol. Rev.* 119, 888–892. doi: 10.1037/a0028004

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., and Townsend, J. T. (2014). Systems factorial technology with R. *Behav. Res. Methods* 46, 307–330. doi: 10.3758/s13428-013-0377-3

Ingvalson, E. M., and Wenger, M. J. (2005). A strong test of the dual-mode hypothesis. *Percept. Psychophys.* 67, 14–35. doi: 10.3758/BF03195010

Johnson, S. A., Blaha, L. M., Houpt, J. W., and Townsend, J. T. (2010). Systems factorial technology provides new insights on global–local information processing in autism spectrum disorders. *J. Math. Psychol.* 54, 53–72. doi: 10.1016/j.jmp.2009.06.006

Kellen, D., Klauer, K. C., and Broder, A. (2013a). Recognition and memory models and binary-response ROCs: a comparison by minimum description length. *Psychon. Bull. Rev.* 20, 693–719. doi: 10.3758/s13423-013-0407-2

Kellen, D., Klauer, K. C., and Singmann, H. (2013b). On the measurement of criterion noise in signal detection theory: Reply to Benjamin (2013). *Psychol. Rev.* 120, 727–730. doi: 10.1037/a0033141

Koop, G. J., and Johnson, J. G. (2011). Response dynamics: a new window on the decision process. *Judgm. Decis. Making* 6, 750. Available online at: http://search.proquest.com.ezproxy.gvsu.edu/docview/1011297217?accountid=39473

Lachmann, T., and Geissler, H.-G. (2002). Memory search instead of template matching? Representation-guided inference in same-different performance. *Acta Psychol.* 111, 19–43. doi: 10.1016/S0001-6918(02)00055-0

Lachmann, T., and van Leeuwen, C. (2004). "Memory-guided inference in same-different comparison tasks," in *Psychophysics Beyond Sensation: Laws and Invariants of Human Cognition,* eds C. Kaernbach, E. Schröger, and H. J. Müller (Hilldale, NJ: Lawrence Erlbaum Associates), 199–221

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychon. Bull. Rev.* 15, 1–15. doi: 10.3758/PBR.15.1.1

Little, D. R., Nosofsky, R. M., and Denton, S. E. (2011). Response-time tests of logical-rule models of categorization. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1–27. doi: 10.1037/a0021330

Little, D. R., Nosofsky, R. M., Donkin, C., and Denton, S. E. (2013). Logical rules and the classification of integral-dimension stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 801–820. doi: 10.1037/a0029667

Liu, C. C., and Smith, P. L. (2009). Comparing time-accuracy curves: beyond goodness-of-fit measures. *Psychon. Bull. Rev.* 16, 190–203. doi: 10.3758/PBR.16.1.190

Maddox, W. T. (1999). On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Percept. Psychophys.* 61, 354–375. doi: 10.3758/BF03206893

Marewski, J. N., Gaissmaier, W., Schooler, L. J., Goldstein, D. G., and Gigerenzer, G. (2010). From recognition to decisions: Extending and testing recognition-based models for multialternative inference. *Psychon. Bull. Rev.* 17, 287–309. doi: 10.3758/PBR.17.3.287

Miller, J. O. (1978). Multidimensional same-different judgments: evidence against independent comparisons of dimensions. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 411–422. doi: 10.1037/0096-1523.4.3.411

Myung, I. J., Kim, C., and Pitt, M. A. (2000). Toward and explanation of the power of law artifact: insight from response surface analysis. *Mem. Cognit.* 28, 832–840. doi: 10.3758/BF03198418

Nickerson, R. S. (1965). Response times for "same"-"different" judgments. *Percept. Mot. Skills* 20, 15–18. doi: 10.2466/pms.1965.20.1.15

Nickerson, R. S. (1969). "Same"-"different" response times: a model and a preliminary test. *Acta Psychol. Amsterdamn* 30, 501–517.

Nosofsky, R. M., Little, D. R., Donkin, C., and Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychol. Rev.* 118, 280–315. doi: 10.1037/a0022494

Pachur, T., Hertwig, R., and Wolkewitz, R. (2014). The affect gap in risky choice: affect-rich outcomes attenuate attention to probability information. *Decision* 1, 64–78. doi: 10.1037/dec0000006

Posner, M. I. (1978). *Chronometric Explorations of Mind*. Hillsdale, NJ: Erlbaum.

Posner, M. I., and Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychol. Rev.* 74, 392–409. doi: 10.1037/h0024913

Proctor, R. W. (1981). A unified theory for matching-task phenomena. *Psychol. Rev.* 88, 291–326. doi: 10.1037/0033-295X.88.4.291

Regenwetter, M., Grofman, B., Popova, A., Messner, W., Davis-Stober, C., and Cavagnaro, D. (2009). Behavioural social choice: a status report. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 833–843. doi: 10.1098/rstb.2008.0259

Rouder, J. N., and Lu, J. (2005). An introduction to bayesian hierarchial models with an application in the theory of signal detection. *Psychon. Bull. Rev.* 12, 573–604. doi: 10.3758/BF03196750

Schmidt, K.-D., and Ackermann, B. (1990). "The structure of internal representations and reaction time related matching task phenomena," in *Psychophysical Explorations of Mental Structures,* ed H.-G. Geissler (in collaboration with M. H. Mueller, and W. Prinz) (Göttingen: Hogrefe and Huber), 278–289.

Schweickert, R. (1978). A critical path generalization of the additive factor methods analysis of a stroop task. *J. Math. Psychol.* 26, 105–139. doi: 10.1016/0022-2496(78)90059-7

Schweickert, R. (1985). Separable effects of factors on speed and accuracy: memory scanning, lexical decision, and choice tasks. *Psychol. Bull.* 97, 530–546. doi: 10.1037/0033-2909.97.3.530

Schweickert, R., Giorgini, M., and Dzhafarov, E. (2000). Selective influence and response time cumulative distribution functions in serial-parallel task networks. *J. Math. Psychol.* 44, 504–535. doi: 10.1006/jmps.1999.1268

Schweickert, R., and Townsend, J. T. (1989). A trichotomy: interactions of factors prolonging sequential and concurrent mental processes in stochastic discrete mental (PERT) networks. *J. Math. Psychol.* 33, 328–347. doi: 10.1016/0022-2496(89)90013-8

Soto, F. A., Vucovich, L., Musgrave, R., and Ashby, F. G. (2014). General recognition theory with individual differences: a new method for examining perceptual and decisional interactions with an application to face perception. *Psychon. Bull. Rev.* doi: 10.3758/s13423-014-0661-y. [Epub ahead of print].

Sternberg, S. (1966). High-speed scanning in human memory. *Science* 153, 652–654. doi: 10.1126/science.153.3736.652

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychol. Amsterdam* 30, 276–315. doi: 10.1016/0001-6918(69)90055-9

Sternberg, S. (1998). *Discovering Mental Processing Stages: The Method of Additive Factors. Methods, Models, and Conceptual Issues: An Invitation to Cognitive Science,* Vol. 4. Cambridge, MA: The MIT Press. 703–863.

Sung, K. (2008). Serial and parallel attentive visual searches: evidence from cumulative distribution functions of response times. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1372–1388. doi: 10.1037/a0011852

Townsend, J. T. (1969). "Mock parallel and serial models and experimental detection of these," in *Paper Presented at the Purdue Centennial Symposium on Information Processing* (West Lafayette, IN).

Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. *Percept. Psychophys.* 10, 161–163. doi: 10.3758/BF03205778

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *J. Math. Psychol.* 28, 363–400. doi: 10.1016/0022-2496(84)90007-5

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychol. Sci.* 1, 46–54. doi: 10.1111/j.1467-9280.1990.tb00067.x

Townsend, J. T., and Ashby, F. G. (1983). *The Stochastic Modeling of Elementary Psychological Processes.* Cambridge: Cambridge Univ. Press.

Townsend, J. T., and Fific, M. (2004). Parallel versus serial processing and individual differences in human memory. *Percept. Psychophys.* 66, 953–962. doi: 10.3758/BF03194987

Townsend, J. T., Fific, M., and Neufeld, R. W. J. (2007). "Assessment of mental architecture in clinical/cognitive research," in *Psychological Clinical Science: Papers in Honor of Richard M. McFall,* ed R. W. J. Neufeld (New York, NY: Psychology Press), 223–258.

Townsend, J. T., and Nozawa, G. (1995). Spatio-temporal properties of elementary perception: an investigation of parallel, serial, and coactive theories. *J. Math. Psychol.* 39, 321–359. doi: 10.1006/jmps.1995.1033

Townsend, J. T., and Thomas, R. D. (1994). Stochastic dependencies in parallel and serial models: Effects on systems factorial interactions. *J. Math. Psychol.* 38, 1–34. doi: 10.1006/jmps.1994.1001

Turner, B. M., Dennis, S., and Van Zandt, T. (2013). Likelihood-free bayesian analysis of memory. *Psychol. Rev.* 120, 667–668. doi: 10.1037/a0032458

Wenger, M. J., and Townsend, J. T. (2001). "Faces as gestalt stimuli: process characteristics," in *Computational, Geometric, and Process Perspectives on Facial Cognition: Contexts and Challenges. Scientific Psychology Series,* eds M. J. Wenger and J. T. Townsend (Mahwah, NJ: Lawrence Erlbaum Associates, Publishers), 229–284

Wenger, M. J., and Townsend, J. T. (2006). On the costs and benefits of faces and words: Process characteristics of feature search in highly meaningful stimuli. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 755–779. doi: 10.1037/0096-1523.32.3.755

Yang, C. (2011). Relative saliency in change signals affects perceptual comparison and decision processes in change detection. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1708–1728. doi: 10.1037/a0024257

Yang, C., Chang, T., and Wu, C. (2013). Relative change probability affects the decision process of detecting multiple feature changes. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 1365–1385. doi: 10.1037/a0030693

Yang, H., Fific, M., and Townsend, J. T. (2014). Survivor interaction contrast wiggle predictions of parallel and serial models for an arbitrary number of processes. *J. Math. Psychol.* 59, 82–94. doi: 10.1016/j.jmp.2014.04.001

Using a multinomial tree model for detecting mixtures in perceptual detection

Richard A. Chechile *

Psychology Department, Tufts University, Medford, MA, USA

In the area of memory research there have been two rival approaches for memory measurement—signal detection theory (SDT) and multinomial processing trees (MPT). Both approaches provide measures for the quality of the memory representation, and both approaches provide for corrections for response bias. In recent years there has been a strong case advanced for the MPT approach because of the finding of stochastic mixtures on both target-present and target-absent tests. In this paper a case is made that perceptual detection, like memory recognition, involves a mixture of processes that are readily represented as a MPT model. The Chechile (2004) 6P memory measurement model is modified in order to apply to the case of perceptual detection. This new MPT model is called the Perceptual Detection (PD) model. The properties of the PD model are developed, and the model is applied to some existing data of a radiologist examining CT scans. The PD model brings out novel features that were absent from a standard SDT analysis. Also the topic of optimal parameter estimation on an individual-observer basis is explored with Monte Carlo simulations. These simulations reveal that the mean of the Bayesian posterior distribution is a more accurate estimator than the corresponding maximum likelihood estimator (MLE). Monte Carlo simulations also indicate that model estimates based on only the data from an individual observer can be improved upon (in the sense of being more accurate) by an adjustment that takes into account the parameter estimate based on the data pooled across all the observers. The adjustment of the estimate for an individual is discussed as an analogous statistical effect to the improvement over the individual MLE demonstrated by the James–Stein shrinkage estimator in the case of the multiple-group normal model.

Keywords: signal detection theory, multinomial processing tree models, perceptual learning, mixture detection, shrinkage estimators

1. INTRODUCTION

The title of this special issue implies two very different questions. The first question is: how should perceptual decision-making be modeled? The second question is: how should individual differences be estimated? This paper addresses both of these questions from a perspective that has been informed by research in the area of model-based memory measurement. The recommendations from this perspective result in some novel techniques for examining perceptual detection data.

Signal detection theory (SDT) is the classic method for measuring the perceived strength of a stimulus (Tanner and Swets, 1954; Green and Swets, 1966). The original applications of SDT typically dealt with cases of detecting the presence of a slight intensity increase on a single sensory dimension such as the loudness of white noise or an increase in the brightness of a color patch. The data from these studies are multinomial frequencies that are used to estimate either a signal sensitivity measure (d') associated with the separation between two presumed distributions on a psychological strength continuum, or a non-parametric measure such as A' associated with the area under the receiver-operator characteristic (ROC) curve. For such applications there has been a general consensus that SDT is valid, accurate and useful. SDT has also been extended to the case of multiple dimensions (e.g., Ashby and Townsend, 1986).

Egan (1958) first noted that the target-present versus target-absent test trials used in a yes/no recognition memory study correspond to the signal-present versus signal-absent tests used in a sensory-based signal detection task. It therefore followed that SDT provided a method for measuring memory strength. In fact Macmillan and Creelman (2005) observed that contemporary applications of SDT in the memory area outnumbered the psychophysical applications. Malmberg (2008) and Yonelinas (2002) provide extensive reviews of recognition memory from the perspective of strength-based SDT models. Yet despite the widespread use of the SDT approach toward recognition memory measurement, there also has been substantial criticism of this approach (Chechile, 1978, 2013; Bröder and Schütz, 2009; Kellen et al., 2013). These critics argue instead for the use of multinomial process tree (MPT) models for a variety of reasons. MPT models have a number of desirable statistical properties and can result in measurements of important latent cognitive processes. For example Chechile and Meyer (1976) first used MPT models

for recognition memory data as well as recall data in order to obtain separate probability measures for trace storage and for the retrieval of stored traces, because forgetting was more suitably described in terms of either storage failures or retrieval failures rather than simply a change in "memory strength." The implicit-explicit separation (IES) model is another example of a MPT model rather than a SDT model for memory (Chechile et al., 2012). With the IES model separate probability measures are estimated for explicit storage, implicit storage, fractional storage and non-storage. In these examples, the MPT modeler deliberately prefers to measure cognitive processes other than a SDT strength measure. See Erdfelder et al. (2009) and Batchelder and Riefer (1999) for additional examples of MPT models in psychology.

MPT models are mixture models because with this approach it is assumed that there are possibly different knowledge states that have differential consequence for behavior. For example, sometimes there is enough information stored in memory that the individual can reproduce the target event entirely, provided that the information is accessible at the time of test. But for other tests, the requisite information is either incomplete or totally missing. In the Chechile (2004) 6P model there are separate tree pathways for these two different knowledge states. The overall proportion of traces that are sufficiently stored is defined as the storage probability θ_S. The θ_S parameter is thus a mixture component. Similarly the other parameters in the 6P model are also probabilities and can be regarded as conditional mixture probabilities. Chechile (2013) provided strong evidence for the necessity of considering mixtures for both target-present memory tests as well as for target-absent tests. Evidence was also provided that mixtures are difficult to detect, i.e., data can be generated where a mixture is present but where conventional density plots or quantile–quantile plots fail to detect the mixture. In contrast MPT models are an excellent method for detecting mixtures. Moreover, the absence of a mixture is a special case of a MPT model where the tree paths have probabilities of either 0 or 1[1].

While there is an ongoing debate about SDT and MPT models in the memory literature, there has not been a corresponding contemporary debate in perceptual psychology about the relative merits of SDT and MPT approaches. Yet the possibility of stochastic mixtures is quite plausible for perceptual detection studies, so there are reasons for considering MPT models for perceptual detection.

One rationale for suspecting that there are mixtures comes from the Stevens (1957, 1961) distinction between prothetic and metathetic continua. Stevens (1961); Stevens (p. 41) illustrated a prothetic dimension with loudness and distinguished it from pitch, which is regarded as a metathetic continuum:

> … it is interesting that some of the better known prothetic continua seem to be mediated by an *additive* mechanism at the physiological level, whereas the metathetic continua appear to involve *substitutive* processes at the physiological level. Thus we experience a change in loudness when excitation is added to excitation already present on the basilar membrane, but we note a change in pitch when new excitation is substituted for excitation that has been removed, i.e., the pattern of excitation is displaced

The Stevens distinction stresses the difference between changes in intensity on a single dimension and changes in qualities. A homogeneous process (as opposed to a mixture) is more likely when dealing with a prothetic continuum; although DeCarlo (2002, 2007) has pointed out that trial-by-trial shifts in attention or phasic alertness can produce a stochastic mixture even in a perceptual detection task on a single dimension. However, if the stimuli are complex and possess qualitative features, then stochastic mixtures are even more likely. Consider, for example, a sonar operator attempting to detect any enemy threats. The operator might detect a clear auditory pattern that is a prototypical signal of a particular class of an enemy submarine. With training and experience the sonar operator can be highly skilled in detecting the complex set of features that are associated with an enemy threat; after all perceptual learning is a well established fact (Kellman, 2002). From this framework, the operator might confidently detect a target, not because of a greater strength or intensity, but because the metathetic pattern exhibited by the stimulus is linked through training to a particular type of target. Yet there might be other cases when a threat is present, but the sonar signal is too poorly defined to be identified as a threat. The operator has to guess in these cases. Hence, from this perspective targets stimuli can be considered a mixture of occasions where the target is confidently and correctly identified and other occasion where the operator guesses. A mixture is also possible over all the target-absent cases. For example, a sonar operator might decide that the stimulus is something other than an enemy threat (e.g., a party boat, or a whale), but for other target-absent events the signal might be too poorly defined for the sonar operator to confidently identify. In this paper, a variation of a MPT model will be advanced for perceptual-detection applications in order to capture the possibility that there are mixtures reflected in the data.

The second focus for this paper concerns the relative accuracy of various statistical procedures for modeling individual differences in terms of the key parameters of a perceptual detection MPT model. There is a widespread belief that the maximum likelihood estimates (MLE) of model parameters, done on an individual basis, is the optional method for obtaining estimates of individual differences. This belief is mistaken; there is now considerable evidence that the MLE can be non-optimal and biased for a number of important practical cases. Even in the case of the Gaussian model with more than two conditions, the MLE estimates are known to be biased and "inadmissible" due to the Stein paradox (Stein, 1956; James and Stein, 1961; Efron and Morris, 1977). These insights have led to empirical Bayes, James–Stein estimators, and other shrinkage estimators as improvements to the MLE (Efron and Morris, 1973; Gruber, 1998). Moreover, based on Monte Carlo simulations of multinomial data,

---

[1] Some MPT models have been characterized as threshold models by the authors of the model (e.g., the two high-threshold model of Snodgrass and Corwin, 1988). A threshold is an activation level on an underlying strength continuum that triggers the memory to be in a given state. The assumption of thresholds in MPT models has been vigorously challenged by researchers who prefer a SDT perspective (viz. Dube and Rotello, 2012). However, the concept of a mixture over different knowledge states does not require the assumption of a threshold. For example in the Chechile (2004) 6P model, the knowledge states discussed above are not driven by an underlying strength, but rather it is based simply on the existence or not of specific memory content.

Chechile (2009) found that the averaging of individual parameter estimates resulted in greater error than pooling the multinomial data across individuals and fitting the MPT model once. This finding foreshadows a relatively surprising result that is similar to the James–Stein shrinkage estimate for individual model parameter estimates.

2. THE PERCEPTUAL-DETECTION (PD) MPT MODEL

2.1. DATA STRUCTURE AND TREE MODEL

The Perceptual-Detection (PD) model is essentially the Chechile (2004) 6P model for old/new recognition test trials. The 6P model for storage and retrieval components of memory also has a recall test that is not a part of the perceptual-detection task. The data categories for target-present and target-absent trials as well as the notation for the corresponding population proportions for each response category are shown in **Figure 1**. The PD tree is displayed in **Figure 2**. The MPT model has five parameters; the 6P model had an additional retrieval parameter that is not relevant for perceptual detection. The subscripts for the five parameters have been labeled differently in order to better match the perceptual detection context. The θ_d parameter is the proportion of target-present tests when the operator clearly and confidently detects the target stimulus; this parameter corresponds to the sufficient storage parameter θ_S in the 6P model. The θ_{nt} parameter is the proportion of the target-absent trials when the operator can confidently identify a stimulus that is different than the target; this parameter corresponds to the knowledge-based foil rejection parameter θ_k in the 6P model.

The θ_d and $1 - \theta_d$ parameters are mixing rates for target-present trials. When the target is not clearly detected, the observer can still decide that the stimulus is a target (with conditional probability θ_g) by a secondary process that is simply labeled as a guessing process. Similarly on target-absent tests, the operator (with probability $1 - \theta_{nt}$) fails to confidently identify a non-target but can still guess (with probability $\theta_{g'}$) that the stimulus is more likely a non-target than a target. The two guessing parameters in the PD model are the same as the guessing parameters in the 6P model. Finally the θ_h parameter is a "nuisance" parameter because it is a conditional probability that is only important as a correction for overly confident guessing. This parameter corresponds to the θ_1 parameter in the 6P model.

2.2. PARAMETER ESTIMATION AND A RADIOLOGY EXAMPLE

A great deal is known about the 6P model, and this information directly transfers to the PD model. For example, Chechile (2004)

formally proved that the model is likelihood identifiable, i.e., each configuration of the model parameters results in a unique multinomial likelihood function[2]. Chechile (2004) also showed how the maximum likelihood estimates (MLE) are obtained for the model parameters. In that same paper, an exact Bayesian method for drawing random vectors of values from the posterior distribution was described; the method is called the population parameter mapping (PPM) method (see Chechile, 1998, 2010a). With the PPM method there is a full probability distribution for each model parameter, and there is a probability for the coherence of the model itself. Software also exists for obtaining random vectors from an approximate Bayesian posterior distribution by means of a Markov chain Monte Carlo (MCMC) sampling system[3]. For both the PPM method and the MCMC method, there is a point estimate for each parameter along with

---

[2] See Chechile (1977, 1998, 2004) for a more detailed discussion of model identifiability.

[3] The MCMC method is an implementation of the Metropolis–Hastings algorithm after an initial "burn in" period of 300,000 cycles for sampling each model parameter.

FIGURE 2 | Process tree for the PD model for (A) target-present test trials and (B) target-absent test trials.

Response Categories for Target-Present Trials			
No (high conf.)	No (low conf.)	Yes (low conf.)	Yes (high conf.)
ϕ_1	ϕ_2	ϕ_3	ϕ_4

Response Categories for Target-Absent Tests			
No (high conf.)	No (low conf.)	Yes (low conf.)	Yes (high conf.)
ϕ_5	ϕ_6	ϕ_7	ϕ_8

FIGURE 1 | Data categories and population proportions for the PD model.

a Bayesian posterior probability distribution[4]. The PPM method has several advantages over the MCMC method. First, it does not require a "burn in" period. Second, the posterior distribution is exact as opposed to asymptotically exact. Third, the samples from the posterior distribution are not autocorrelated. Fourth, the PPM method has a probability for the coherence of the model itself.

As an example of parameter estimation for the PD model, let us consider the actual case of the detection characteristics of a single radiologist who was assessing 109 CT scans in order to detect abnormal versus normal scans. Hanley and McNeil (1982) provided the frequencies in four response categories. The categories were labeled as (1) "definitely normal," (2) "probably normal," (3) "probably abnormal," and (4) "definitely abnormal." There were a total of 58 patients who were later determined to be normal, and 51 patients who were determined later to have an abnormality. The frequencies in these four respective categories for the normals (target-absent) are (33, 9, 14, 2)[5]. The corresponding frequencies for the abnormals (target-present) are (3, 3, 12, 33)[6]. The PPM, MCMC, and MLE point estimates for each parameter in the PD model are displayed in **Table 1**.

The PD model point estimates fit the multinomial frequencies very well as indicated by a non-significant goodness-of-fit difference between the observed and predicted frequencies, i.e., $G^2(1) = 0.262$. In addition to the point estimates, the two Bayesian methods have a posterior probability distribution for each model parameter, and these distributions provide a method for testing some important questions about the radiologist. One of the central ideas in the PD model is the concept that there is

a mixture of states for both target-present cases (abnormals) and for target-absent cases (normals). From the posterior distribution of the θ_d parameter, it can be stated that the probability exceeds 0.95 that the θ_d parameter is at least 0.39, i.e., $P(\theta_d > 0.39) > 0.95$. Similarly the posterior distribution for the θ_{nt} parameter results in the high probability statement that θ_{nt} is at least 0.37, i.e., $P(\theta_{nt} > 0.37) > 0.95$.

Using a standard SDT model analysis of the radiological data results in an estimate of $d' = 2.332$ and a ratio of the standard deviations between the signal and noise conditions of $\frac{\sigma_S}{\sigma_N} = 1.409$. This model also fits the data well as indicated by a non-significant difference between the observed and expected frequencies, $G^2(1) = 0.220$. However, the SDT model does not posit that there are mixtures, so the finding that the θ_d and θ_{nt} parameters are reliably different than zero demonstrates that the conventional signal detection model is missing an important feature exhibited by the radiologist. If there were an absence of mixtures, then the PD model would have estimated the θ_d and θ_{nt} parameters as approximately 0.

For MPT models, the mean of the Bayesian posterior distribution for a parameter is usually a different value than the MLE. Chechile (2004) conducted a series of Monte Carlo simulations to see which of these estimates is more accurate for the 6P model; these simulations directly apply to the PD model. For each Monte Carlo run, a random configuration of the model parameters was selected. These parameter values became the true values that are compared later to the estimated values. Also based on the true values, there is a corresponding set of true multinomial cell proportions, i.e., the ϕ_i values in **Figure 1**. From the multinomial likelihood distributions, n random "observations" were drawn for the target-present frequencies and another n random observations were drawn for the target-absent frequencies[7]. Using the cell frequencies, the PPM and MLE parameter estimates are computed. For each estimate there is thus an error score based on the absolute value difference between the estimated value and the true value for that particular Monte Carlo run. For each sample size there was a total of 10,000 Monte Carlo runs. The mean absolute value across the 10,000 runs for PPM and MLE methods are denoted respectively as MAE(ppm) and MAE(mle). The standard deviation of the absolute value errors was also found for both estimation methods. Representative results from these Monte Carlo simulations are shown in **Table 2** for the θ_d parameter.

The Bayesian PPM estimates are more accurate for all the sample sizes. Although the MLE and PPM errors are approaching each other, the rate of approach is relatively slow. Notice that even for the case of $n = 1000$, there is still a smaller standard deviation of the errors for the PPM estimates. The greater accuracy for

---

[4]There is a difference in the prior distributions used for the MCMC method and for the PPM method. For the MCMC approach, a flat prior is assumed for each of the PD model parameters, i.e., the $(\theta_d, \theta_{nt}, \theta_g, \theta_{g'}, \theta_h)$ parameters. However, for the PPM method the prior is a flat distribution for the multinomial cell proportions shown in **Figure 1**, i.e., the (ϕ_i) parameters. The joint posterior distribution for the (ϕ_i) parameters is a product of two Dirichlet distributions. With the PPM method, random samples of (ϕ_i) values are taken from the posterior distribution, and each vector of (ϕ_i) values is mapped to a corresponding vector of the PD model parameters.

[5]There were six cases for the normals where the radiologist used another category called questionable. Three of these cases are assigned here to the second category (probably normal), and three cases were assigned here to the third category (probably abnormal).

[6]There were two CT scans for the abnormals that the radiologist gave the response of questionable. One of these cases was assigned here to the second category, and one was assigned here to the third response category.

Table 1 | PPM, MCMC, and MLE values for the PD model parameters from 109 CT scans by one radiologist reported in the Hanley and McNeil (1982) **study.**

Parameter	PPM	MCMC	MLE
θ_d	0.552	0.555	0.578
θ_{nt}	0.496	0.507	0.523
θ_g	0.734	0.711	0.721
$\theta_{g'}$	0.405	0.438	0.421
θ_h	0.250	0.259	0.227

[7]Given the values for $p_1 = \phi_1$, $p_2 = \phi_1 + \phi_2$, and $p_3 = \phi_1 + \phi_2 + \phi_3$ there are three decision points for randomly assigning a simulated "observation" to one of the four cells. For each simulated observation, a random score is sampled from a uniform distribution on the (0, 1) interval. If the random score is less than p_1, then the observation is for cell 1. If the random score is in the $[p_1, p_2]$ interval, then it is an observation for cell 2. If the random score is in the $[p_2, p_3]$ interval, then it is an observation for cell 3. If the random score is greater or equal to p_3, then it is an observation in cell 4.

Table 2 | The mean absolute value error (MAE) for the θ_d parameter for both the PPM and MLE methods.

n	MAE(ppm)	MAE(mle)	SDE(ppm)	SDE(mle)
10	0.129	0.198	0.090	0.174
20	0.102	0.143	0.076	0.135
30	0.090	0.124	0.070	0.123
40	0.082	0.112	0.064	0.115
50	0.075	0.099	0.061	0.104
100	0.059	0.071	0.050	0.074
300	0.039	0.043	0.035	0.047
600	0.029	0.030	0.027	0.031
1000	0.023	0.023	0.022	0.025

Also shown are the standard deviations of the errors (SDE). Each entry is based on 10,000 Monte Carlo runs from Chechile (2004).

the Bayesian PPM estimates has been also demonstrated for other MPT models (Chechile, 2009, 2010a).

2.3. INTERPRETING THE GUESSING PARAMETERS

The θ_g and $\theta_{g'}$ parameters have actually been used in memory applications since the original storage-retrieval separation paper by Chechile and Meyer (1976). In the memory context it was hypothesized that the guessing parameters involve a mixture of processes that include the possibility of partial storage as well as response bias factors. For memory applications, these parameters are both typically greater than $\frac{1}{2}$, (viz. Chechile and Ehrensbeck, 1983; Chechile and Meyer, 1976; Chechile, 1987, 2004, 2010b; Chechile and Roder, 1998). If the guessing parameters were strictly response bias, then both parameters should not exceed $\frac{1}{2}$, but if there is sometimes partial storage, then that information can be helpful and result in the two guessing parameters exceeding $\frac{1}{2}$. Although the possibility of partial storage was likely, it was not possible to estimate fractional storage with only the yes/no recognition data along with confidence ratings. Later Chechile and Soraci (1999) and Chechile et al. (2012) used different test protocols that enabled the measurement of partial storage. These other MPT models did find evidence for partial storage on some test trials; consequently, the finding of both guessing parameters being greater than $\frac{1}{2}$ is a reasonable outcome.

For the PD model, there is a counterpart to the educated guessing based on partial storage. For the perceptual detection task, there might be occasions where a stimulus is judged more likely a target than not but the quality of the perception is not good enough to constitute a confident classification. On other occasions, the stimulus might be judged more likely a particular "non-target" than a target, but again because the stimulus quality is degraded, the observer is uncertain. For both cases the stimulus is not in a clear detection state, but nonetheless, the person is still able to make informed decisions above a random guessing level.

An interesting special case is when the guessing in both target-present and target-absent conditions are purely response bias, i.e., when $\theta_g = 1 - \theta_{g'}$. However, if there is something like the partial storage found for some memory studies, then the stimulus is more likely to yield a yes response in the target-present condition than in the target-absent condition. Note that the radiologist

measured with the PD model exhibited guessing better than pure response bias because $\theta_g = 0.734 > 1 - \theta_{g'} = 0.595$. These results are consistent with the interpretation that the radiologist was relatively conservative because the doctor guessed that the patient had an abnormality at a rate of 0.595 for the subset of difficult scans from healthy patients. Nonetheless for the subset of difficult scans from patients with an abnormality, the rate for deciding on the abnormal categorization increased to 0.734. Consequently on these more challenging CT scans the physician did have some differential tendency to use the abnormal classification when in fact the CT scan came from a patient with an abnormality.

2.4. PROPERTIES OF THE ROC FOR THE PD MODEL

The Receiver Operator Characteristic (ROC) in SDT is a curved plot of the hit rate versus the false alarm rate. In standard SDT, any point on the ROC is a possible operating point depending on the decision criterion used by the subject. Hence in standard SDT, the ROC is an iso-sensitivity curve. In standard SDT, the points $(0, 0)$ and $(1, 1)$ are on the ROC curve; these points are the extrema. If the subject had no ability to detect the target, and the data are identical in the target-absent and target-present conditions, then the ROC would be the line of slope 1 connecting the extrema. If there is some greater tendency to detect the target in the target-present condition, then in standard SDT the ROC is a smooth curve in the region of the unit square where $y \geq x$.

Empirical ROC plots have been used in numerous experimental papers as a method for comparing theories, but it is challenging to statistically discriminate between models based on only a few points on the empirical ROC. However, given the historical interest in the ROC in psychology, it is instructive to consider the theoretical ROC for the PD model. See **Figure 3** for a general ROC illustration for the PD model. Also see **Table 3** for the PD model equations that are linked to key operating points. The table caption describes the definition of the three discrete points illustrated by the open squares in **Figure 3**, i.e., points P_2, P_3, and P_4. These three points and the two extreme points for the PD model, P_1 and P_5 are a function of the five parameters in the PD model. If $0 < \theta_d < 1$, $0 < \theta_{nt} < 1$, and $\theta_g > 1 - \theta_{g'}$, then the ROC path is along two linear segments. Note that the single-high threshold model discussed by Macmillan and Creelman (2005) is the special case of the PD model when $\theta_{nt} = 0$ and $\theta_g = 1 - \theta_{g'}$. The double-high threshold model also discussed in Macmillan and Creelman (2005) is another special case of the PD model when $\theta_{nt} = \theta_d$ and $\theta_g = 1 - \theta_{g'}$.

To better understand the PD ROC, consider points P_2 and P_3. If we were to define an affirmative response as strictly a "yes" with high confidence, then the corresponding false alarm rate and hit rate would be illustrated by P_2 and have the values corresponding to the prediction equation shown in **Table 3** for that point. Next we redefine an affirmative response as any "yes" response, then the false alarm rate and hit would be illustrated by P_3 and the corresponding prediction equation in **Table 3**. The slope between P_2 and P_3 is denoted as s_{23} and is given as

$$s_{23} = \frac{(1 - \theta_d)\,\theta_g}{(1 - \theta_{nt})\,(1 - \theta_{g'})}, \tag{1}$$

FIGURE 3 | ROC for the PD model. See **Table 3** for a description of the points P_i, $i = 0, \cdots, 6$. The open squares are the three theoretical "operating" points, and the extrema points are $P_1 = (0, \theta_d)$ and $P_5 = (1 - \theta_{nt}, 1)$.
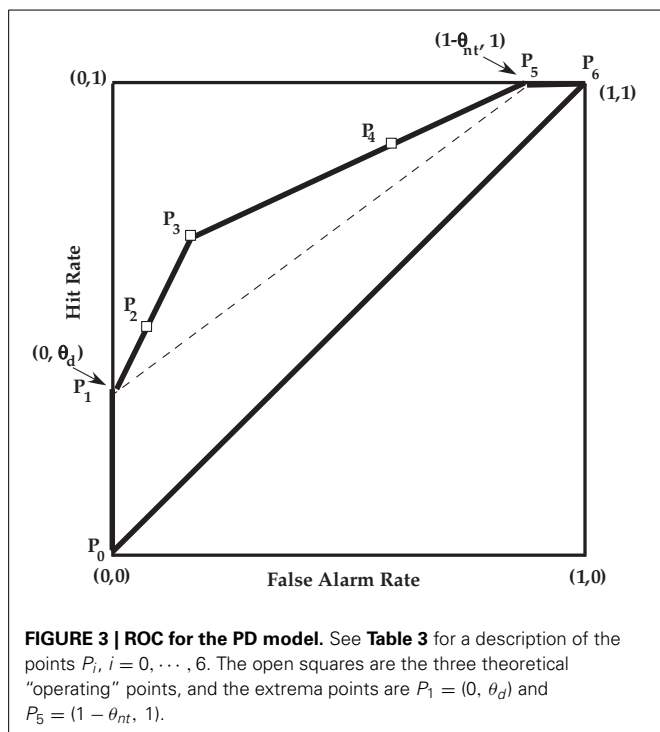
Table 3 | The PD model equations for the key points shown in Figure 3.

Point	x = False alarm	y = Hit
P_0	0	0
P_1	0	θ_d
P_2	$(1 - \theta_{nt})(1 - \theta_{g'})\theta_h$	$\theta_d + (1 - \theta_d)\theta_g\theta_h$
P_3	$(1 - \theta_{nt})(1 - \theta_{g'})$	$\theta_d + (1 - \theta_d)\theta_g$
P_4	$(1 - \theta_{nt})(1 - \theta_{g'}\theta_h)$	$\theta_d + (1 - \theta_d)(1 - (1 - \theta_g)\theta_h)$
P_5	$1 - \theta_{nt}$	1
P_6	1	1

Point P_2 corresponds to the case where a positive response is considered as a high confident yes, but for point P_3 a positive is regarded as any yes response. For point P_4 a positive is considered as any response that is not a high confident no.

and the slope between points P_1 and P_2 is also equal to s_{23}. The linear path from points P_1 and P_3 can be described in terms of a hypothetical variable v that varies on the [0, 1] interval. The false alarm rate x and hit rate y on this path is described by the following equations:

$$x = (1 - \theta_{nt})(1 - \theta_{g'})v, \qquad (2)$$

$$y = \theta_d + (1 - \theta_d)\theta_g v. \qquad (3)$$

The least risky point P_1 corresponds to when $v = 0$. Point P_2 corresponds to the more risky case when $v = \theta_h$. Point P_3 corresponds to the even more risky case of $v = 1$. Of course the only observable points on this path from P_1 to P_3 are P_2 and P_3.

Interestingly the slope from P_3 to P_4 is in general different than the slope from P_1 to P_3. Let us denote the slope from P_3 to P_4 as s_{34}, and it is given as

$$s_{34} = \frac{(1 - \theta_d)(1 - \theta_g)}{(1 - \theta_{nt})\theta_{g'}}. \qquad (4)$$

It is also the case that the slope from P_4 to P_5 is also equal to s_{34}. Moreover, the linear path from P_3 to P_5 can be described in terms of another hypothetical variable w that varies from 0 to 1 as the risk increases. The false alarms x and hits y on this path is characterized by the following equations:

$$x = (1 - \theta_{nt})(1 - \theta_{g'} + \theta_{g'} w), \qquad (5)$$

$$y = \theta_d + (1 - \theta_d)\theta_g + (1 - \theta_d)(1 - \theta_g)w. \qquad (6)$$

The P_3 point corresponds to $w = 0$; whereas the P_4 point corresponds to $w = 1 - \theta_h$ and P_5 corresponds to $w = 1$.

Figure 4 illustrates the PD model ROC path from one extreme point to the other in terms of the v and w variables. As v varies from 0 to 1 it traces points on the P_1 to P_3 line as stipulated by Equations (2, 3). Similarly as w varies from 0 to 1, (Equation 5) and (Equation 6) traces points on the P_3 to P_5 line. Notice that θ_h determines the separation from each of the two extreme ends. This feature is a property of the PD model because there is a common parameter of incorrectly using the high confidence rating when guessing regardless if the guessing is done in either the target-present condition or the target-absent condition. Chechile (2004) also presented another identifiable memory MPT model where there are separate parameters for over confidence when using the "yes" response (θ_2) versus over confidence when using the "no" response (θ_1). This model is the 7B model. Other than the difference in the handling of over confidence, the 7B and 6P models are identical, i.e., the 6P model is the special case of 7B where $\theta_h = \theta_1 = \theta_2$. Model 7B can also be applied to the perceptual detection task (lets denote that model as the PD* model). In the PD* model the θ_2 parameter determines the location for the v variable for the P_2 point, and the θ_1 parameter determines the separation for the w variable from the maximum of 1. Hence, the spacing for the points on the $v - w$ plot is different for the PD* model than the spacing shown in **Figure 4** for the PD model.

In general the slope from P_3 to P_5 is less than the slope from P_1 to P_3. Given Equations (1), and (4) the ratio of the slopes can be written as

$$r = \frac{s_{35}}{s_{13}} = \frac{(1 - \theta_g)(1 - \theta_{g'})}{\theta_g \theta_{g'}}. \qquad (7)$$

If there is some partial or degraded perception, then the tendency to respond "yes" is at least equal or greater in the target-present condition as it is in the target-absent condition. It follows that

$$\frac{\theta_g}{1 - \theta_g} \geq \frac{1 - \theta_{g'}}{\theta_{g'}}. \qquad (8)$$

It also follows from Equations (7, 8) that $r \leq 1$. Consequently, if $\theta_g > 1 - \theta_{g'}$, then the slope from P_1 to P_3 is larger than the slope

FIGURE 4 | Illustration of the relative position of the *v* and *w* variables that determine the points on the PD ROC. See **Figure 3** for the definition of the points.
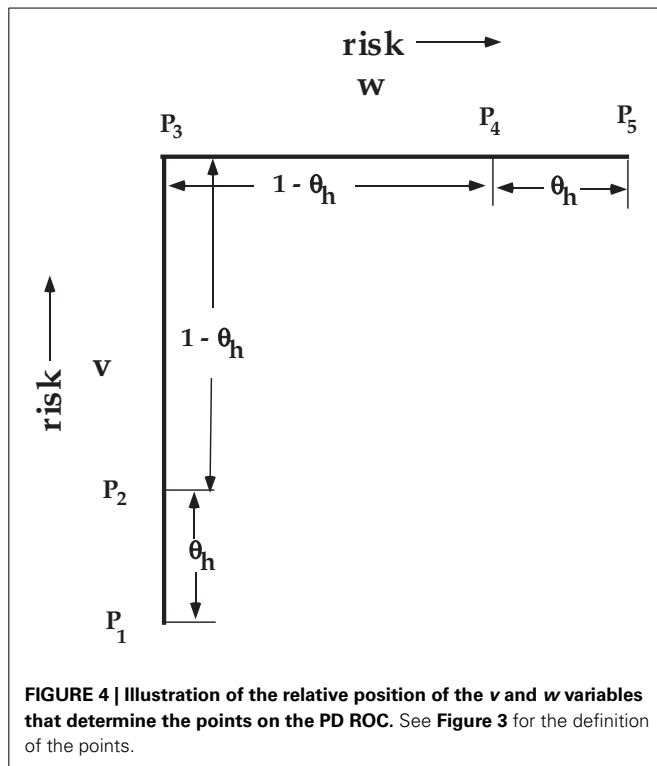
from P_3 to P_5. The case where $r = 1$ corresponds to when $\theta_g = 1 - \theta_{g'}$ or when there is the same "yes" guessing in the target-present condition as in the target-absent condition. In this special case, there is no partial detection, and the ROC does not have two linear components, but there is instead a single line of slope $\frac{1-\theta_d}{1-\theta_{nt}}$ between P_1 and P_5.

The area under the ROC has been used as a measure of sensitivity in standard SDT. It is straightforward to show that area A_c between the P_1-P_5 dashed line in **Figure 3** and the main diagonal line of $y = x$ is $\frac{1}{2}(\theta_d + \theta_{nt} - \theta_d\theta_{nt})$[8]. This region is a function of certain perceptual detection and does not depend on guessing. Because the total area in the upper half of the unit square where $y > x$ is $\frac{1}{2}$, it is advantageous to multiply A_c by 2, so that the area measure of certain detection is placed on a 0 to 1 scale. This measure is defined as a certain detection D_c, and

$$D_c = \theta_d + \theta_{nt} - \theta_d\theta_{nt}. \tag{9}$$

The area of the $P_1 P_3 P_5$ triangle is a function of guessing. This area is denoted as A_g, and it can be found from Heron's formula, i.e., $A_g = \frac{1}{2}(1 - \theta_{nt})(1 - \theta_d)[\theta_g - (1 - \theta_{g'})]$. We can put this measure of effective guessing on a 0 to 1 scale by defining $D_g = 2A_g$ or

$$D_g = (1 - \theta_{nt})(1 - \theta_d)[\theta_g - (1 - \theta_{g'})]. \tag{10}$$

---

[8]Note that the total area above the main diagonal is $\frac{1}{2}$, and the area above the dashed line is $\frac{1}{2}(1 - \theta_d)(1 - \theta_{nt})$, so A_c can be determined by subtracting these quantities.

Thus the total detection measure can be defined as twice the area between the ROC and the main diagonal; this metric is $D = D_c + D_g$ or

$$D = \theta_d + \theta_{nt} - \theta_d\theta_{nt} + (1 - \theta_{nt})(1 - \theta_d)[\theta_g - (1 - \theta_{g'})], \tag{11}$$

As an example, let us compute these area-based metrics for the radiological data discussed in section **2.2**. Using PPM estimates for θ_d and θ_{nt}, it follows from Equation (9) that $D_c = 0.774$. The corresponding D_g measure from Equation (10) is 0.031, so the overall D metric is 0.805.

Although the detection measure D is on a proportional basis, it is, nonetheless, a confounded measure because it does not delineate how the detection was achieved. For example suppose that $\theta_{nt} = 0.805$ and $\theta_d = 0$, then the resulting D value would be the same as for the radiologist discussed above. Clearly the hypothetical observer with $\theta_d = 0$ and $\theta_{nt} = 0.805$ would be very good at recognizing a normal CT scan, but would not be capable of detecting an abnormal scan, which would be a rather serious problem for the diseased patients of that hypothetical radiologist! Consequently, the area-based D metric, along with its component metrics of D_c and D_g, is less informative as the original PD model parameters. The detection of the target increases with the value of the θ_d parameter, and the identification of a non-target increases with the value of the θ_{nt} parameter. Those two types of detection can be quite different. It is also informative to know how the observer does for the unclear cases where there is guessing. The D metric does not pull out the many different perceptual and decision-making characteristics of the observer's behavior. Also the standard SDT metrics of d' and the ratio of the standard deviations do not extract the different properties of the observer's perceptual-detection performance.

3. INDIVIDUAL DIFFERENCE ESTIMATION FOR THE PD MODEL

A fundamental issue that arises in mathematical psychology is the basis for fitting a model. One method is to fit the model separately for each individual and to average individual estimates for the group average. Another method is to aggregate the data across a group of individuals for a particular experimental condition and then fit the model once for that condition[9]. The estimates from these two approaches differ. Although there are applications where each of these pure approaches is reasonable, in this paper a hybrid of these two methods will be recommended. Consequently, the answer to the question as to how to fit a model depends on the purpose of the analysis.

There are several contexts that necessitate the fitting of the model on an individual basis. For example, if the model is a non-linear function of an independent variable, then many investigators have demonstrated that group-averaged data can result in biased fits (Estes, 1956; Sigler, 1987; Ashby et al., 1994). Also

---

[9]A third approach also exists for obtaining individual and group effects by means of a hierarchical Bayesian model similar to the analysis developed for MPT models by Klauer (2010). This method is computationally challenging, and it has not yet been assessed to see if it has improved accuracy relative to the simple model advanced in the present paper.

the theoretical issue being examined can require that the analysis be done on an individual basis. For example, Chechile (2013) examined the memory hazard function to see if there was evidence of a mixture over stimuli. Had that analysis been done on a grouped-data basis, then any results suggesting a mixture could have been a mixture over individuals with different memory properties instead of a mixture over stimuli.

There are also cases when pooling the data prior to the model fit is the preferred analysis (Cohen et al., 2008; Chechile, 2009). Chechile (2009), for example, studied four prototypic MPT models with an extensive series of Monte Carlo simulations in order to examine the relative accuracy of averaging versus data pooling. For any given Monte Carlo run, a group of n_g simulated "subjects" with slightly different true values for the model parameters was constructed, and for each artificial subject there were n_r "observations" that were randomly sampled from the appropriate multinomial likelihood distribution[10]. Based on this set of simulated outcome frequencies, the model was fit in two different ways: (1) the averaging method and (2) the data-pooling method. For the averaging method the MPT model was fit separately for each of the n_g subjects, and these estimates were averaged to obtain an estimate for each model parameter. For an arbitrary model parameter, θ_x, the group average estimate is $\bar{\theta}_x = \frac{1}{n_g} \sum_{i=1}^{n_g} \hat{\theta}_{xi}$ where $\hat{\theta}_{xi}$ is the parameter estimate for the ith subject. For any Monte Carlo run, the absolute value difference was computed between $\bar{\theta}_x$ and the true mean for that parameter $\theta_x(true) = \frac{1}{n_g} \sum_{i=1}^{n_g} \theta_{xi}(true)$. This difference is taken as the error for the averaging method for that one Monte Carlo run. The process was then repeated so that in total there were 1000 separate Monte Carlo runs for each combination of n_g and n_r. Across these separate Monte Carlo runs the model parameters were varied, so the model was simulated over a vast set of configurations of the parameters. The overall error for the averaging method is the mean error across the 1000 Monte Carlo data sets for each combination of n_g and n_r. For the identical data as described above, a corresponding error was also found for the pooling method. For the pooling method the frequencies in each multinomial response category was summed across the n_g subjects in a group, and the model was fit once with the pooled data. The estimate based on pooling for the jth simulated data set is denoted as $\hat{\theta}_{xj}(pooled)$. The absolute value difference between this estimate and the true value for that run is the pooling error for the jth Monte Carlo data set, and mean error across all 1000 data sets is the overall error for the pooling method[11]. For all four models reported in Chechile (2009) and for most combinations of n_g and n_r, the mean error for the pooling method was less than the corresponding error obtained for the averaging method[12]. Consequently, Chechile (2009) reported a pooling

advantage score that was the difference between the mean averaging error and the mean pooling error. For example, a positive value for the pooling advantage score of 0.07 means that the averaging mean error was larger by 0.07 than the corresponding pooling error. A negative pooling advantage score would mean that the averaging method had less error than the pooling method.

One of the models examined in Chechile (2009) was a four-cell MPT model that is identical to the structure of the process trees for either the target-present or the target-absent test conditions with the PD model. Consequently, those Monte Carlo simulations directly apply to the PD model. **Table 4** provides a condensed summary of the Monte Carlo results from Chechile (2009). The θ_d parameter in **Table 4** corresponds to the θ_S parameter in Model A; whereas θ_g and θ_h, respectively, correspond to the θ_g and θ_1 parameters in Model A.

The pooling advantage scores in **Table 4** exhibit a number of interesting properties that were also found with the other MPT models. First, the pooling advantage scores are positive indicating that there is greater accuracy for the pooling method. Second, although the magnitude of the pooling advantage decreases with the number of observations per subject (n_r), there is still a non-trivial advantage for pooling even when $n_r = 400$. It is challenging to do an experiment with large values for n_r. For example, a replication number of 50 is larger than all but two of the memory studies reported from my laboratory. Consequently, the idea of running a large number of replication trials per subject is not a practical option. Third, the size of the pooling advantage increases with group size n_g. This effect is due to the fact that the error for the pooling method decreases rapidly with increasing group size; whereas the error for the averaging method slowly decreases with increasing n_g,

---

[10] Each individual was within ± 0.03 of the group mean.

[11] This whole procedure of estimating the model with both the averaging and pooling method was done for both PPM and MLE estimates for each of the four typical MPT models.

[12] Only eight cases out of 640 cases reported in Chechile (2009) had greater error for the pooling method, and all of these exceptions were when the MLE was used. Generally the MLE was not the optimal estimator for the model parameters because the corresponding Bayesian PPM estimator had greater accuracy.

Table 4 | The difference in mean error between averaging and pooling for n_g individuals in a group and for n_r trials in the target-present condition.

n_g n_r	Pooling Advan. score		
	θ_d	θ_g	θ_h
20 20	0.069	0.076	0.078
20 50	0.045	0.050	0.051
20 100	0.034	0.035	0.034
20 400	0.015	0.013	0.013
40 20	0.078	0.086	0.087
40 50	0.054	0.059	0.059
40 100	0.037	0.040	0.040
40 400	0.017	0.015	0.015
80 20	0.087	0.096	0.098
80 50	0.059	0.065	0.064
80 100	0.043	0.043	0.045
80 400	0.020	0.016	0.016

This difference is a pooling advantage score. Positive values indicate less error for the pooling method. Monte Carlo simulations from Chechile (2009).

so the net effect is that the pooling advantage score increases with n_g.

It might not seem intuitive as to why the pooling of data results in superior estimates for the group mean. This result is more reasonable when viewed from a Bayesian perspective. From Bayes theorem it does not matter if the data are examined in aggregate or one observation at a time, provided that the same starting prior probability is used. Suppose we use a uniform distribution as the prior distribution for each combination of the parameters $(\theta_d, \theta_g, \theta_h)$. Let us call this prior the "vague" prior. Furthermore suppose we examine the model parameters for the first individual in the group via Bayes theorem to yield a posterior distribution. The posterior distribution after the first individual should then be the prior distribution for examining the data for the second subject, i.e., it is no longer appropriate to maintain the vague prior after examining the first subject. Similarly the prior distribution for Subject 3 should be the posterior distribution after considering the first two subjects. This one-subject-at-a-time method eventually yields a posterior distribution that is the same as the posterior distribution achieved by pooling the multinomial categories and applying Bayes theorem once. Had the Bayesian analyst used a vague prior for each of the n_g subjects and averaged the estimates, then the analysis would not be consistent in the application of Bayes theorem. The averaging of separate estimates is not an operation by which probability distributions are revised via Bayes theorem. In terms of this framework, the findings in **Table 4** are quite reasonable. The pooling method should be more accurate, and the pooling advantage should grow with the size of the group.

Despite the above demonstration of a pooling advantage for estimating the group mean, it is still an open question as to what should be the basis for estimating the model parameters for an individual. Two choices seem reasonable. One method is simply to use the data for just the individual, e.g., for the θ_d parameter it would be $\hat{\theta}_{d\,i}$ for the ith observer. For the second method the data for the individual is used but there is a fixed correction so that the mean across all observers is equal to the pooled estimate for the group. For the θ_d parameter this estimate is denoted as $\hat{\theta}_{d\,i}^{(a)}$ and is defined as

$$\hat{\theta}_{d\,i}^{(a)} = \hat{\theta}_d(pooled) - \bar{\theta}_d + \hat{\theta}_{d\,i}. \tag{12}$$

Note that the two methods have estimates that are perfectly correlated because the adjusted estimate $\hat{\theta}_{d\,i}^{(a)}$ is a constant plus the individual estimate $\hat{\theta}_{d\,i}$. The constant correction term is equal to $\hat{\theta}_d(pooled) - \bar{\theta}_d$. The correction makes the mean of the adjusted estimates equal to the pooling method estimate because

$$\frac{1}{n_g} \sum_{i=1}^{n_g} \hat{\theta}_{d\,i}^{(a)} = \hat{\theta}_d(pooled) - \bar{\theta}_d + \bar{\theta}_d = \hat{\theta}_d(pooled).$$

The estimate based on Equation (12) is similar in principle to a James–Stein estimator used for the linear model for Gaussian random variables because the estimate for the individual is shifted based on properties of the group.

Another Monte Carlo simulation was designed for a widely different group of simulated observers in order to assess the relative accuracy of the two methods for estimating the parameters for individuals. The group consisted of 10 observers for each of the 3×3 combinations of values for θ_d and θ_{nt}. The three values were 0.2, 0.5, and 0.8. For each of the 90 simulated observers the values for θ_h were randomly selected from a beta distribution with coefficients of 2 and 4, and the θ_g and $\theta_{g'}$ parameters were randomly selected from a beta distribution with coefficients of 28 and 14. Consequently true scores were established for each simulated observer. For each observer, 20 simulated observations were randomly sampled for the target-present condition, and another 20 observations were randomly sampled for the target-absent condition. These observations were based on the appropriate multinomial likelihood distribution for each subject. The PD model was then estimated by each method described above. Because θ_d and θ_{nt} are the two key parameters of interest in the PD model, the root mean square (rms) error was found between the true score point $\{\theta_{d\,i}(true), \theta_{nt\,i}(true)\}$ and the estimated point for the individual $\{\hat{\theta}_{d\,i}, \hat{\theta}_{nt\,i}\}$. The rms error for the adjusted score point $\{\hat{\theta}_{d\,i}^{(a)}, \hat{\theta}_{nt\,i}^{(a)}\}$ was also found. The rms errors for the individual and the adjusted method are respectively 0.1671 and 0.1385. Thus, the adjusted estimates based on Equation (12) resulted in a 17% reduction in the rms error. This simulation illustrates the improvement in the accuracy of model estimation by the use of the adjusted score method.

4. DISCUSSION

In this paper the Chechile (2004) 6P memory measurement model was modified and applied to perceptual detection. The resulting PD model is a MPT model that has two mixture rate parameters (θ_d and θ_{nt}) that measure the proportion of times that the observer confidently detects something that belongs to an identifiable category. The categories are different for targets and non-targets, but in both cases something is being identified. The measurement of these detection rates is an important part of the psychometric assessment of perceptual performance. The PD model also has three other parameters that come into play when the observer is unable to confidently classify the stimulus.

The PD model differs from standard SDT on the issue of stochastic mixtures. MPT models, like the PD model, are essentially probability mixture models. In contrast, SDT developed in the context of assuming separate but homogeneous distributions for target-present and target-absent conditions. The success of the PD model in accounting for the radiological judgments described earlier in this paper occurred because the PD model was sensitive to the fact the radiologist was able to know sometimes that a CT scan was normal and to know at other times that a CT scan revealed an identifiable abnormality. This attribute of categorical and sophisticated perception is not an isolated property of experts. More than 120 years ago William James discussed the importance of perceptual learning; in fact perception according to James differed from a pure sensation because of the information that the person associates and adds to the sensation (James, 1890). There is now a vast literature describing the improvement in

perception with practice (Kellman, 2002). With experience people can develop refined perceptual categories that sharpen their ability to process and to interpret stimuli.

It is noteworthy that the prototypic experiments in the early history of SDT used stimuli that were designed to be featureless and varied on only a single prothetic intensity dimension. For example the stimulus-absent stimulus for some experiments was white noise; whereas the target-present stimulus was a louder white noise (Tanner et al., 1956). Perceptual categories and perceptual learning is limited for such impoverished stimuli. SDT is expected to be quite successful for such applications, but SDT is expected to be problematic when stimuli possess rich perceptual features and when the observer has some experience with the class of stimuli. For those applications, the PD model would be a more suitable cognitive psychometric tool for assessing the properties of the observer.

The PD model is a minimalistic model that intentionally eschews delineating any specific cognitive representation of the stimulus. Like other MPT models, there are probability measures for specific states. The states for the PD model are: (1) a state of certain target recognition, which occurs on θ_d proportion of the target-present trials, and (2) the state of certain identification of something other than a target, which occurs on θ_{nt} proportion of the target-absent trials. These probability measures provide for a characterization of the observer's detection ability.

MPT models have many desirable statistical properties and can be estimated by a variety of methods. Monte Carlo simulations with large sample sizes demonstrated that the MLE and the Bayesian posterior mean for the PD model were very close, but the accuracy of these estimates differed more substantially for smaller sample sizes. When the estimates differ, the Bayesian mean was found to be more accurate. In addition, an improved estimate was found for the individual observer when the estimate based on the individual's data was adjusted. The adjustment was a fixed amount for all observers, and it equated the mean of the adjusted scores to the mean of the estimate based on pooled data. This adjustment was discussed as an analogous adjustment to the James–Stein shrinkage improvements to the MLE found for the multiple-group Gaussian model.

REFERENCES

Ashby, F. G., Maddox, W. T., and Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychol. Sci.* 5, 144–151. doi: 10.1111/j.1467-9280.1994.tb00651.x

Ashby, F. G., and Townsend, J. T. (1986). Varieties of perceptual independence. *Psychol. Rev.* 93, 154–179. doi: 10.1037/0033-295X.93.2.154

Batchelder, W. H., and Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychon. Bull. Rev.* 6, 57–86. doi: 10.3758/BF03210812

Bröder, A., and Schütz, J. (2009). Recognition ROCs are curvilinear - or are they? On premature arguments against the two-high-threshold model of recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 587–606. doi: 10.1037/a0015279

Chechile, R. A. (1977). Likelihood and posterior identification: implications for mathematical psychology. *Br. J. Math. Stat. Psychol.* 30, 177–184. doi: 10.1111/j.2044-8317.1977.tb00737.x

Chechile, R. A. (1978). Is d'_S a suitable measure of recognition memory "strength"? *Bull. Psychon. Soc.* 12, 152–154. doi: 10.3758/BF03329655

Chechile, R. A. (1987). Trace susceptibility theory. *J. Exp. Psychol. Gen.* 116, 203–222. doi: 10.1037/0096-3445.116.3.203

Chechile, R. A. (1998). A new method for estimating model parameters for multinomial data. *J. Math. Psychol.* 42, 432–471. doi: 10.1006/jmps.1998.1210

Chechile, R. A. (2004). New multinomial models for the Chechile-Meyer task. *J. Math. Psychol.* 48, 364–384. doi: 10.1016/j.jmp.2004.09.002

Chechile, R. A. (2009). Pooling data versus averaging model fits for some prototypical multinomial processing tree models. *J. Math. Psychol.* 53, 562–576. doi: 10.1016/j.jmp.2009.06.005

Chechile, R. A. (2010a). A novel Bayesian parameter mapping method for estimating the parameters of an underlying scientific model. *Commun. Stat. Theor. Methods* 39, 1190–1201. doi: 10.1080/03610920902859615

Chechile, R. A. (2010b). Modeling storage and retrieval processes with clinical populations with applications examining alcohol-induced amnesia and Korsakoff amnesia. *J. Math. Psychol.* 54, 150–166. doi: 10.1016/j.jmp.2009.03.006

Chechile, R. A. (2013). A novel method for assessing rival models of recognition memory. *J. Math. Psychol.* 57, 196–214. doi: 10.1016/j.jmp.2013.07.002

Chechile, R. A., and Ehrensbeck, K. (1983). Long-term storage losses: a dilemma for multistore models. *J. Gen. Psychol.* 109, 15–30. doi: 10.1080/00221309.1983.9711505

Chechile, R. A., and Meyer, D. L. (1976). A Bayesian procedure for separately estimating storage and retrieval components of forgetting. *J. Math. Psychol.* 13, 269–295. doi: 10.1016/0022-2496(76)90022-5

Chechile, R. A., and Roder, B. (1998). "Model-based measurement of group differences: an application directed toward understanding the information-processing mechanism of developmental dyslexia," in *Perspectives on Fundamental Processes in Intellectual Functioning: Volume 1- A Survey of Research Approaches*, eds S. Soraci and W. J. McIlvane (Stamford, CN: Ablex), 91–112.

Chechile, R. A., Sloboda, L. N., and Chamberland, J. R. (2012). Obtaining separate measures for implicit and explicit memory. *J. Math. Psychol.* 56, 35–53. doi: 10.1016/j.jmp.2012.01.002

Chechile, R. A., and Soraci, S. A. (1999). Evidence for a multiple-process account of the generation effect. *Memory* 7, 483–508. doi: 10.1080/741944921

Cohen, A. L., Sanborn, A. N., and Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychon. Bull. Rev.* 15, 692–712. doi: 10.3758/PBR.15.4.692

DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: theoretical developments with applications to recognition memory. *Psychol. Rev.* 109, 710–721. doi: 10.1037/0033-295X.109.4.710

DeCarlo, L. T. (2007). The mirror effect and mixture signal detection theory. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 18–33. doi: 10.1037/0278-7393.33.1.18

Dube, C., and Rotello, C. M. (2012). Binary ROCs in perception and recognition memory are curved. *Exp. Psychol. Learn. Mem. Cogn.* 38, 130–151. doi: 10.1037/a0024957

Efron, B., and Morris, C. (1973). Stein's estimation rule and its competitors. *J. Am. Stat. Assoc.* 65, 117–130. doi: 10.1080/01621459.1973.10481350

Efron, B., and Morris, C. (1977). Stein's paradox in statistics. *Sci. Am.* 236, 119–127. doi: 10.1038/scientificamerican0577-119

Egan, J. P. (1958). *Recognition Memory and the Operating Characteristic*. Technical Note AFCRC-TN-58-51. Hearing and Communication Laboratory. Indiana University.

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., and Nadarevic, L. (2009). Multinomial processing tree models: a review of the literature. *J. Psychol.* 217, 108–124. doi: 10.1027/0044-3409.217.3.108

Estes, W. K. (1956). The problem of inference from curves based on grouped data. *Psychol. Bull.* 53, 134–140. doi: 10.1037/h0045156

Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: Wiley.

Gruber, M. H. J. (1998). *Improving Efficiency by Shrinkage: The James-Stein and Ridge Regression Estimators*. New York, NY: Marcel Dekker Inc.

Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under the Receiver Operating Characteristic (ROC) curve. *Radiology* 143, 29–36.

James, W. (1890). *The Principles of Psychology*, Vol. 2. New York, NY: Holt and Company. doi: 10.1037/11059-000

James, W., and Stein, C. (1961). "Estimation with quadratic loss," in *Proceeding of the Fourth Berkeley Symposium on Mathematics and Statistics* (Berkeley, CA), 361–379.

Kellen, D., Klauer, K. C., and Bröder, A. (2013). Recognition memory models and binary-response ROCs: a comparison by minimum description length. *Psychon. Bull. Rev.* 20, 693–719. doi: 10.3758/s13423-013-0407-2

Kellman, P. J. (2002). "Perceptual learning," in *Steven's Handbook of Experimental Psychology*. Learning, motivation, and emotion, Vol. 3, eds R. Gallistel and H. Pashler (New York: Wiley), 259–299.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: a latent-trait approach. *Psychometrika* 75, 70–75. doi: 10.1007/s11336-009-9141-0

Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide*, 2nd Edn. Mahwah, NJ: Erlbaum.

Malmberg, K. J. (2008). Recognition memory: a review of the critical findings and an integrated theory for relating them. *Cogn. Psychol.* 57, 335–384. doi: 10.1016/j.cogpsych.2008.02.004

Sigler, R. S. (1987). The perils of averaging data over strategies: an example from children's addition. *J. Exp. Psychol. Gen.* 121, 278–304.

Snodgrass, J. G., and Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *J. Exp. Psychol. Gen.* 117, 34–50. doi: 10.1037/0096-3445.117.1.34

Stein, C. (1956). "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in *Proceedings of the Third Berkeley Symposium on Mathematics, Statistics and Probability* (Berkeley, CA), 197–206.

Stevens, S. S. (1957). On the psychophysical law. *Psychol. Rev.* 64, 163–181. doi: 10.1037/h0046162

Stevens, S. S. (1961). Toward a resolution of the Fechner-Thurstone legacy. *Psychometrika* 26, 35–47. doi: 10.1007/BF02289683

Tanner, W. P., and Swets, J. A. (1954). A decision-making theory of visual detection. *Psychol. Rev.* 61, 401–409. doi: 10.1037/h0058700

Tanner, W. P., Swets, J. A., and Green, D. M. (1956). *Some General Properties of the Hearing Mechanism.(Electronic Defense Group, Technical Report No. 30)*. Ann Arbor, MI: University of Michigan.

Yonelinas, A. P. (2002). The nature of recollection and familarity: a review of 30 years of research. *J. Mem. Lang.* 46, 441–517. doi: 10.1006/jmla.2002.2864

Time-varying boundaries for diffusion models of decision making and response time

Shunan Zhang[1]*, Michael D. Lee[1], Joachim Vandekerckhove[1], Gunter Maris[2] and Eric-Jan Wagenmakers[2]

[1] Department of Cognitive Sciences, University of California, Irvine, Irvine, CA, USA
[2] Psychological Methods, University of Amsterdam, Amsterdam, Netherlands

Diffusion models are widely-used and successful accounts of the time course of two-choice decision making. Most diffusion models assume constant boundaries, which are the threshold levels of evidence that must be sampled from a stimulus to reach a decision. We summarize theoretical results from statistics that relate distributions of decisions and response times to diffusion models with time-varying boundaries. We then develop a computational method for finding time-varying boundaries from empirical data, and apply our new method to two problems. The first problem involves finding the time-varying boundaries that make diffusion models equivalent to the alternative sequential sampling class of accumulator models. The second problem involves finding the time-varying boundaries, at the individual level, that best fit empirical data for perceptual stimuli that provide equal evidence for both decision alternatives. We discuss the theoretical and modeling implications of using time-varying boundaries in diffusion models, as well as the limitations and potential of our approach to their inference.

Keywords: accumulator model, collapsing bounds, model equivalence, sequential sampling, first-passage time

1. INTRODUCTION

Being able to make a timely choice between two alternatives is a cornerstone of human cognition, and a long-standing focus of experimentation and theorizing in cognitive psychology. One widely used approach to modeling the time course of decision making comes from the class of *sequential sampling* models (Link and Heath, 1975; Ratcliff, 1978; Vickers, 1979; Luce, 1986; Busemeyer and Townsend, 1993; Usher and McClelland, 2001; Ratcliff and McKoon, 2008). In these models, people are assumed to gather information, piece by piece, until they have accrued enough evidence in favor of one or other alternative to justify that decision. The most prominent and popular sequential sampling models are *diffusion models*, which make the assumption that the samples of evidence come from a Gaussian distribution, and are accumulated according to a random walk that becomes a diffusion process as the time-step between samples approaches a limit of zero (Ratcliff, 1980, 1985, 1988, 2013; Ratcliff and Rouder, 1998, 2000; Ratcliff et al., 1999).

The basic diffusion model assumptions and operation are shown graphically in **Figure 1A**. Evidence values are sampled from a Gaussian with mean μ and standard deviation σ. These values are accumulated in a single tally until the tally reaches either the upper or lower boundaries shown by solid black lines. Once the tally reaches a boundary, evidence accumulation stops, and the model makes the decision associated with the boundary that was reached, with a response time corresponding to the number of samples taken. **Figure 1A** shows 10 example tallies by thin gray lines. It also shows by histograms at the boundaries the distribution of response times for each decision.

When applied to account for human decision-making, diffusion models are usually extended beyond the basic form shown in **Figure 1A**. Most often, additional parameters are added, introducing variability to the evidence accrual process, or incorporating encoding and retrieval processes, or processes that cause leakage or drift in the tallies (e.g., Ratcliff, 1978; Busemeyer and Townsend, 1992; Usher and McClelland, 2001; Ratcliff and McKoon, 2008). In these expanded forms, diffusion models have been widely applied to model human decision-making for a variety of tasks, including: many simple perceptual decisions like coherent motion detection, line length comparison, and brightness discrimination (e.g., Ratcliff and Rouder, 1998; Ratcliff et al., 2003); simple cognitive tasks, like lexical decision (e.g., Ratcliff et al., 2004a; Wagenmakers et al., 2008); basic information processing tasks like choice reaction time (e.g., Laming, 1968; Link and Heath, 1975); memory processes (e.g., Ratcliff et al., 2004b; White et al., 2009); and a range of more complex cognitive decision tasks, including categorization and classification (e.g., Nosofsky and Palmeri, 1997), heuristic decision-making (e.g., Lee and Cummins, 2004; Lee and Zhang, 2012), and judgment and choice (e.g., Wallsten and Barton, 1982; Busemeyer and Rapoport, 1988; Busemeyer and Townsend, 1993; Diederich, 1997).

One area that has been under-explored in diffusion modeling involves the use of time-varying boundaries. The vast majority of diffusion models in psychology use constant boundaries, as shown in **Figure 1A**. Constant boundaries were originally motivated by optimality properties, in the sense that setting a boundary corresponds to setting a Type I error rate, as in the sequential probability ratio test (Wald and Wolfowitz, 1948). Some previous

FIGURE 1 | The basic drift diffusion sequential sampling model of two-choice decision-making and response times (A), and variants involving time-varying boundaries (B–E). In each case, evidence values are sampled from the same Gaussian distribution with mean μ and standard deviation σ, but different boundaries lead to different response time distributions for the two alternative decisions.

diffusion models, however, have considered within-trial changes in boundaries, usually in the form of that converge over time (e.g., Pickett, 1968; Rapoport and Burkheimer, 1971; Clay and Goel, 1973; Viviani, 1979; Hockley and Murdock, 1987; Busemeyer and Rapoport, 1988; Heath, 1992; Frazier and Yu, 2008; Milosavljevic et al., 2010). Considering time-varying boundaries has become an active area of research recently, both in the context of models that combine neuro-psychological data with formal modeling of decision processes (e.g., Cisek et al., 2009; Gluth et al., 2012; Ratcliff and Frank, 2012; Thura et al., 2012), and in the context of studying the theoretical relationships between, and the falsifiability of, sequential-sampling models (Jones and Dzhafarov, 2014).

Figures 1B–E show examples of different time-varying boundaries, and the distributions of decisions and response times they produce for the same Gaussian evidence distribution. It is clear that allowing this flexibility in diffusion models makes them capable of capturing both qualitatively and quantitatively different decision and response time patterns. One reason for wanting this flexibility is to accommodate patterns seen in empirical data, especially arising from experimental task demands. Time-varying boundaries could be regarded, for example, as implementing time pressure, urgency-gating, or deadlines within a single decision trial (Ditterich, 2006; Frazier and Yu, 2008; Cisek et al., 2009). Another reason for considering time-varying boundaries is to broaden the types of optimality in decision-making that can be considered by diffusion models (e.g., Drugowitsch et al., 2012; Ratcliff and Frank, 2012). While constant boundaries, as noted above, optimize single decisions with respect to a fixed Type I

error rate, this is not the only possible criterion decision makers might optimize. For example, in some situations—such as when there is not fixed number of decisions to be made, but rather a fixed length of time in which any number of decisions can be made—it might be more important to optimize the *rate* at which correct decisions are made, rather than focus on the correctness of each individual trial. A specific example is provided by Drugowitsch et al. (2012, Figure 3C), who showed that the optimal boundaries for the Wiener diffusion model are decreasing when there are multiple levels of difficulty and intermixed trials in a 2-alternative-forced-choice (2AFC) task[1]. It is when there is only one level of difficulty in the task that the SPRT Optimality Theorem guarantees that the Wiener process with constant boundaries (among all possible models) maximize any reward criteria that are monotonically non-increasing with respect to the response time (e.g., Bogacz et al., 2006, A.1.1). Many real-world decision-making situations are more general, and so afford possibility that time-varying boundaries may be optimal. In general, different time-varying boundaries can often be interpreted as optimizing different sorts of criteria relevant to different decision-making situations.

In this paper, we develop a computational method for finding time-varying boundaries from response time distributions

---

[1]Simulations in Khodadadi et al. (2014), on the other hand, show that a Wiener process with constant boundaries is optimal for 2AFC with multiple difficulty levels, when a cue is added before each trial indicating the difficulty level of the upcoming trial.

that does not constrain their form and does not commit to specific theoretical assumptions about optimality. Our method is motivated by relevant results from statistics that relate pattern of decisions and response times to diffusion models with time-varying boundaries. Our method does not constrain the time-varying boundaries to a parametric family, but does require knowing the mean and standard deviation of the Gaussian evidence distribution.

To demonstrate our method, we apply it to two concrete problems. The first problem involves equating diffusion models with an alternative class of sequential sampling models, known as accumulator models, and requires applying our method to simulated data. The second problem involves finding the time-varying boundaries in a perceptual decision-making task in the case where the visual stimulus provides the same level of evidence in favor of either decision alternative. Applying our method at the individual level, this second application allows us to consider basic individual differences in the thresholds people use to make a simple perceptual decision. We conclude with a discussion of the theoretical and modeling implications of using time-varying boundaries for diffusion models, as well as considering the limitations and potential of our method.

2. FINDING TIME-VARYING BOUNDARIES

We approach the problem of finding time-varying boundaries as one of solving an inverse problem numerically. There are three important elements to our approach. The first element is having a method for generating the decision and response time distributions that are produced by a known Gaussian evidence distribution and known time-varying boundaries. The second element is a theoretical result that guarantees that any decision and response time distribution, for a given Gaussian evidence distribution, is generated by unique time-varying boundaries. The third element is a numerical method for finding those boundaries, given the Gaussian evidence distribution and decision and response time distribution. In this section, we present each of these three elements in turn.

2.1. GENERATING DATA FROM DIFFUSION MODELS WITH TIME-VARYING BOUNDARIES

We study a diffusion model sampling evidence from the Gaussian distribution with constant mean μ and standard deviation σ, but with the additional flexibility of having time-varying boundaries. This model generates a decision probability p^{diff} and response time distributions r_A^{diff} and r_B^{diff} for the two decisions. Denoting the decision boundaries as a_A and a_B for the two decisions, where a_A and a_B are both time-dependent functions, the diffusion model can be conceived as a mapping

$$m^{\text{diff}} : \left(\mu, \sigma, a_A, a_B\right) \rightarrow \left(p^{\text{diff}}, r_A^{\text{diff}}, r_B^{\text{diff}}\right). \tag{1}$$

The mapping m^{diff} has been studied in the statistics literature, and an effective approach using the analysis of renewal equations has been developed (Durbin, 1971; Buonocore et al., 1987, 1990). Buonocore et al. (1990) provide an efficient algorithm to compute the response time distributions for time-varying boundaries. A summary of these methods well-suited for psychologists

is given by Smith (2000). In particular, data can be generated from a diffusion model with flexible boundaries using general Markov process methods. Because (Smith, 2000) does not provide results for exactly the diffusion model we use (we use a special case of a more general one that is provided), we give explicitly the details needed to reproduce our results.

The basic idea is to specify how sample evidence paths $X(t)$ are generated, and then use existing results that give the first passage time distributions through arbitrary boundaries that are continuously differentiable. The diffusion model we study corresponds to a Wiener process with a constant drift ξ and infinitesimal variance s^2.[2] Specifying the sample paths for this process is done by specifying the transition density

$$f\left(x, t \mid y, \tau\right) = \frac{\mathrm{d}}{\mathrm{d}x} F\left(x, t \mid y, \tau\right) = \frac{1}{\sqrt{2\pi s^2 \left(t - \tau\right)}} \exp$$

$$\left(-\frac{\left(x - y - \xi \left(t - \tau\right)\right)^2}{2s^2 \left(t - \tau\right)}\right) \tag{2}$$

where $F\left(x, t \mid y, \tau\right)$ is the probability of the tally being less than or equal to x at time t, given its value at an earlier time τ was y. Notice that both f and F are the densities when there is no boundary.

The first passage time densities through the time-varying absorbing boundaries, a_A and a_B, are denoted by $g_A(a_A(t), t|x_0, t_0)$ and $g_B(a_B(t), t|x_0, t_0)$, where x_0 and t_0 are the initial state and time. Analysis using the renewal equation (e.g., Durbin, 1971) yields the *Volterra equations* of the relationship between the transition density and the first passage time densities (Smith, 2000, Equation 41):

$$f(a_A(t), t|x_0, t_0) = \int_{t_0}^t g_A(a_A(\tau), \tau|x_0, t_0)f(a_A(t), t|a_A(\tau), \tau)d\tau$$

$$+ \int_{t_0}^t g_B(a_B(\tau), \tau|x_0, t_0)f(a_A(t), t|a_B(\tau), \tau)d\tau$$

$$f(a_B(t), t|x_0, t_0) = \int_{t_0}^t g_B(a_B(\tau), \tau|x_0, t_0)f(a_B(t), t|a_B(\tau), \tau)d\tau$$

$$+ \int_{t_0}^t g_A(a_A(\tau), \tau|x_0, t_0)f(a_B(t), t|a_A(\tau), \tau)d\tau \tag{3}$$

In principle, these equations are soluble, but $f\left(x, t \mid y, \tau\right)$ is singular as t approaches τ, therefore Equation 3 needs to be transformed stably for practical approximation methods. A detailed description of the equation and the singularity issue can be found in Smith (2000, pp. 430–432). The kernels of the transformed equations can be found using the method developed by Buonocore et al. (1987, 1990) and detailed by Smith (2000, pp. 441–446). By letting $\mu(s) = \mu = constant$ in Equation 57 of Smith (2000), the proper function is

---

[2]We use μ and σ to denote the mean and standard deviation of the evidence distribution, or *incremental distribution*, when we discretize the process to take samples from $\mathcal{N}\left(\mu, \sigma\right)$. We use the standard notation ξ and s for the drift and the diffusion coefficient for the corresponding continuous drift diffusion process.

$$\Psi\left(a\left(t\right), t \mid y, \tau\right) = \frac{f\left(a\left(t\right), t \mid y, \tau\right)}{2}\left(a'\left(t\right) - \frac{a\left(t\right) - y}{t - \tau}\right) \quad (4)$$

where $a(t)$ takes the form of a_A or a_B, and $a'(t)$ denotes the first derivative of the boundary. With these results in place, diffusion model data can be produced directly from the first passage time densities, g_A and g_B, which are the same as g_1 and g_2 in Equations 47a and 47b of Smith (2000).

2.2. THEORETICAL RESULTS FOR THE INVERSE PROBLEM

The inverse first passage time problem—finding the boundaries, given the evidence distribution and decision and response time distribution—is much harder than the first passage time problem. It has, however, been studied in the fields of applied mathematics and statistics (e.g., Capocelli and Ricciardi, 1972; Cheng et al., 2006; Chen et al., 2011).

Analytic expressions for the boundaries are rarely available and previous research has usually focused on developing numerical methods for computing the boundary. Theoretical work has been relatively scarce. Early work by Capocelli and Ricciardi (1972) addressed the problem of under what conditions an arbitrary density function can be interpreted as the first passage density function for a continuous one-dimensional Markov process with constant boundaries and a known starting value. Some relevant results, in the context of the types of sequential sampling models used to model human decision-making, were obtained. In particular, Capocelli and Ricciardi (1972, corollary 2.2) found the technical conditions that guarantee the uniqueness of the solution, if it exists, for the Wiener-Lévy and the Ornstein-Uhlenbeck diffusion processes with specified initial condition.

Cheng et al. (2006) were the first to study the well-posedness—that is, the existence and uniqueness—of a specific inverse first-passage time problem close to that of interest in our study. Cheng et al. (2006) addressed the case where a diffusion model has a single boundary, so that there is only one possible decision, and the response time for that decision is being measured. For that case, they proved that for any probability density function q, there exists a unique *viscosity* solution to the inverse-first-passage-time problem (i.e., a unique boundary exists under weak assumptions of differentiability). Analogous results for the two-boundary case of direct interest remain an open (and active) research question in the statistics literature. To date, there is no proof that the numerical method developed in the next section of the paper always finds a unique solution.

2.3. A NUMERICAL METHOD FOR FINDING TIME-VARYING BOUNDARIES

Zucca and Sacerdote (2009) and Song and Zipkin (2011) developed numerical methods for finding time-varying boundaries in the one-boundary case. Because we are interested in diffusion models with two time-varying boundaries, we rely on the approach used by Buonocore et al. (1990). In essence, our method applies this approach, previously used as a forward method only, to the problem of finding two time-varying boundaries.

Algorithm 1 presents the main part of our numerical method for computing the time-varying boundaries as pseudo code. The aim of the algorithm is to find the two boundaries such that the

---

Algorithm 1 | Compute the discretized boundaries $a_A(n)$ and $a_B(n)$, $n = 1, 2, \cdots$, with input μ, σ, $P_{A,n}$, and $P_{B,n}$.

Discretize $[0, 1]$ into I small intervals (grid for the boundary)
for $n = 1$ to N **do**
  Compute $P_{A,n}$ and $P_{B,n}$
  for $i = 1$ to I **do**
    $c_A(i) = i/I$
    $c_B(i) = -i/I$
    $q_A(i) \leftarrow g_A(c_A(i), n\lambda \mid x_0 = 0, t_0 = 0)$
    $q_B(i) \leftarrow g_B(c_B(i), n\lambda \mid x_0 = 0, t_0 = 0)$ g_A, g_B as in
Smith (2000), Equation 47
  end for
  $a_A(n) \leftarrow \arg\min_i \left|q_A(i)\lambda - P_{A,n}\right|/I$
  $a_B(n) \leftarrow -\arg\min_i \left|q_B(i)\lambda - P_{B,n}\right|/I$
end for

---

first passage time densities of the process through those boundaries are equal to two desired specific density functions. The algorithm sets the interval between sampling steps to be a small value λ, and calculates the probabilities $P_{A,n}$ and $P_{B,n}$ that decision alternatives "A" and "B," respectively, will be chosen after n samples. In practice, $P_{A,n}$ and $P_{B,n}$ can be obtained by discretizing the empirical RT distributions for the two alternatives. For the diffusion model discretized to the same sampling interval λ, and using the same Gaussian evidence distribution, the drift rate is $\xi = \mu/\lambda$ and the diffusion coefficient is s, where $s^2 = \sigma^2/\lambda$. The first-order derivative of the boundary at step n can be approximated by $a'(n) = [a(n) - a(n-1)]/\lambda$. These values allow the calculation of Equations 2 and 4 above.

The algorithm finds the time-varying boundary through a point-wise approach to its construction, receiving samples from the same Gaussian evidence distribution with mean μ and standard deviation σ. Because the boundaries scale with σ without changing shape, and our assumption that the decision process starts without bias, the initial values of the boundaries can be fixed at $+1$ and -1, without loss of generality.

The algorithm now sets the equalities $g_A(2)\lambda = P_{A,2}$ and $g_B(2)\lambda = P_{B,2}$, allowing for the solution of the boundaries at the second sample $a_A(2)$ and $a_B(2)$. These steps of the algorithm are now repeated for all of the samples, to find both boundaries in their entirety. Once $a_A(1), \ldots a_A(n)$, and $a_B(1), \ldots a_B(n)$ are available, it is possible to solve for $a_A(n+1)$ and $a_B(n+1)$ by setting the first passage time densities to be equal, so that $g_A(n+1)\lambda = P_{A,n+1}$ and $g_B(n+1)\lambda = P_{B,n+1}$.

Our algorithm solves the equations at each sample using a simple grid search approach. Values between 0 and 1 are examined by a small increment $l = 0.01$ up to N, where N is a large number chosen such that the value of the response time distribution at $N\lambda$ is negligibly small for both decisions.

The recursive nature of the algorithm means that numerical precision errors accumulate as the sample being considered progresses. In practice, we found this sometimes necessitates a second corrective part to our numerical method. For later samples beyond a critical value, we fit the boundary a piece-wise linear curve, each segment containing 2–3 steps, minimizing the

deviation between the simulated and the target first passage time distributions. The boundary that is found is thus a combination of the values returned by the algorithm up to the critical step, and brute-force piece-wise linear curve fitting.

3. APPLICATIONS OF OUR ALGORITHM

In this section, we apply our algorithm to two problems. The first problem is theoretical, and involves the relationship between diffusion classes of sequential-sampling models. The second problem is empirical, and involves finding the time-varying boundaries for individual subjects from their behavioral data in key trials of a simple perceptual decision-making task.

3.1. EQUATING ACCUMULATOR AND DIFFUSION MODELS

Within the sequential sampling framework, an alternative to the class of diffusion model is the class of *accumulator* models (Vickers, 1970, 1979). As shown in **Figure 2**, accumulator models maintain two separate evidence tallies, one for each alternative decision. Each sampled piece of evidence favors one or the other decision, and only those samples that favor a decision are added to their corresponding tally. The first tally to reach the boundary results in that decision being made, and the response time is the number of samples required for this to happen.

Because of their different evidence accrual mechanisms, diffusion and accumulator model are usually regarded as being qualitatively different, and treated as competing accounts of human decision making. Empirically, the standard conclusion is that diffusion models are superior accounts of data (e.g., Ratcliff and Smith, 2004), although there are some studies that find in favor of accumulator models (e.g., Lee and Corlett, 2003). Bogacz et al. (2006) compare diffusion and accumulator models theoretically, in terms of a set of optimality properties, and conclude that accumulator models cannot be reduced to diffusion models.

Complementing this focus on the two models as competing accounts of human decision-making, a natural application of our method is to find the time-varying boundaries that make a diffusion model *equivalent* to an accumulator model with constant boundaries and the same Gaussian evidence distribution. This goal can be seen as a natural extension of the long-standing equivalence result presented by Pike (1968) between random-walk and race models, which are the discrete analogs, respectively, of diffusion and accumulator models. Pike (1968, Section 4.3)

showed that, when the evidence samples are unit increments or decrements, simple time-varying boundaries, decreasing one unit in each time step, make the random-walk decisions and response-time distributions equivalent to the race model.

Formally, we consider the accumulator model sampling evidence from the Gaussian distribution with mean μ and standard deviation σ, and with a fixed starting point 0 and symmetric thresholds. This model generates a decision probability p^{acc} for choosing decision A, and response time distributions r_A^{acc} and r_B^{acc} for the two decisions. Thus, the accumulator model can be conceived as the mapping

$$m^{\text{acc}} : (\mu, \sigma) \rightarrow \left(p^{\text{acc}}, r_A^{\text{acc}}, r_B^{\text{acc}}\right). \tag{5}$$

Equating accumulator and diffusion models requires finding the boundaries $a_A(n)$ and $a_B(n)$, such that $\left(p^{\text{acc}}, r_A^{\text{acc}}, r_B^{\text{acc}}\right) = \left(p^{\text{diff}}, r_A^{\text{diff}}, r_B^{\text{diff}}\right)$.

The mapping m^{acc} has been well-studied. Smith and Vickers (1988) provided an analytical expression, in the form of convolutions of the evidence distribution. For Gaussian evidence distributions, there is no closed-form solution, but a discrete approximation method is provided by Smith and Vickers (1989). In particular, we used the method detailed by Smith and Vickers (1989, Appendix A). Their Equations A3a and A3b define $P_{A,n}$ and $P_{B,n}$ which are, respectively, the probability the accumulator model will choose alternative "A" or "B" after n samples.

Figure 3 shows four examples of the boundaries found by our algorithm. Each example corresponds to a different Gaussian evidence distribution, using means of $\mu = 0.01$ and $\mu = 0.05$ and standard deviations of $\sigma = 0.1$ and $\sigma = 0.12$. For these parameter combinations, we generated response-time distributions from an accumulator model. These distributions provided the input to our algorithm.

The boundaries found by the algorithm are shown in the main left-hand panel for each example in **Figure 3**. The part of the boundary found by the main algorithm is shown as a solid line, while the part found by the piece-wise approximation is shown as a broken line.[3] The basic result is that the decision probabilities

---

[3]The Appendix provides more detail on the piece-wise approximation in this application.
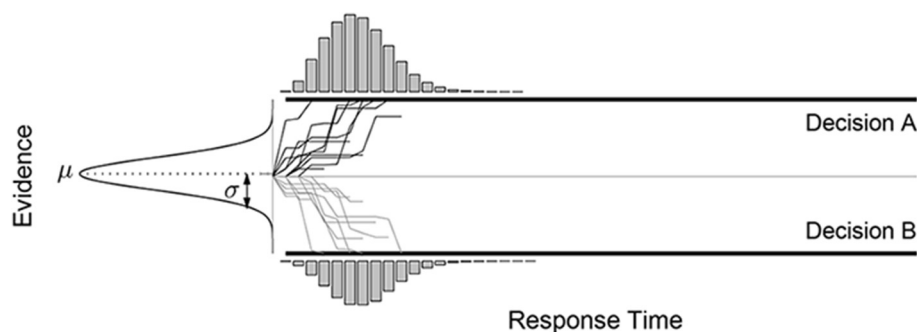
FIGURE 2 | The accumulator sequential sampling model.

FIGURE 3 | Examples of the boundaries that equate a diffusion evidence accrual process with an accumulator model using constant boundaries. The four examples correspond to four different evidence distributions, with values for the means μ and standard deviations σ indicated in the panels. In each example, the left-hand panel shows the time-varying boundaries found by our algorithm, with the part of the boundary found by the main algorithm shown as a solid line, while the part found by the piece-wise approximation shown as a broken line. The right-hand panels show the response time distributions for the two boundaries, weighted by the decision probability for each alternative. The accumulator distributions are shown as solid lines, and the diffusion distributions are shown by gray histograms. The values in the top-right corners show the choice probabilities.

and response-time distributions generated by accumulator models correspond to those generated by a diffusion evidence accrual process with time-varying boundaries.

The right-hand panels in **Figure 3** correspond to the two-decision alternatives, and show the accumulator and diffusion response-time distributions, as solid lines and gray histograms, respectively. These distributions are weighted by the decision probabilities, and so capture all of the aspects of model behavior that need to be equated. It is clear that the decision probabilities and response times generated by the diffusion evidence accrual process with the time-varying boundaries are very close to the target accumulator model distributions.

The four evidence distributions illustrated in **Figure 3** span the interesting range of possibilities. They include cases where the response time distributions are skewed as well as symmetric, and cases where the mean response times for the two decisions are very different as well as very similar. They also include a wide range of decision probabilities, ranging from close to 50% down to about 1%.

The basic result is that diffusion models with time-varying boundaries, of the type shown in **Figure 3**, produce the same decisions and response time distributions as accumulator models with constant boundaries. An important aspect of this result is that the boundaries are established before any particular evidence sequence is encountered. The nature of the boundaries is not developed or changed as evidence is sampled within a trial.

While establishing equivalence dynamically by adapting to current evidence is an interesting research problem in its own right (e.g., Hockley and Murdock, 1987), the current results establish a more general equivalence. They show what sorts of time-varying boundaries make the diffusion approach to evidence accrual the same as standard accumulator approaches.

An interesting aspect of the results in **Figure 3** is that it is clear that the time-varying boundaries are, in general, asymmetric. For example, when the evidence distribution is a Gaussian with $\mu = 0.05$ and $\sigma = 0.10$, the lower boundary converges to zero more quickly than the upper boundary. **Figure 4** presents a follow-up analysis, exploring how important symmetry is to equate accumulator and diffusion approaches to evidence accrual. **Figure 4** shows the response-time distributions for the same examples considered in **Figure 3**, but using a modified algorithm that constrains the boundaries to be symmetric. For the evidence distributions with mean $\mu = 0.01$ there is still close agreement between the accumulator and diffusion response time distributions. For the more extreme examples with mean $\mu = 0.05$, the qualitative properties of different mean response times and negative skew are preserved, but there is quantitative disagreement between the accumulator and diffusion distributions.

3.2. BOUNDARIES FOR AMBIGUOUS PERCEPTUAL STIMULI
One of the most intuitive motivations for considering diffusion models with time-varying boundaries relates to the case
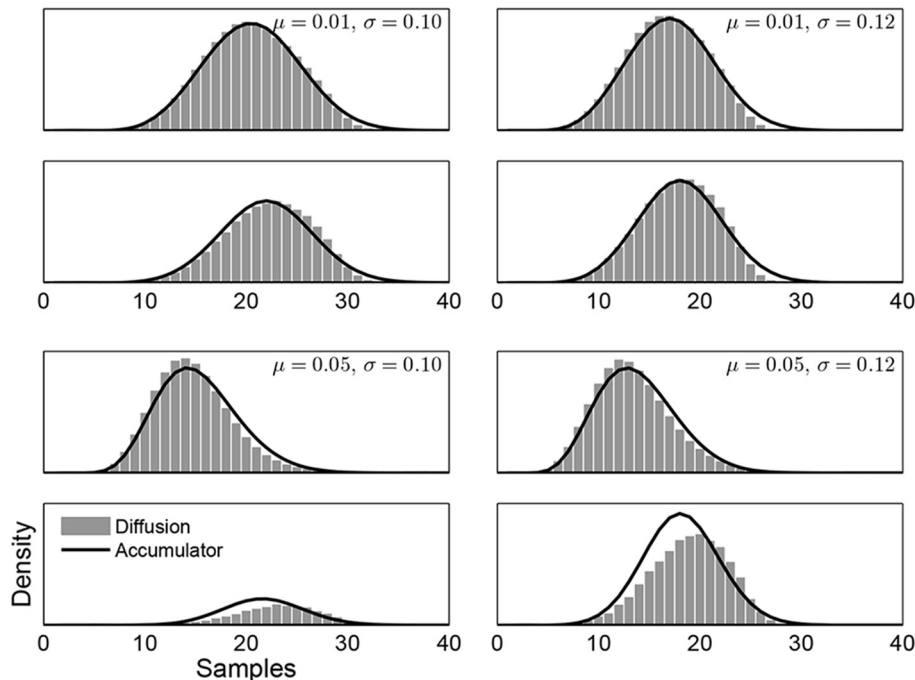
FIGURE 4 | Accumulator and diffusion response time distributions, under the constraint of symmetric boundaries, for four example evidence distributions.

of non-evidential stimuli. These are stimuli that provide equal evidence for both response alternatives, and so the expectation of the evidence distribution is zero (i.e., $\mu = 0$). For these stimuli, constant boundaries predict at least some extremely long response times, even though there is no information to be gained from repeated sampling from the stimulus. This prediction seems problematic, both empirically and theoretically, and has even led to sequential sampling models of human decision-making being lambasted in non-psychological literatures (Lamport, 2012). Converging boundaries provide a natural mechanism for ensuring a decision is made in a reasonable time, without needing to invoke additional psychological assumptions like over-riding termination processes.

Against this background, one interesting application of our method is to find the type of boundaries consistent with behavioral data for non-evidential stimuli. We consider data collected and analyzed by Ratcliff and Rouder (1998), which have also been examined by a number of other authors (e.g., Brown and Heathcote, 2005; Vandekerckhove et al., 2008). The Ratcliff and Rouder (1998) data involve three individual subjects each doing about 8000 trials over 11 days on a brightness discrimination task, under both speed and accuracy instructions. The stimuli consist of visual arrays of black and white dots, with the number of black and white dots controlling the evidence they provide for the choice alternatives bright and dark. Of the 33 different levels of brightness considered by Ratcliff and Rouder (1998), we focus on just those stimuli with equal numbers of black and white dots that (objectively) provide no evidence for either response alternative.

To apply our algorithm to these data, we had to make a number of simplifying assumptions. First, we assumed that the

drift rate was zero, because of the objective properties of the stimuli. Obviously, it is possible that psychologically the stimuli are perceived as favoring one alternative or the other, through some form of bias. Secondly, we shifted the response time distributions according to the smallest response time observed for each individual in each condition. This is a simple empirical approach that probably only roughly approximates the underlying time to encode and respond that requires the shift. Finally, because our method proved unstable with respect to the multi-modalities inherent in binned characterizations of the data, we first fit a Weibull function to the response time distributions, and applied our algorithm to samples from these distributions.

Figure 5 shows the results of our method on the Ratcliff and Rouder (1998) data, as applied to the accuracy condition.[4] We used the Pearson's Chi-square tests standardly used in this literature[5] to evaluate the goodness-of-fit of the Weibull distributions, binning the response times by decile, $d.f. = 7$. For subject "JF," the Chi-square statistics and corresponding p-values for both alternatives are 7.14 ($p = 0.41$) and 13.02 ($p = 0.07$); for subject "KR," they are 4.69 ($p = 0.70$) and 9.01 ($p = 0.25$); for "NH," they are 10.02 ($p = 0.19$) and 10.99 ($p = 0.14$). The three rows in **Figure 5** correspond to the three individual subjects:"JF," "KR," and "NH." The main panels on the left show the boundaries found by our algorithm, with

---

[4]We focused only on the accuracy condition, because we found the Weibull to be an inadequate characterization of the response time distributions in the speed condition.
[5]We are aware of the limitations of both the chi-square statistics and the use of p-values on which this analysis is based.

FIGURE 5 | Results of applying our algorithm to the three subjects from the Ratcliff and Rouder (1998) **data, when viewing non-evidential perceptual stimuli in a brightness discrimination task under accuracy instructions.** The rows correspond to the three individual subjects: "JF," "KR," and "NH." The main panels on the left show the boundaries found by our algorithm for the two decisions. The smaller panels on the right show the distribution of empirical response times (as gray histograms) and the distributions of response times generated by the boundaries (as solid lines) for the two decisions. The values in the top-right corners show the choice probabilities.

respond to discretized samples of 0.01 s duration. Here, we assume that every subject has the same evidence distribution, arbitrarily chosen to be $N(0, 0.01)$, thus the starting values of the boundaries are now free parameters. The smaller panels on the right show the distributions of empirical response times (as gray histograms) and the distributions of response times generated by the time-varying boundaries found by our algorithm (as solid lines) for the two decision alternatives, measured in seconds. There is reasonably good agreement between these distributions, although it is better for some subjects (e.g., "JF") than others. It is also clear that there are significant individual differences between the subjects, with "KR" taking longer to make decisions for these non-evidential stimuli.

Most interestingly, **Figure 5** shows, once again, that the boundaries found are ones that converge asymmetrically. After an extended period of requiring the same level of evidence, both boundaries drop sharply toward zero and converge. They commence their descents at different times, though, with the lower boundary always converging first, but less sharply. Intuitively, when the stimulus favors neither alternative, symmetric boundaries should be able to fit the data well. We calculate the boundaries using the algorithm with the symmetry constraint in Appendix B, and find that the restricted algorithm finds

boundaries close to the boundaries found by the original algorithm.

4. DISCUSSION

Sequential sampling models are compelling accounts of the time course of human decision-making, based on the simple assumption that people sample information from a stimulus until they have enough evidence to make a decision. The default assumption in psychological modeling has been that the level of evidence required to make a decision does not change during this sampling process. The more general idea that the level of evidence might change during sampling is an appealing one, and the possibility that the evidence boundaries triggering decisions converge over time is an important one.

Most previous work dealing with time-varying boundaries has either involved assuming a parametric form for time-varying boundaries and fitting them to data (e.g., Milosavljevic et al., 2010; Ratcliff and Frank, 2012), fitting more general stochastic processes (e.g., Viviani, 1979), or making theoretical assumptions about optimality from which boundaries are derived by methods like dynamic programming (e.g., Frazier and Yu, 2008). In this paper, we have taken the first steps toward a more general approach that places minimal constraints on the form of time-varying boundaries, with the aim of finding their form from the response time distributions they produce.

We developed a method for finding time-varying boundaries that tries to solve the inverse problem of finding the boundaries that generate a given response time distribution for a known Gaussian evidence distribution. This method is related to current theoretical and practical work in statistics (e.g., Capocelli and Ricciardi, 1972; Cheng et al., 2006; Zucca and Sacerdote, 2009; Chen et al., 2011; Song and Zipkin, 2011). There remain important theoretical and practical gaps in these links, however, that future work should address. Theoretically, guarantees for the existence of time-varying boundaries being able to generate any response time distribution are available only for the single-boundary case. Practically, our current approach of solving an inverse problem can and should be generalized to one of solving an inference problem, placed priors on the time-varying boundaries that are possible, and expressing uncertainty over those possibilities based on available data. Our current algorithm, for example, does not allow for any characterization, such as a credible interval, of the uncertainty inherent in the fitted boundaries. Future work should aim to approach the problem as one of inference rather than inversion to provide this important information.

For these reasons, we think the two applications we presented of our method highlight the potential of the general approach, but constitute a starting point rather than a mature method. The theoretical application of our method showed that diffusion processes for accruing evidence, when allowed time-varying boundaries, produce the same behavior as the alternative class of accumulator accrual processes. This result is important, because it encourages a more general modeling perspective than seeing diffusion and accumulator models as incommensurable rivals. It also raises theoretical challenges, such as understanding the difference between what standard diffusion models with

constant boundaries and standard accumulators are optimizing, and understanding the asymmetry of the boundaries that are inferred.

One interpretation of the asymmetry and its behavioral consequences is that accumulator evidence accrual is, in fact, fundamentally different from diffusion evidence accrual, in those situations where the decision-maker must be able to specify decision boundaries before a trial starts. This is because there is no way of knowing *a priori* which decision is favored by the stimulus, and so symmetry of the decision boundaries is a basic requirement. A counter-argument is that **Figure 4** shows that imposing symmetry on the time-varying boundary still leads to close mimicry, and retains agreement on the fundamental qualitative features of the decisions and response times. Thus, it might be argued that there is a practical equivalence, in which empirical data might be equally well-explained by either model. In this sense, our analysis of the asymmetry raised more theoretical questions than it answered, but these questions would not have arisen or be able to be addressed without the capability to examine time-varying boundaries. Thus, we view this application of our method as one of those results that serves to sharpen the theoretical questions, and so usefully advances the field.

Similarly, our analysis of the response time distributions people produce when faced with perceptual stimuli that favored neither alternative is incomplete. We had to make a number of strong simplifying assumptions to apply our algorithm, and we think the boundaries we found should be treated as indicative rather than definitive. But this application did constitute a first productive step toward the important general goal of being able to find time-varying boundaries for diffusion models directly from individual-level behavioral data. The ultimate goal is an approach in which all of the relevant parameters, including properties of the evidence distribution, biases, encoding and responding times, and other properties of the decision-making process can be inferred simultaneously with unconstrained time-varying boundaries needed to account for a large set of empirical data varying across stimuli, task instructions, and other relevant manipulations.

Sequential sampling models are a powerful, popular, and important approach to understanding human decision-making. Extending these models to allow for time-varying boundaries has the potential to enhance greatly what they might help us learn about nature of human decision-making. We hope that the method developed and applied in this paper constitutes a first step toward realizing that potential.

ACKNOWLEDGMENTS

REFERENCES

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced choice tasks. *Psychol. Rev.* 113, 700–765. doi: 10.1037/0033-295X.113.4.700

Brown, S. and Heathcote, A. (2005). A ballistic model of choice response time. *Psychol. Rev.* 112, 117–128. doi: 10.1037/0033-295X.113.4.700

Buonocore, A., Giorno, V., Nobile, A. G., and Ricciardi, L. (1990). On the two-boundary first-crossing-tim problem for diffusion processes. *J. Appl. Probab.* 27, 102–114. doi: 10.2307/3214598

Buonocore, A., Nobile, A., and Ricciardi, L. (1987). A new integral equation for the evaluation. *Adv. Appl. Probab.* 19, 784–800. doi: 10.2307/1427102

Busemeyer, J. R. and Rapoport, A. (1988). Psychological models of deferred decision making. *J. Math. Psychol.* 32, 91–134. doi: 10.1016/0022-2496(88)90042-9

Busemeyer, J. R. and Townsend, J. T. (1992). Fundamental derivations from decision field theory. *Math. Soc. Sci.* 23, 255–282. doi: 10.1016/0165-4896(92)90043-5

Busemeyer, J. R. and Townsend, J. T. (1993). Decision field theory: a dynamic–cognitive approach to decision making. *Psychol. Rev.* 100, 432–459. doi: 10.1037/0033-295X.100.3.432

Capocelli, R. M. and Ricciardi, L. M. (1972). On the inverse of the first passage time probability problem. *J. Appl. Probab.* 9, 270–287. doi: 10.2307/3212798

Chen, X., Cheng, L., Chadam, J., and Saunders, D. (2011). Existence and uniqueness of solutions to the inverse boundary crossing problem for diffusions. *Ann. Appl. Probab.* 21, 1663–1693. doi: 10.1214/10-AAP714

Cheng, L., Chen, X., Chadam, J., and Saunders, D. (2006). Analysis of an inverse first passage problem from risk management. *SIAM J. Math. Anal.* 38, 845–873. doi: 10.1137/050622651

Cisek, P., Puskas, G. A., and El-Murr, S. (2009). Decisions in changing conditions: the urgency-gating model. *J. Neurosci.* 29, 11560–11571. doi: 10.1523/JNEUROSCI.1844-09.2009

Clay, J. R. and Goel, N. S. (1973). Diffusion models for firing of a neuron with varying threshold. *J. Theor. Biol.* 39, 633–644. doi: 10.1016/0022-5193(73)90074-X

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *J. Math. Psychol.* 41, 260–274. doi: 10.1006/jmps.1997.1167

Ditterich, J. (2006). Evidence for time-variant decision making. *Eur. J. Neurosci.* 24, 3628–3641. doi: 10.1111/j.1460-9568.2006.05221.x

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., and Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *J. Neurosci.* 32, 3612–3618. doi: 10.1523/JNEUROSCI.4010-11.2012

Durbin, J. (1971). Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov-Smirnov test. *J. Appl. Probab.* 8, 431–453. doi: 10.2307/3212169

Frazier, P. and Yu, A. J. (2008). "Sequential hypothesis testing under stochastic deadlines," in *Advances in Neural Information Processing Systems 20*, eds J. Platt, D. Koller, Y. Singer, and S. Roweis (Cambridge, MA: MIT Press), 465–472.

Gluth, S., Rieskamp, J., and Büchel, C. (2012). Deciding when to decide: time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. *J. Neurosci.* 32, 10686–10698. doi: 10.1523/JNEUROSCI.0727-12.2012

Heath, R. A. (1992). A general nonstationary diffusion model for two-choice decision-making. *Math. Soc. Sci.* 23, 283–309. doi: 10.1016/0165-4896(92)90044-6

Hockley, W. E. and Murdock, B. B. J. (1987). A decision model for accuracy and response latency in recognition memory. *Psychol. Rev.* 94, 341–358. doi: 10.1037/0033-295X.94.3.341

Jones, M., and Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychol. Rev.* 121, 1–32. doi: 10.1037/a0034190

Khodadadi, A., Fakhari, P., and Busemeyer, J. (2014). Learning to maximize reward rate: a model based on semi-markov decision processes. *Front. Neurosci.* 8:101. doi: 10.3389/fnins.2014.00101

Laming, D. R. J. (1968). *Information Theory of Choice Reaction Time*. London: Academic Press.

Lamport, L. (2012). Buridan's principle. *Found. Phys.* 42, 1056–1066. doi: 10.1007/s10701-012-9647-7

Lee, M. D. and Corlett, E. Y. (2003). Sequential sampling models of human text classification. *Cogn. Sci.* 27, 159–193. doi: 10.1207/s15516709cog2702_2

Lee, M. D. and Cummins, T. D. R. (2004). Evidence accumulation in decision making: unifying the "take the best" and "rational" models. *Psychon. Bull. Rev.* 11, 343–352. doi: 10.3758/BF03196581

Lee, M. D. and Zhang, S. (2012). Evaluating the process coherence of take-the-best in structured environments. *Judge. Decis. Making* 7, 360–372.

Link, S. W. and Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika* 40, 77–105. doi: 10.1007/BF02291481

Luce, R. D. (1986). *Response Times*. New York, NY: Oxford University Press.

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., and Rangel, A. (2010). The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judge. Decis. Making* 5, 437–449. doi: 10.2139/ssrn.1901533

Nosofsky, R. M. and Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychol. Rev.* 104, 266–300. doi: 10.1037/0033-295X.104.2.266

Pickett, R. M. (1968). "The visual perception of random line segment texture," in *Paper read at Ninth Meeting of the Psychonomic Society* (St. Louis, MO).

Pike, A. (1968). Latency and relative frequency of response in psychophysical discrimination. *Br. J. Math. Stat. Psychol.* 21, 161–182. doi: 10.1111/j.2044-8317.1968.tb00407.x

Rapoport, A. and Burkheimer, G. J. (1971). Models for deferred decision making. *J. Math. Psychol.* 8, 508–538. doi: 10.1016/0022-2496(71)90005-8

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108. doi: 10.1037/0033-295X.85.2.59

Ratcliff, R. (1980). A note on modelling accumulation of information when the rate of accumulation changes over time. *J. Math. Psychol.* 21, 178–184. doi: 10.1016/0022-2496(80)90006-1

Ratcliff, R. (1985). Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychol. Rev.* 92, 212–225. doi: 10.1037/0033-295X.92.2.212

Ratcliff, R. (1988). Continuous versus discrete information processing: modeling the accumulation of partial information. *Psychol. Rev.* 95, 238–255. doi: 10.1037/0033-295X.95.2.238

Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychol. Rev.* 120, 281–292. doi: 10.1037/a0030775

Ratcliff, R. and Frank, M. J. (2012). Reinforcement-based decision making in corticostriatal circuits: mutual constraints by neurocomputational and diffusion models. *Neural Comput.* 24, 1186–1229. doi: 10.1162/NECO_a_00270

Ratcliff, R. and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* 20, 873–922. doi: 10.1162/neco.2008.12-06-420

Ratcliff, R. and Rouder, J. N. (1998). Modeling response times for two–choice decisions. *Psychol. Sci.* 9, 347–356. doi: 10.1111/1467-9280.00067

Ratcliff, R. and Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 127–140. doi: 10.1037/0096-1523.26.1.127

Ratcliff, R. and Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychol. Rev.* 111, 333–367. doi: 10.1037/0033-295X.111.2.333

Ratcliff, R., Thapar, A., Gomez, P., and McKoon, G. (2004a). A diffusion model analysis of the effects of aging in the lexical–decisiontask. *Psychol. Aging* 19, 278–289. doi: 10.1037/0882-7974.19.2.278

Ratcliff, R., Thapar, A., and McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Percept. Psychophys.* 65, 523–535. doi: 10.3758/BF03194580

Ratcliff, R., Thapar, A., and McKoon, G. (2004b). A diffusion model analysis of the effects of aging on recognition memory. *J. Mem. Lang.* 50, 408–424. doi: 10.1016/j.jml.2003.11.002

Ratcliff, R., Van Zandt, T., and McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychol. Rev.* 106, 261–300. doi: 10.1037/0033-295X.106.2.261

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: a foundational primer. *J. Math. Psychol.* 44, 408–463. doi: 10.1006/jmps.1999.1260

Smith, P. L. and Vickers, D. (1988). The accumulator model of two-choice discrimination. *J. Math. Psychol.* 32, 135–168. doi: 10.1016/0022-2496(88)90043-0

Smith, P. L. and Vickers, D. (1989). Modeling evidence accumulation with partial loss in expanded judgment. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 797–815. doi: 10.1037/0096-1523.15.4.797

Song, J.-S. and Zipkin, P. (2011). An approximation for the inverse first passage time problem. *Adv. Appl. Probab.* 43, 264–275. doi: 10.1239/aap/1300198522

Thura, D., Beauregard-Racine, J., Fradet, C. W., and Cisek, P. (2012). Decision making by urgency gating: theory and experimental support. *J. Neurophysiol.* 108, 2912–2930. doi: 10.1152/jn.01071.2011

Usher, M. and McClelland, J. L. (2001). On the time course of perceptual choice: the leaky competing accumulator model. *Psychol. Rev.* 108, 550–592. doi: 10.1037/0033-295X.108.3.550

Vandekerckhove, J., Tuerlinckx, F., and Lee, M. D. (2008). "A Bayesian approach to diffusion process models of decision-making," in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, eds V. Sloutsky, B. Love, and K. McRae (Austin, TX: Cognitive Science Society), 1429–1434.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics* 13, 37–58. doi: 10.1080/00140137008931117

Vickers, D. (1979). *Decision Processes in Visual Perception*. New York, NY: Academic Press.

Viviani, P. (1979). A diffusion model for discrimination of temporal numerosity. *J. Math. Psychol.* 19, 108–136. doi: 10.1016/0022-2496(79)90015-4

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., and McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *J. Mem. Lang.* 58, 140–159. doi: 10.1016/j.jml.2007.04.006

Wald, A. and Wolfowitz, J. (1948). Optimal character of the sequential probability ratio test. *Ann. Math. Stat.* 19, 326–339. doi: 10.1214/aoms/1177730197

Wallsten, T. S. and Barton, C. (1982). Processing probabilistic multidimensional information for decisions. *J. Exp. Psychol. Learn. Mem. Cogn.* 8, 361–384. doi: 10.1037/0278-7393.8.5.361

White, C., Ratcliff, R., Vasey, M., and McKoon, G. (2009). Dysphoria and memory for emotional material: a diffusion model analysis. *Cogn. Emot.* 23, 181–205. doi: 10.1080/02699930801976770

Zucca, C. and Sacerdote, L. (2009). On the inverse first-passage-time problem for a wiener process. *Ann. Appl. Probab.* 19, 1319–1346. doi: 10.1214/08-AAP571

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

APPENDIX A

Figure A1 shows the numerical problem in the main algorithm that requires the addition of the piece-wise linear correction. It shows the results of applying the unmodified **Algorithm 1** to the response time distribution generated by an accumulator model with with Gaussian evidence distribution parameters $\mu = 0.01$ and $\sigma = 0.1$ considered in the top-left of **Figure 3**. The left hand panel of **Figure A1** shows the boundaries found, which differ from those in **Figure 3** after the 26th sample, as indicated by the broken lines. The right hand panel of **Figure A1** shows the target response time distributions generated by the accumulator model as solid lines, and the distributions generated from the boundary found by the unmodified algorithm as a line with asterisk markers. Using a small tolerance for the difference between these expected and generated distributions, it is possible to identify the critical point, highlighted by the magnification in the right hand panel, beyond which the piece-wise linear correction is applied.

APPENDIX B

Figure A2 shows the results of applying a modified version of our algorithm that is constrained to find symmetric boundaries to the data from non-evidential stimuli for the three subjects considered by Ratcliff and Rouder (1998).
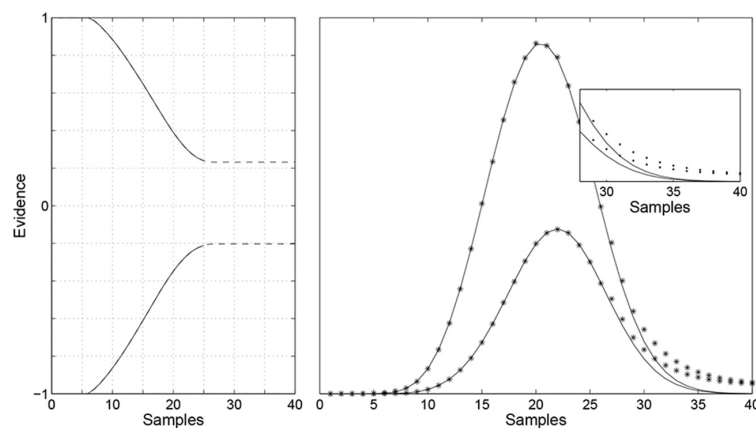
FIGURE A1 | Illustration of numerical problems with the basic algorithm The left-hand panel shows the boundary returned by the algorithm without correction. The right-hand panel shows with asterisks the response time distributions generated by the boundaries in the left-hand panel. The target densities generated by the accumulator model are shown in solid lines.
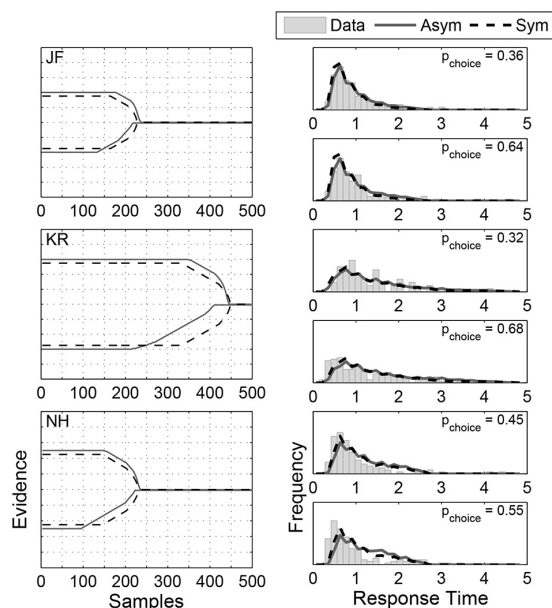
FIGURE A2 | Symmetric boundaries (shown as the broken lines) found for the three subjects using the restricted algorithm, superimposed with the asymmetric boundaries as in Figure 5.

Category variability effect in category learning with auditory stimuli

Lee-Xieng Yang[1]* and Yueh-Hsun Wu[2]

[1] Department of Psychology and Research Center for Mind, Brain, and Learning, National Chengchi University, Taipei, Taiwan
[2] Department of Psychology, National Chengchi University, Taipei, Taiwan

The category variability effect refers to that people tend to classify the midpoint item between two categories as the category more variable. This effect is regarded as evidence against the exemplar model, such as GCM (Generalized Context Model) and favoring the rule model, such as GRT (i.e., the decision bound model). Although this effect has been found in conceptual category learning, it is not often observed in perceptual category learning. To figure out why the category variability effect is seldom reported in the past studies, we propose two hypotheses. First, due to sequence effect, the midpoint item would be classified as different categories, when following different items. When we combine these inconsistent responses for the midpoint item, no category variability effect occurs. Second, instead of the combination of sequence effect in different categorization conditions, the combination of different categorization strategies conceals the category variability effect. One experiment is conducted with single tones of different frequencies as stimuli. The collected data reveal sequence effect. However, the modeling results with the MAC model and the decision bound model support that the existence of individual differences is the reason for why no category variability effect occurs. Three groups are identified by their categorization strategy. Group 1 is rule user, placing the category boundary close to the low-variability category, hence inducing category variability effect. Group 2 takes the MAC strategy and classifies the midpoint item as different categories, depending on its preceding item. Group 3 classifies the midpoint item as the low-variability category, which is consistent with the prediction of the decision bound model as well as GCM. Nonetheless, our conclusion is that category variability effect can be found in perceptual category learning, but might be concealed by the averaged data.

Keywords: category variability effect, sequence effect, perceptual category learning, memory and comparison, decision bound model

The seminal study of Rips (1989) showed that people tend to classify an item (e.g., a 3-inches circular object) at the midpoint between two categories (e.g., QUATER and PIZZA) as the category with a larger variability (i.e., PIZZA), although the middle item is more similar to the low-variability category (i.e., QUATER). This finding attracts many researchers' attention, for it indicates that category variability is one of the sources for categorization and challenges the exemplar-based model, specifically GCM (Generalized Context Model; Nosofsky, 1986, 1987). Since the exemplars of low-variability category vary in a smaller range than those of high-variability category, the total distance from exemplars to the middle item is shorter for the low-variability category than the high-variability category. Thus, the middle item is more similar to the low-variability category. Based on similarity, GCM would always classify the middle item as the low-variability category. Only when the two categories in the same psychological space have different specificities for similarity computation, can GCM predict Rips (1989)' finding (see Nosofsky and Johansen, 2000).

In contrast, the famous rule-based model GRT (Generalized Recognition Theory; Ashby and Townsend, 1986; Ashby and Gott, 1988; Ashby and Maddox, 1992; Maddox and Ashby, 1993) is thought to be able to account for this phenomenon. According to GRT, learning categories is to generate a category boundary. The boundary divides the psychological space into different regions, each of which corresponds to a category. An item would be classified as a category, if its percept is located in the region corresponding to that category. Each category is assumed to be represented as a normal distribution with the mean location having the largest likelihood to be classified as that category. The optimal boundary between two categories is located on where the percept of item has an equally high likelihood to be classified as either category. According to the nature of normal distribution, the likelihood of a value is a function of the distribution variance. Thus, the optimal category boundary will be influenced by the variance of category distribution and always close to the low-variability category. This is why the middle item would be predicted as the high-variability category by GRT.

Although this phenomenon is observed in conceptual category learning, it is not often reported in the studies of perceptual category learning. Thus, the purpose of this study is to examine whether the variability of category would influence perceptual

categorization. Specifically, how the midpoint item between two categories would be classified is our focus. For the convenience of discussion, we follow Stewart and Chater (2002) to call this phenomenon category variability effect (CVE). In the later sections, we review the past studies, discussing the possible reasons for the low reliability of them, including the sequence effect in category learning and individual differences, and then introduce our experiment, discussing the empirical data, and modeling results.

1. CATEGORY VARIABILITY EFFECT IN PERCEPTUAL CATEGORY LEARNING

In the study of Cohen et al. (2001), two categories were defined as high-variability and low-variability categories by their covering range on the stimulus dimension. In the learning phase, the participants learned to correctly classify the exemplars of these two categories. In the transfer phase, the critical item was presented for the participants to predict its category label. The results showed that the probability of high-variability category for the critical item became higher when the exemplar number of high-variability category increased from two to seven, with the exemplar number of low-variability category fixed to one. However, the probability of high-variability category for the critical item is still not significantly larger than 0.50, namely no CVE occurred.

Stewart and Chater (2002) used a circle with a dot attaching on its periphery as stimulus. The dot position was the stimulus dimension and the high-variability and low-variability categories, respectively, cover a larger and a smaller portion of the periphery. Their results showed no CVE when the participants were presented with one stimulus on each trial. However, when all exemplars of each category were presented together to the participants in the learning phase, CVE was observed. Thus, it seems critical to CVE that people should be aware of the variability of category.

Similar to Cohen et al. (2001), Hsu and Griffiths (2010) also used lines of different lengths as stimuli to examine CVE. In the discrimination condition, the participants were instructed to predict the category label, given the current line length. In the generation condition, the participants were instructed to predict which category would be more likely to have a line of this length. The results showed no CVE in the discrimination condition and a clear CVE in the generation condition. These behavioral results were correctly simulated by their Bayes network models. For the generative condition, the model aimed to estimate the probability distribution over the input given the category label, namely $p(x|c)$. However, for the discrimination condition, the model aimed to find a direct mapping between inputs and category labels, namely $p(c|x)$. The success of their models implied that the occurrence of CVE demands the knowledge about candidate categories. Together with the findings of Stewart and Chater (2002), this knowledge should include the variability of each category.

According to the previous review, it is not clear whether CVE would occur in category learning. To figure out why the past studies did not observe CVE is the purpose of this study. We seek for the answer by checking out the nature of category learning task, instead of testing people in some new experimental design.

Our focus is on the sequence effect and individual differences in category learning.

2. SEQUENCE EFFECT IN CATEGORY LEARNING

Normally, the category representation (i.e., rule or exemplars) is assumed to be quite stable during category learning, as it is the representation of category structure, which would not change throughout the experiment. Thus, with the stable category representation, one item would be classified to the same category under any circumstances. However, recent studies show that the same item might be classified as different categories when following different items (Stewart et al., 2002; Stewart and Brown, 2004). This finding instead suggests the possibility of short-term representation (i.e., the information of the preceding item) to be adopted in category learning. Inspired by this finding, in the case of CVE, the midpoint item may be classified as one category when following a certain items and the other category when following some other items. Accordingly, when mixing up these conditions, the averaged result would show no CVE. If this is true, we should expect some sequence effect in the experiment for examining CVE.

The sequence effect of our interest is suggested by Stewart et al. (2002)'s MAC (Memory and Comparison) strategy for categorization. The MAC strategy is very simple. Suppose we know that one category takes larger values and the other takes smaller values, just like the one-dimensional category structure used for examining CVE. When item $n - 1$ is from the large category and item n is even larger than it, $X_n \geq X_{n-1}$, item n must be the large category. Likewise, when item $n - 1$ item is from the small category and $X_n \leq X_{n-1}$, item n must also be the small category. That is, when the sign of the difference between successive items can guarantee the category of the latter one, the probability to repeat the preceding category label as the response for the latter item is 1.00. When this heuristic cannot be applies to categorization, that is $X_n < X_{n-1}$ when item $n - 1$ is from the large category or $X_n > X_{n-1}$ when item $n - 1$ is from the small category, the probability to repeat the preceding category as the current response is the similarity between item $n - 1$ and item n. The MAC model can be expressed as

$$p = \begin{cases} 1.00 \\ exp^{-c|X_n - X_{n-1}|}, \end{cases} \quad (1)$$

where c is the specificity, when c is large, items would be less similar and vice versa. The similarity between item $n - 1$ and item n is exponentially transferred from their psychological distance. The smaller the distance, the larger the similarity.

According to the MAC strategy, we define the sequence effect as the tendency to repeat the preceding category label as current response. In this study, we would like to examine the sequence effect in categorization. Specifically, we would like to check if this effect is the reason for the inconsistent reports about CVE in the past studies.

3. INDIVIDUAL DIFFERENCES IN CATEGORY LEARNING

In addition to sequence effect, whether there are individual differences concealed in the averaged data is our second concern. In the literature of category learning, heaps of individual

differences are reported, providing us clues to understand individual participant's categorization strategy (Nosofsky et al., 1989; Johansen and Palmeri, 2002) and to evaluate models (Maddox and Ashby, 1993; Nosofsky et al., 1994). For instance, Yang and Lewandowsky (2004) examined human's category learning with multi-dimensional stimuli. Among all dimensions, one was the context dimension. In their experimental design, no matter the context dimension was attended to or not, participants could get perfect learning performance. The results showed a clear difference on categorization strategy. One group of participants learned to attended to the context dimension for categorization, whereas the other group did not. The modeling results further showed that ATRIUM (Erickson and Kruschke, 1998) (with rule plus exemplar) can account for the performance of both groups, whereas ALCOVE (Kruschke, 1992) (with exemplar only) had difficulty doing so. Thus, these authors suggested that multiple representations are used in categorization.

Perhaps, the most salient contribution of individual-difference analysis is to turn over our understanding of an old phenomenon. For instance, in order to examine the allocation of attention over stimulus dimensions during category learning, Lee and Wetzels (2010) reanalyzed the data of Kruschke (1993) study. In the condensation condition of this study, the category structure could be perfectly learned, if the information from two stimulus dimensions were integrated for categorization. Lee and Wetzels (2010) first fit GCM to the averaged data. The estimated attention weight on one dimension was about 0.55, suggesting that the participants did spread their attention equally on the two dimensions. However, when fitting GCM to the individual data, clear individual differences were observed. One group of participants focused their attention on one dimension, whereas the other group strongly attended to the other dimension. The averaged data disguised this fact and erroneously suggested that people evenly divided attention on the two dimensions when learning the condensed category structure. Therefore, the individual differences provide a more transparent understanding about how attention can be allocated during category learning.

Back to the issue of CVE. The past studies all reported the averaged data. As shown by Lee and Wetzels (2010)'s work, the averaged data might be not too much informative. Thus, it is reasonable to suspect that the non-CVE result reported in the past studies might actually contain the positive evidence of CVE as well. Thus, in this study, we would also like to examine the occurrence of CVE via the analysis of individual differences.

4. EXPERIMENTS

According to previous discussions, we proposed two hypotheses to address the question why CVE was not consistently observed in category learning. First, there might be some individual differences buried under the averaged data. Perhaps those non-CVE reports actually included some participants who did show CVE and some others did not. Second, the classification for the midpoint item might be influenced by the preceding item, namely the sequence effect. As a result, the midpoint item may be classified as the high-variability category following some precedent and not following some others. In order to get rid of confounding from the regimen of experiment, we conducted this experiment

in the conventional feedback-learning paradigm. All participants were asked to do the learning phase and then the transfer phase. The emphasis of data analysis was placed on verifying these two hypotheses.

In addition, we used single tones varying in frequency as stimuli in this experiment. In order to make the scale of stimuli equal in distance from one another, we transferred the frequency f to the psychological scale mel, $mel = 1127 log_e(\frac{f}{700} + 1)$ (Steinberg, 1937; Stevens et al., 1937). The category structure was shown in **Figure 1**. There were five items in each category. The low-variability category (called Category 1) took the region between 480 and 520 mel and the high-variability category (called Category 2) took the region between 670 and 970 mel. The interval between the members of Category 1 was 10 mel and that of Category 2 was 75 mel. The critical item was the tone of 595 mel, which was denoted as the white bar in **Figure 1**. Therefore, if the probability of Category 1 for the critical item was less than 0.50, CVE occurred. All tones were played at a constant amplitude of 60 dB.

4.1. METHODS

4.1.1. Participants and apparatus

In total, 41 undergraduate students from National Chengchi University aged from 18 to 30 were recruited in this experiment. The whole experiment was conducted in a quiet dim booth. The display of stimuli, the procedure of testing, and the data collection were all controlled by the scripts of MATLAB on an IBM compatible PC. On average, every participant would finish this experiment in 30 min and got reimbursed with NTD\$ 60 ($\simeq$ US\$ 2) for their time and travel expense. Before doing the experiment, all participants were confirmed to be able to hear two extreme tones (i.e., 470 and 980 mel, covering the range of stimulus tones) each presented twice in a headset.

FIGURE 1 | The category structures. HV, high variability; LV, low variability. Axis X represents tones and axis Y shows the mel of the stimuli.
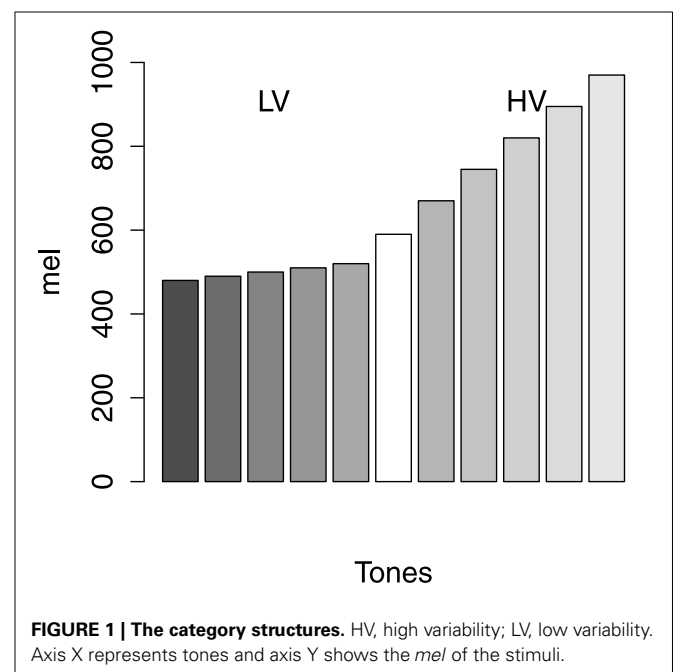
4.1.2. Materials and procedure

Following the design of past studies (e.g., Sakamoto et al., 2008; Hsu and Griffiths, 2010), the two categories were defined as two uniform distributions. In the low-variability category, there were 5 tones equally spreading from 480 *mel* to 520 *meal* with 10 *mel* as the interval. In the high-variability category, there were also 5 tones equally spreading from 670 to 970 *mel* with 75 *mel* as the interval. There were 5 learning blocks, each of which was followed by a transfer block. Therefore, there were in total 10 blocks in this experiment. In the learning block, the 10 tones of the two categories were presented twice in random order. In the transfer block, the transfer stimuli consisted of 2 tones randomly sampled from each category and 1 critical item, which was 595 *mel* at the mid point between the edges of two categories. The transfer stimuli were presented once, except that the critical item was presented twice. Thus, there were 6 trials in total in the transfer block, which of course were presented in random order.

On each learning trial, a tone was presented to the participants from a headset for 1 s. After the stimulus disappeared, the participants were asked to predict which alien (i.e., Category 1 or Category 2) would make this sound by pressing the "s" key or the ";" key. Once the response was made, a "correct" or "wrong" feedback signal was presented on the computer screen for 500 ms. After 2 s, next trial began. The participants were instructed to do this task as accurately as they can. On each transfer trial, the procedure was the same as on the learning trial, except that there was no corrective feedback.

4.2. RESULTS

4.2.1. Learning phase

The participants learn the categories quite well. The accuracy in the first block is as high as 0.86 and it increases significantly to 0.95 in the fifth block, $F_{(4, 160)} = 13.08$, MSe = 0.003, $p < 0.01$. Clearly, this task is very easy to the participants.

4.2.2. Transfer phase

The mean probability of Category 1 on transfer item across five transfer blocks is shown in **Figure 2**. Axis X denotes the item *mel* and axis Y the probability of Category 1 predicted by the participants. For the items which have been presented in the learning phase, a Category (2) × Block (5) within-subject ANOVA shows that they are correctly classified as their own categories [$F_{(1, 40)} = 3786$, MSe = 0.02, $p < 0.01$]. However, the overall tendency to make a Category 1 response is influenced by the transfer block [$F_{(4, 160)} = 6.564$, MSe = 0.018, $p < 0.01$]. This is because there is a drop on the mean probability of Category 1 (0.46) in the final block. With no doubt, the response for the item from each category is not changed in different blocks [$F_{(4, 160)} = 1.364$, MSe = 0.016, $p = 0.25$]. Thus, the participants' categorization for the learning items is accurate and consistent through the transfer blocks.

Of most interest is how the participants would predict the category of the critical item. The mean probability of Category 1 on the critical item across five blocks is 0.54, which is not significantly different from 0.50 [$t_{(40)} = 0.96$, $p = 0.34$]. Thus, there is no evidence of CVE, as the critical item is not significantly classified as Category 2 (the high-variability category). However, the
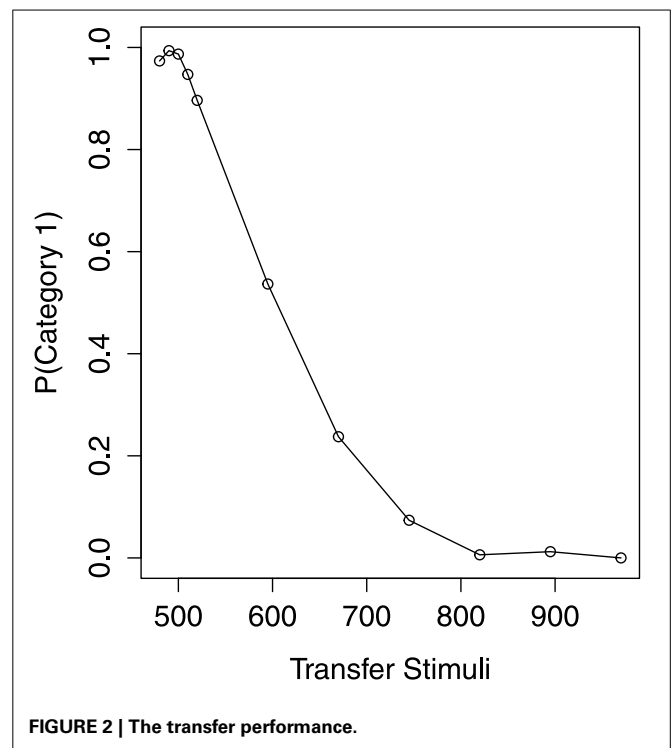
FIGURE 2 | The transfer performance.

critical item is decreasingly classified as Category 1 in a linear trend from the first block [$p_{(Category1)} = 0.68$] to the fifth block [$p_{(Category1)} = 0.41$], with $F_{(1, 40)} = 15.63$, MSe = 0.15, $p < 0.01$. In the final block, the probability of Category 1 for the critical item is still not different from 0.50 [$t_{(40)} = -1.42$, $p = 0.16$].

4.2.3. Sequence effect

As discussed in the previous section, how to classify an item might depend on which item it follows. In this experiment, the critical item is presented twice in every transfer block, once following a different item. Thus, it is reasonable to suspect that the critical item actually be classified as the high-variability category in one time, but as the low-variability category in another, so the aggregated result shows no CVE.

In order to verify this hypothesis, we examine for any sequence effect in our transfer data. Following the idea of the MAC strategy, we redefine the trials to four cases (C1+Up, C1+Down, C2+Up, and C2+Down)[1], according to the category of the preceding item (Category 1 or Category 2) and the change of direction on frequency from the preceding item to the current one (Up or Down). One point is worth noting. In the transfer phase, there is no feedback, hence no correct answer for the preceding item. We substitute the participants' response for the category answer, due to the high learning accuracy they made in the experiment (mean = 0.94). If the participants rely on some long-term representations to do categorization (i.e., rule or exemplars of the two categories), the preceding category has nothing to do with the current response and so is the direction change between the frequencies of successive tones.

---

[1] Thus, the first trial is omitted.

FIGURE 3 | Sequence effect in learning phase, transfer phase, and on the critical item.
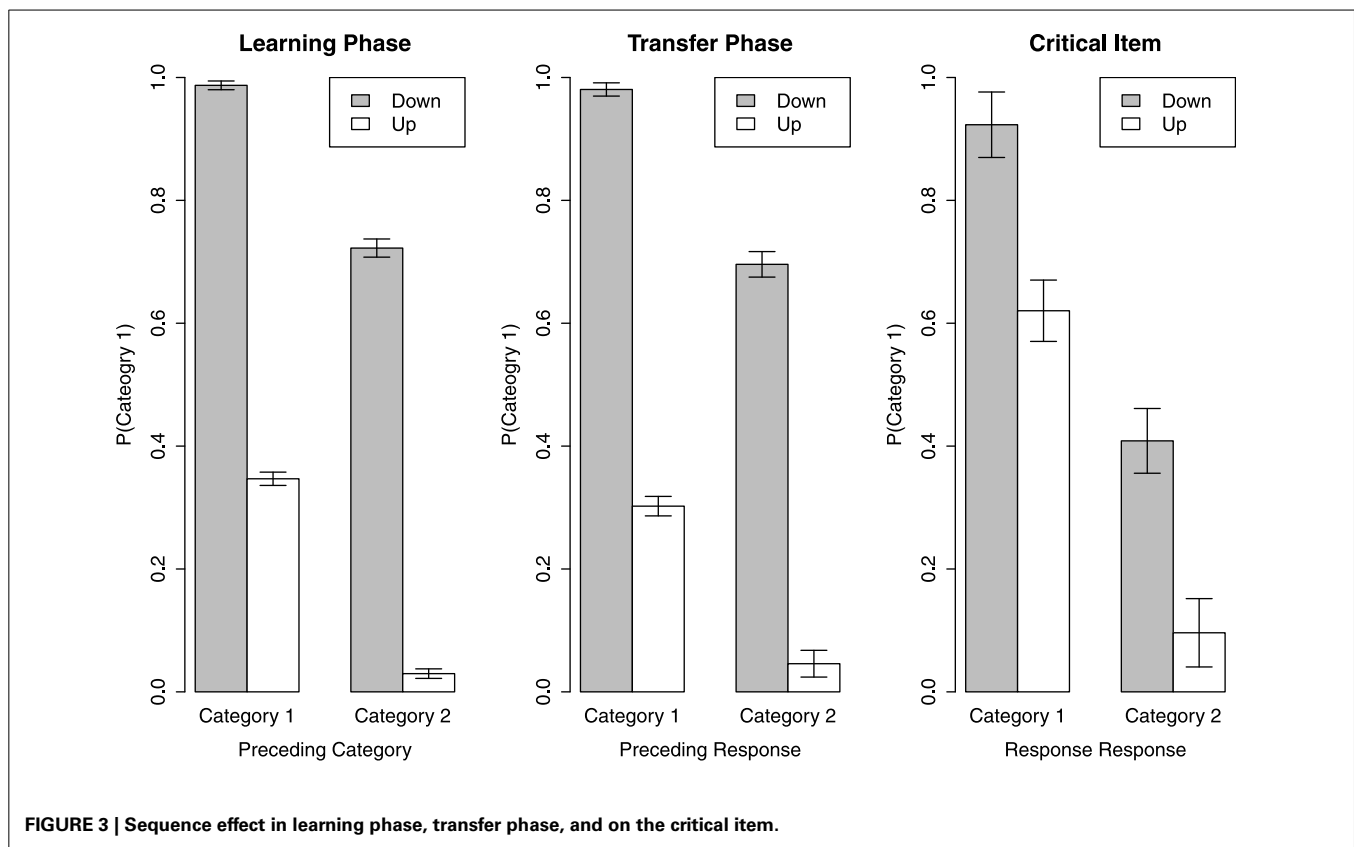
The results are revealed in **Figure 3**. See the left and middle panels for the overall results in the learning and transfer phases. Visual inspection shows when the direction of frequency change provides sufficient information, namely the cases of C1+Down and C2+Up, the participants strongly repeat the preceding category as the current response. For the cases of C1+Up and C2+Down, this tendency is not as strong. Across the learning and transfer phases, when the tone sounds higher than the preceding one, the participants tend to make a Category 2 response and when the tone sounds lower, they tend to make a Category 1 response $[F_{(1, 40)} = 2695, MSe = 0.005, p < 0.01]$. However, regardless of the direction of frequency change, the participants seem to repeat the preceding category as the current response to a certain extent that the main effect of the preceding category is significant $[F_{(1, 40)} = 513.60, MSe = 0.01, p < 0.01]$. The overall mean probability of Category 1 made for the current item is not different between the learning phase and the transfer phase $[F_{(1, 40)} = 3.29, MSe = 0.006, p = 0.07]$. The response pattern of the cases when the preceding item is from different categories is not different in different phases $[F_{(1, 40)} = 1.03, MSe = 0.01, p = 0.32]$. Also, the response pattern of the cases when the frequency change in different directions is not different in different phases $[F_{(1, 40)} < 1]$. There is no significant interaction effect between the preceding category and the direction of frequency change across all phases $[F_{(1, 40)} < 1]$. However, the three-way interaction effects between the experiment phase, the preceding category, and the frequency change direction is significant $[F_{(1, 40)} = 5.55, MSe = 0.006, p < 0.05]$.

We also examine sequence effect on the critical item. See the right panel in **Figure 3**. Recall the critical item is actually higher in frequency than the members of Category 1 and lower than the members of Category 2. Thus, the cases of C1+Down and C2+Up theoretically do not exist. The two bars for these two cases represent the response made for the current critical item when the preceding item was also the critical item. However, for some participants who have never seen the critical item being presented twice in turn, we substitute the mean of the rest participants' data for the missing value. A Category (2) × Direction (2) within-subject ANOVA shows that the prediction for the critical item is influenced by the preceding response $[F_{(1, 40)} = 139.2, MSe = 0.08, p < 0.1]$ and the change in direction of frequency from the preceding item $[F_{(1, 40)} = 46.57, MSe = 0.08, p < 0.01]$. However, there is no interaction effect between Category and Direction $[F_{(1, 40)} < 1]$.

Although these results seem to be the evidence of sequence effect, the two cases C1+Down and C2+Up are actually not that informative. As the current item in these two cases can also be correctly categorized by a rule or by all exemplars of categories. Thus, the cases of C1+Up and C2+Down are our focus. It is clear that the participants tend to predict the current item as the category which is contrasting to the preceding one in the left and middle panels in **Figure 3**. However, this pattern is not held in the right panel for the critical item. In fact, the participants seem to predict the critical item as the same category of the preceding item, although the tendency is not strong. This is not surprising. When the critical item is in the C1+Up or C2+Down cases,

similarity to the preceding item is the sole basis to predict its category. As the critical item is at the center position of all stimuli, the similarity between it and any other item is on average higher than that between any other pairs. Thus, in the C1+Up and C2+Down cases, the critical item would be more likely classified as the category of its preceding item. The analysis of sequence effect seems to suggest that the negative finding of CVE results from mixing all different influences brought by the preceding items in different testing situations. However, this conclusion is better not made quickly until we check out the individual differences.

4.3. INDIVIDUAL DIFFERENCES ANALYSIS

The sequence effect on the critical item may provide an explanation to why there is no CVE observed in the averaged data. However, we do not know whether this is a general case for all participants or there are some rule-use strategies[2] mixed up in the averaged data. In fact, it is hard to detect those rule users by simply looking at the averaged sequence effect data. This is because their predictions for the critical item would be independent of the preceding item, that makes their influence as a constant added to the four categorization conditions. Therefore, we intend to investigate the individual differences by fitting the MAC model and the decision bound model[3] to each participant's data. If the MAC model provides a better fit, the participant is regarded as a MAC strategy user. If the decision bound model provides a better fit, the participant is regarded as a rule user. Presumably, the participants who show CVE must be in the group of rule user. We can check out the probability of high-variability category predicted for the critical item to identify them. If there exist rule users, especially those who show CVE, the sequence effect should not be regarded as the reason for not observing CVE.

For each participant, we fit these two models to the transfer data separately. For the MAC model, only the specificity c is freely estimated. If the preceding item is from Category 2, the output will be transferred to $p(1) = 1 - p(2)$ to make sure all MAC predictions are the probability of Category 1. For the decision bound model, the probability of Category 2 for item X is transferred from the area below the percept of X on the normal distribution with category boundary b as mean and perceptual error ϵ as standard deviation. The larger the covered area, the larger probability of Category 2 is[4]. The parameters b and ϵ are freely estimated. The stimulus values are normalized between 0 and 1 for modeling. The aim of parameter estimation is to maximize the likelihood of the model to predict the observed probability of Category 1 in the four categorization conditions (i.e., C1+Up, C1+Down, C2+Up, and C2+Down). The goodness of fit is $AIC = -2LogL + 2N$ (Aakike's Information Criterion; Akaike, 1974) with $N =$ parameter number. The smaller AIC the better fit. The log likelihood is

---

[2]Since the exemplar model such as GCM is evident to have difficulties accounting for CVE, in order to detect the participants who actually show CVE, we search for them in rule users.

[3]GCM is precluded, as GCM is known unable to predict CVE.

[4]For fitting the participants data, probability of Category 2 would be transferred to probability of Category 1 by $p(1) = 1 - p(2)$.

$$LogL = \sum_i log \left(\sum_k f_{ik} \right)! - \sum_i \sum_k (logf_{ik})! + \sum_i \sum_k f_{ik} log (p_{ik}),$$
(2)

where k is the number of categories and i is the number of the categorization conditions.

According to the modeling results, the participants can be divided into three groups. See the *AIC* of each model for each group in **Table 1**. The group number is made in accordance with the tendency to predict the critical item as Category 1. Group 1 ($n = 11$) and Group 3 ($n = 12$) are consistent with the decision bound model, except that Group 1 tends not to classify the critical item as Category 1 and Group 3 strongly predicts the critical item as Category 1. Group 2 ($n = 18$) is identified as the MAC strategy user. The observed probability of Category 1 on the critical item made for each group is shown as the bars in **Figure 4**. Here we present the data collected in the C1+Up and C2+Down conditions. This is because the critical item is in between the two categories and it is always larger than a preceding item from Category 1 and smaller than a preceding item from Category 2. Thus, the cases of C1+Down and C2+Up are nearly impossible to happen for the critical item.

Group 1 strongly classifies the critical item as Category 2, mean $p(1) = 0.26$, in either the C1+Up or C2+Down case. The performance of Group 1 in these two cases is not significantly different [$t_{(10)} = 0.10$, $p = 0.92$]. This result is better accommodated by the decision bound model. See the triangle in **Figure 4**, which represents the prediction of the winning model. For Group 1, the winning model is the decision bound model. See **Table 2** for

Table 1 | Model performance (AIC) on fit to transfer performance.

	MAC	Decision bound
Group 1	52.34	43.50
Group 2	74.59	106.29
Group 3	61.84	53.34

FIGURE 4 | The observed and predicted group difference on the critical item.

Table 2 | Best-fitting parameter values.

	MAC	Decision bound	
	c	b	ϵ
Group 1	1.32	0.137	0.035
Group 2	1.10	0.079	0.007
Group 3	1.06	0.244	0.029

the parameter values, which provide best fits. The mean best-fit boundary b is 0.137, which equals 543 *mel*, locating in between the highest edge of Category 1 (520 *mel*) and the critical item (595 *mel*). Consequently, Group 1 shows CVE with no doubt.

Group 2 clearly shows sequence effect on classifying the critical item. On classifying the critical item, when following a Category 1 item $[p(1) = 0.71$ for C1+Up], Group 2 tends to make a response of Category 1, whereas when following a Category 2 item, Group 2 tends to make a response of Category 2 $[p(1) = 0.28$ for C2+Down]. The difference on probability of Category 1 between these two cases is significant $[t_{(17)} = 3.89, p < 0.01]$. The mean probability of Category 1 is about 0.50. The triangle shown for Group 2 in **Figure 4** is the prediction of the MAC model.

Group 3 is a bit tricky, as these participants predict the critical item as Category 1 in the C1+Up case $[p(1) = 0.81]$ and the C2+Down case $[p(1) = 0.75]$. For Group 3, the tendency to make classification for the critical item is not different in different categorization conditions $[t_{(11)} = 0.91, p = 0.38]$. The performance of Group 3 is better fit by the decision bound model. The mean best-fit boundary b is 0.244, which equals 599.56 *mel*. This boundary is larger than the critical item, hence predicting the critical item as Category 1. The decision bound model's prediction for Group 3 can be seen in **Figure 4**. However, this result presumably can also be accommodated by GCM. Since GCM would always predict the critical item as the low-variability category (i.e., Category 1), it is hard to say that Group 3 relies on rule or exemplars for categorization. One thing for sure is that Group 3 does not show CVE and does not rely on some short-term representation for categorization.

To sum up, a number of interesting findings in this experiment are listed as follow. First, CVE does occur in perceptual category learning (i.e., Group 1). Second, although some participants show CVE, some others do not, suggesting clear individual differences. Third, among those participants who do not show CVE, some take on the MAC strategy for categorization (i.e., Group 2) and some can be realized as doing categorization without considering the category variability (i.e., Group 3).

5. GENERAL DISCUSSION

In this study, we would like to figure out why CVE is seldom reported in the past studies. The analysis for the averaged data shows that there is no CVE. This is the same as what is reported in the past studies. We further examine two hypotheses for this result. One hypothesis is that the sequence effect in four categorization conditions, when being combined, would conceal CVE. The other is that the non-CVE report results from mixing up

the uses of different categorization strategies, including the one which shows CVE. Although we find clear sequence effect, individual differences seem to provide a better account for why CVE is seldom reported. We fit the MAC model and the decision bound model to participants' transfer data with the attempt to detect any individual differences. The modeling results show three different groups. Group 1 shows CVE and is consistent with the decision bound model. Group 2 obviously adopts the MAC strategy, as supported by the clear sequence effect. Group 3 again is fit better by the decision bound model. However, this group tends to classify the critical item as the low-variability category.

In spite of positive evidence for CVE, a few constraints of this study need to mention. First, although it should be clear that Group 1 adopts rule for categorization and Group 2 adopts the MAC strategy, it is still not clear which representation, rule or exemplars, Group 3 forms for categorization. Second, we use only one item, namely the critical item, as the probe to examine CVE, that might decrease the power of our experiment. Instead of using one item, a line of novel items between two categories might be better as transfer items. Third, due to the randomization of trial orders, we cannot guarantee that the odds of each of the four categorization conditions (C1+Up, C1+Down, C2+Up, and C2+Down) are the same. Nonetheless, the implications of this study are discussed as follow.

5.1. INDIVIDUAL DIFFERENCES

Of our great interest is the individual differences revealed in this study. Group 1 classifies the critical item as the high-variability category, Group 2 classifies it as both categories depending on which item precedes it, and Group 3 classifies the critical item as the low-variability category. The reason why we have these individual differences might be relevant to the design of category structure and the individual participant's cognitive capacity. As to the category structure, the two categories in our experiment can be perfectly distinguished by a category boundary located on anywhere between them. When the boundary is put close to the low-variability category, we have Group 1, whereas when the boundary is put close to the high-variability category, we have Group 3.

Similarly, the study of Yang and Lewandowsky (2004) showed clear individual differences with a particular category structure, which could be represented by at least two different ways. The categories were constructed in a three-dimensional space, in which one dimension was context and could not directly predict the categories. The perfect learning performance could be achieved via either focusing on the related dimensions, ignoring the context dimension, to generate the true rule for categorization, or generating two different partial 2-D rules for categorization in different contexts. The participants did not know in advance this tricky part of the experiment, yet some of them learned to ignore context and some others learned to apply different rules for categorization in different contexts.

In a following study, the participants who relied on context to generate different rules for categorization were found to have a larger working memory capacity (operational span) than those who ignore context (Yang et al., 2006). This is reasonable, as attending more information does require more cognitive

resource. In addition, a psychometric-approach study provides evidence that working memory capacity which is measured by the tasks of operational span, sentence span, memory updating, and spatial short-term memory is correlated with learning accuracy ($r = 0.44$) (Lewandowsky et al., 2012). Therefore, it is reasonable to expect that working memory capacity might have something to do with the individual differences we observed in this study. At least, we can expect that Group 2 might have a smaller working memory capacity than the other two groups. This is because they only need to retain the preceding item's information for current categorization, that consumes not too much cognitive capacity. The other two groups might need more efforts to generate the rule, which should be suitable for classifying all items. In the future study, the relationship between working memory capacity and category learning performance is worth investigating in more detail.

5.2. SHORT-TERM vs. LONG-TERM CATEGORY REPRESENTATION

Most of contemporary models for category learning posit that categorization is accomplished by some long-term representation. For Group 1 and Group 3 in our study, it is true that some long-term representation must be formed for categorization. It could be a rule or exemplars of categories. Although Group 3 is fit better by the decision-bound model than the MAC model, it presumably is consistent with the prediction of GCM. Nonetheless, for Group 2, it is implied that the short-term exemplar memory might be relied on for categorization. Also, we should be able to find the evidence for the use of short-term representation in other experiments, as long as more one test trial is adopted. In fact, Navarro et al. (2013) recently ask the participants to learn the category structure, which varies along with learning trials. The task is not easy to learn, yet the participants' performance is above the chance level. They also report that the conventional exemplar model and prototype model cannot account for their data. Instead, their data can be fit by a heuristic model, which based on the preceding item to predict the category boundary for the next item. That is, the category boundary keeps shifting from one trail to the next. Together with their finding, the role of short-term representation in categorization should be more emphasized.

5.3. CONCEPTUAL vs. PERCEPTUAL PROCESSING IN CATEGORIZATION

Although the present study provides insights to why CVE was not reported in the perceptual categorization task, we do not think that these findings can properly benefit the conceptual categorization task, as the conceptual and perceptual processing differs substantially. In perceptual categorization, a rule can be defined mathematically as a boundary in the psychological space. Thus, as which category an item would be classified depends on which region in the psychological space the percept of this item locates in.

However, in conceptual categorization, a rule is often a logical statement such as "If necessary feature Y, then category X." For example, an animal with a feature of "being born of cat parents" must be a cat, as our lay theory of animals demands that they must be of the same species as their parents. In the study of Rips (1989), the rule might state "If an object is more than 1 inch in diameter,

it must be a PIZZA," since quarters are severely restricted in size but pizzas are not. The feature "3-inches in diameter" is not characteristic of either PIZZA or QUARTER, but diagnostic of PIZZA, as a pizza can be as small as 3 inches in diameter. As shown in the study of Smith and Sloman (1994), when no characteristic features of QUARTER (e.g., silver colored) are present, the rule-based categorization is triggered and classifies the circular object with a 3-inches diameter as PIZZA. Clearly, CVE with conceptual categories is construed in a very different way.

In addition, in our study, the understanding of each category is established in the trial-by-trial learning experience, whereas the structure of conceptual category reflects our common knowledge of the world, which is acquired out of laboratory. Thus, the MAC strategy is not possible to be applied in the conceptual categorization task. On the other hand, it is expected that the sequence effect or the MAC strategy can be observed in other perceptual category learning tasks.

To sum up, our study provides evidence for the individual differences on classifying the critical item. This is regarded as one reason for why some studies report CVE but some others do not. Also, sequence effect is clearly observed in our experiment, which suggests the use of short-term representation for categorization. However, the success of the decision bound model suggests that long-term representation would also be used for categorization. Therefore, we find evidence for both short-term and long-term representation in a single study. However, it is still not clear why these individual differences occur, or how to induce a particular categorization strategy. These issues need to be addressed in future studies.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control.* 19, 716–723. doi: 10.1109/TAC.1974.1100705

Ashby, F. G., and Gott, R. E. (1988). Decision rules in perception and categorization of multidimensional stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 33–53. doi: 10.1037/0278-7393.14.1.33

Ashby, F. G., and Maddox, W. T. (1992). Complex decision rules in categorization: contrasting novice and experienced performance. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 50–71. doi: 10.1037/0096-1523.18.1.50

Ashby, F. G., and Townsend, J. T. (1986). Varieties of perceptual independence. *Psychol. Rev.* 93, 154–179. doi: 10.1037/0033-295X.93.2.154

Cohen, A., Nosofsky, R., and Zaki, S. (2001). Category variability, exemplar similarity, and perceptual classification. *Mem. Cogn.* 29, 1165–1175. doi: 10.3758/BF03206386

Erickson, M. A., and Kruschke, J. K. (1998). Rules and exemplars in category learning. *J. Exp. Psychol. Gen.* 127, 107–140. doi: 10.1037/0096-3445.127.2.107

Hsu, A. S., and Griffiths, T. L. (2010). "Effects of generative and discriminative learning on use of category variability," in *Proceedings of 32nd Annual Conference of the Cognitive Science Society* (Portland, OR).

Johansen, M., and Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cogn. Psychol.* 45, 482–553. doi: 10.1016/S0010-0285(02)00505-4

Kruschke, J. K. (1992). Alcove: an exemplar-based connectionist model of category learning. *Psychol. Rev.* 99, 22–44. doi: 10.1037/0033-295X.99.1.22

Kruschke, J. K. (1993). Human category leanring: impications for backpropagation models. *Connect. Sci.* 5, 3–36. doi: 10.1080/09540099308915683

Lee, M., and Wetzels, R. (2010). "Individual differences in attention during category learning," in *Proceedings of The 32nd Annual Conference of Cognitive Science Society* (Portland, OR: Cognitive Science Society), 387–392.

Lewandowsky, S., Yang, L.-X., Newell, B. R., and Kalish, M. (2012). Working memory does not dissociate between different perceptual categorization tasks. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 881–904. doi: 10.1037/a0027298

Maddox, W. T., and Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Percept. Psychophys.* 53, 49–70. doi: 10.3758/BF03211715

Navarro, D. J., Perfors, A., and Vong, W. K. (2013). Learning time-varying categories. *Mem. Cogn.* 41, 917–927. doi: 10.3758/s13421-013-0309-6

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol. Gen.* 115, 39–57. doi: 10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1987). Attention and learning processes in the identification and classification of integral stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 87–108. doi: 10.1037/0278-7393.13.1.87

Nosofsky, R. M., Clark, S. E., and Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 282–304. doi: 10.1037/0278-7393.15.2.282

Nosofsky, R. M., and Johansen, M. K. (2000). Exemplar-based accounts of multiple-system phenomena in perceptual categorization. *Psychon. Bull. Rev.* 7, 375–402.

Nosofsky, R. M., Palmeri, T. J., and McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychol. Rev.* 101, 53–79. doi: 10.1037/0033-295X.101.1.53

Rips, L. J. (1989). *Similarity, Typicality, and Categorization.* New York, NY: Cambridge University Press.

Sakamoto, Y., Jones, M., and Love, B. C. (2008). Putting the psychology back into psychological model: mechanistic versus rational approaches. *Mem. Cogn.* 36, 1057–1065. doi: 10.3758/MC.36.6.1057

Smith, E. E., and Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Mem. Cogn.* 22, 377–386. doi: 10.3758/BF03200864

Steinberg, J. C. (1937). Positions of stimulation in the cochlea by pure tones. *J. Acoust. Soc. Am.* 8, 176–180. doi: 10.1121/1.1915891

Stevens, S. S., Volkmann, J., and Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *J. Acoust. Soc. Am.* 8, 185–190. doi: 10.1121/1.1915893

Stewart, N., and Brown, G. D. A. (2004). Sequence effects in the categorization of tones varying in frequency. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 416–430. doi: 10.1037/0278-7393.30.2.416

Stewart, N., Brown, G. D. A., and Chater, N. (2002). Sequence effects in categorization of simple perceptual stmuli. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 3–11. doi: 10.1037/0278-7393.28.1.3

Stewart, N., and Chater, N. (2002). The effect of category variability in perceptual categorization. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 893–907. doi: 10.1037/0278-7393.28.5.893

Yang, L.-X., and Lewandowsky, S. (2004). Knowledge partitioning in categorization: constraints on exemplar models. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1045–1064. doi: 10.1037/0278-7393.30.5.1045

Yang, L.-X., Lewandowsky, S., and Jheng, W. (2006). "Working memory and knowledge partitioning in categorization," in *The 4th International Conference on Memory* (Sydney, NSW).

Logical-rules and the classification of integral dimensions: individual differences in the processing of arbitrary dimensions

Anthea G. Blunden[†], Tony Wang[†], David W. Griffiths and Daniel R. Little*

Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, VIC, Australia

A variety of converging operations demonstrate key differences between separable dimensions, which can be analyzed independently, and integral dimensions, which are processed in a non-analytic fashion. A recent investigation of response time distributions, applying a set of logical rule-based models, demonstrated that integral dimensions are pooled into a single coactive processing channel, in contrast to separable dimensions, which are processed in multiple, independent processing channels. This paper examines the claim that arbitrary dimensions created by factorially morphing four faces are processed in an integral manner. In two experiments, 16 participants completed a categorization task in which either upright or inverted morph stimuli were classified in a speeded fashion. Analyses focused on contrasting different assumptions about the psychological representation of the stimuli, perceptual and decisional separability, and the processing architecture. We report consistent individual differences which demonstrate a mixture of some observers who demonstrate coactive processing with other observers who process the dimensions in a parallel self-terminating manner.

Keywords: integrality, separability, serial vs. parallel, coactivation, holistic processing, categorization, computational modeling, reaction time

INTRODUCTION

Understanding how our perceptual systems process multidimensional stimuli provides fundamental insights into basic cognitive operations such as categorization (Ashby and Gott, 1988; Fifić et al., 2010; Little et al., 2011), object representation (Folstein et al., 2013), and recognition memory (Nosofsky et al., 2011, 2012). Of critical importance is the difference between stimuli that consist of either separable or integral perceptual dimensions. Separable dimensions are those which can be attended to and analyzed in isolation, such as size and shape (Attneave, 1950; Torgenson, 1958; Shepard, 1964; Garner, 1974, 1978). In contrast, integral dimensions are thought to be psychologically "fused," such that one integral dimension cannot be attended to at the expense of the other; both must be processed together (Garner, 1974; Burns and Shepp, 1988).

Although many stimulus dimensions have been studied in the information processing literature, research demonstrating the integrality of stimulus dimensions has focused primarily on the dimensions of brightness and saturation of Munsell colors for visual stimuli (Shepard and Chang, 1963; Garner, 1974; Nosofsky, 1987; Shepard, 1987; Burns and Shepp, 1988; Nosofsky and Palmeri, 1996; Fifić et al., 2008; Little et al., 2013) and pitch and loudness for auditory stimuli (Grau and Kemler-Nelson, 1988). Though these dimensions meet several empirical criteria for integrality (defined further below), there is also a sense in which these dimensions are easily used to form a mental representation of the stimuli; that is, given a set of stimuli which vary in brightness and saturation, individuals are likely to form a psychological

representation of the stimuli using dimensions which correspond to brightness and saturation. Consequently, these dimensions are *psychologically privileged* and fall short of Grau and Kemler-Nelson's (1988) notion of the "extreme-end" of integrality, where the individual dimensions are unable to be accessed at all.

More recently, Goldstone and Steyvers (2001; see also Gureckis and Goldstone, 2008; Hendrickson et al., 2010; Folstein et al., 2012; Jones and Goldstone, 2013) have utilized a set of morph dimensions which are thought to have no perceivable dimensional structure yet still meet the empirical criteria for integrality; consequently, these arbitrarily-defined morph stimuli may fulfill Grau and Kemler-Nelson's (1988) notion of an "extreme" integral stimulus. This renders these morphs useful for studying the difference between integral and separable dimensions. In this paper, we test whether these arbitrarily-defined morph dimensions demonstrate evidence of integrality in a task which goes beyond the classic converging operations by utilizing not only mean response time (RT) and choice comparisons, but also analysis of the full RT distributions and the time course of information processing. Our measure thus provides a more nuanced understanding of integrality than previous empirical criteria.

CONVERGING EMPIRICAL OPERATIONS FOR INTEGRALITY

There are a number of converging operations suggesting that integral dimensions are processed differently from separable dimensions (Garner, 1974):

(1) The distances between stimuli derived from proximity estimates (e.g., similarity ratings, identification confusions and so on) using multidimensional scaling (MDS) are better described by an Euclidean distance metric if the dimensions are integral but by a city-block distance metric if the dimensions are separable (Attneave, 1950; Torgenson, 1958; Shepard, 1964, 1987; Nosofsky, 1992).

(2) People tend to sort integral-dimensioned stimuli based on overall similarity but separable-dimensioned stimuli based on individual dimensions (Imai and Garner, 1968; Handel and Imai, 1972; Garner, 1974).

(3) Learning to attend to important attributes takes place more efficiently for separable-dimensioned stimuli (Shepard et al., 1963; Posner, 1964; Nosofsky, 1986) than for integral-dimensioned stimuli (Shepard and Chang, 1963; Nosofsky, 1987; Nosofsky and Palmeri, 1996).

(4) Integral dimensions, but not separable dimensions, tend to interfere with each other if one of the dimensions must be ignored, but tend to facilitate one another if the dimensions are varied in a correlated manner (Lockhead, 1966; Egeth, 1967; Garner, 1969; Garner and Felfoldy, 1970; Biederman and Checkosky, 1970; Garner, 1974).

Each of these operations suggests that integral dimensions are processed as an entire object (Lockhead, 1966, 1972), but separable dimensions are processed as independent, component parts of an object.

Despite this wealth of converging operations, Cheng and Pachella (1984) argue that integrality may be an artifact of testing perceptual dimensions which do not correspond to an observer's psychological representation. For example, results showing a failure of converging operations (e.g., an interference effect between purported integral dimensions but no facilitation effect, (Garner, 1974; see also Biederman and Checkosky, 1970; Levy and Haggbloom, 1971; Gottwald and Garner, 1975; Pomerantz and Sager, 1975; Smith and Kemler, 1978) reduce the "explanatory power" of the concept of integrality (Cheng and Pachella, 1984, p. 283). In order to conclusively demonstrate integrality, Cheng and Pachella (see also Grau and Kemler-Nelson, 1988) argue that one must demonstrate that the experimenter-defined and participant-defined dimensions are commensurate *and* that the dimensions still satisfied the empirical criteria for integrality. Obviously, this presents a problem for empirically justifying the integrality of dimensions at the extreme-end of integrality which are meant to be without perceivable dimensional structure.

ARBITRARY DIMENSIONS AND INTEGRALITY

One possible set of dimensions that might satisfy the criteria of being both integral and having no identifiable dimensional structure, are the factorially-generated morph dimensions shown in **Figure 1** (top panel). These stimuli are created by morphing together four base faces (e.g., Goldstone and Steyvers, 2001). The morphed stimuli vary on two dimensions, with each of these dimensions representing the transition between two of the base faces (faces A–D in **Figure 1**). Hence, each stimulus can be defined by its proportional value on each of the morph dimensions, but the morph dimensions are very difficult to analyze independently.

The dimensions are termed arbitrary because, although each stimulus varies systematically along two face morph axes, the face morph axes do not correspond to any naturally interpretable dimensions.

Goldstone and Steyvers (2001) showed that the morph dimensions demonstrated an interference effect in the filtration condition of the Garner (1974) speeded classification task, supporting the claim that the dimensions are processed in an integral fashion. Furthermore, Folstein et al. (2012) found that there was no advantage for learning an orthogonal boundary compared to a diagonal boundary in a factorially-generated morph space such as the space shown in **Figure 1** (although it is important to note that Folstein et al., used morph cars and not morphed faces). Taken together these results indicate the arbitrary morph dimensions seem to fulfill Grau and Kemler-Nelson's (1988) criteria for the extreme-end of integrality.

Despite the large number of converging operations to identify integrality, we argue that these operations are, in fact, somewhat equivocal with regard to the actual theoretical mechanism underlying the processing of integral dimensions. For example, there have been suggestions that integrality is a continuum from completely integral to completely separable (Torgenson, 1958; Shepard, 1964; Lockhead, 1972; Garner, 1974; Foard and Kemler, 1984; Grau and Kemler-Nelson, 1988; Melara and Marks, 1990) and that separable stimuli, with practice, may become integral over time (Ashby and Maddox, 1991; Goldstone, 2000; Blaha et al., 2009). Consequently, it is unclear whether integral dimensions are always processed in a consistent fashion, especially for those dimensions which, unlike brightness and saturation or pitch and loudness, may not involve "a positive correlation between the ranges of variation of stimuli associated with important consequences" in the environment (Shepard, 1991, p. 68). Indeed, many purportedly integral dimensions are not perfectly described by a Euclidean metric, but instead by a metric somewhere in-between city-block and Euclidean (Grau and Kemler-Nelson, 1988). Hence, the converging operations typically used to identify integrality do not always converge.

Furthermore, some converging operations, such as finding slower RTs in Garner's (1974) classic filtration task when compared to the corresponding control task, are open to multiple interpretations about the underlying processing architecture. For instance, in a filtration task, the number of stimuli is increased from two to four stimuli compared to the control condition. Like the control task, only one of the dimensions is relevant for classification, and the increased RT in the filtration task compared to the control task is taken as evidence that the variation on the irrelevant dimension interferes with selective attention to the relevant dimension. Such a result is used to diagnose integrality. However, rather than reflecting interference due to irrelevant variation, the increase in RT in the filtration task might simply reflect increased confusability due to the increased number of stimuli (Maddox, 1992). Indeed, increased RTs in a filtration task have been reported for stimuli that appear to be nominally separable (Shepp, 1989).

Determining whether the arbitrary morph dimensions are, in fact, processed coactively is a fundamental question, as a number of important learning results are predicated on this assumption
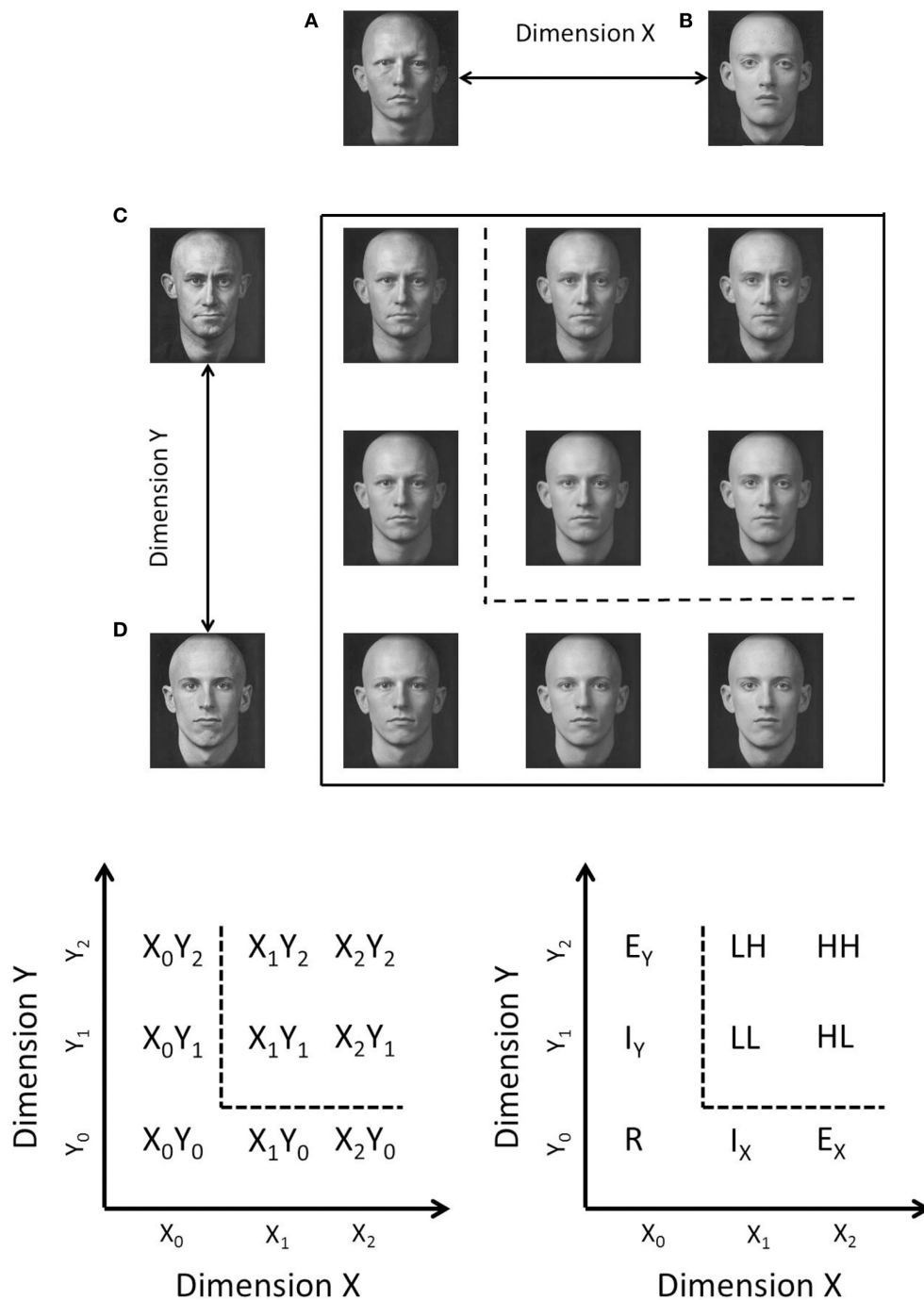
FIGURE 1 | Top: Example of morph stimuli for Experiment 1. Each of the dimensions are created by morphing between two base faces. Each morph stimulus is a proportional mixture of all four base faces. **Bottom:** Schematic illustration of category space indicating the nomenclature used in the text. Stimuli which lie above and to the right of the decision boundary (dotted line), belong to the *target* category (category A), stimuli which low below and to the left of the decision boundary belong to the *contrast* category (category B). Stimuli in the target category are referred to by their salience which can be high (H) or low (L) depending on whether an item is far from or close to the category boundary, respectively. Contrast category items are referred to as internal (I), external (E), and redundant (R) depending on their positions in the stimulus space.

(e.g., Goldstone and Steyvers, 2001; Gureckis and Goldstone, 2008; Hendrickson et al., 2010; Jones and Goldstone, 2013). For example, Goldstone and Steyvers (2001) trained participants to categorize face morphs using a single orthogonal category boundary; then in a second phase, transferred participants to a new boundary which was either a 90° or 45° rotation of the originally trained boundary. Participants were able to perform more accurately with the new 90° boundary than with the 45°

boundary suggesting that the initially integral morph dimensions were differentiated into two orthogonal dimensions which mapped directly onto the dimensions used to create the stimuli. Although, the morph dimensions were not confirmed to be processed separably (e.g., using a Garner interference task) after training, better performance with the 90° boundary rotation than the 45° rotation suggests that the dimensions are "psychologically privileged" after training. This effect provides strong empirical evidence that learning changes perception by creating a featural or dimensional vocabulary which perceptual processes can use for future learning and decision making (Goldstone, 1998; Goldstone et al., 2000, 2008). The emergence of psychologically privileged dimensions, termed *differentiation,* has been suggested as one of the key perceptual changes underlying human development from infancy (Smith, 1989; Goldstone et al., 2011) and the development of expertise (Burns and Shepp, 1988).

This finding is somewhat controversial as other researchers have found that differentiation does not occur with other integral dimensioned stimuli (e.g., "blobs" created via the convolution of sine waves in polar coordinates varying in amplitude and frequency; Op de Beeck et al., 2003) or even other morph dimensions created using a different morphing technique (i.e., by blending four base stimuli rather than factorially combining the base stimuli as in **Figure 1**; see Folstein et al., 2012, for a detailed explanation of the difference). By contrast, Hockema et al. (2005) found that differentiation did occur for blob stimuli if an adaptive learning procedure, which started with categorization of the easiest items and increased the difficulty of the task by gradually moving the selection of items closer to the category boundary, was used.

In this paper, we investigate whether the morph stimuli used to demonstrate differentiation (Goldstone and Steyvers, 2001; Folstein et al., 2012) are initially processed in an integral fashion by examining a more theoretically motivated test of integrality than previously used for these stimuli. We draw on two theoretical frameworks for understanding integrality. The first, General Recognition Theory (GRT; Ashby and Townsend, 1986) grew out of the signal detection theory tradition (Green and Swets, 1966) but allowed for rigorous theoretical definition of several empirically defined notions of independence and separability (both perceptual and decisional). The second, logical rule models of categorization (Fifić et al., 2010), utilizes the representational

concepts from GRT but combines these representations with processing assumptions based on sequential sampling models (Ratcliff, 1978; Busemeyer, 1985) and information processing approaches to response time (Kantowitz, 1974; Townsend and Ashby, 1983; Townsend, 1984). A further aim of this paper is to investigate the combination of assumptions necessary for explaining an individual's categorization decisions using these face morph stimuli.

THEORETICAL FRAMEWORKS FOR UNDERSTANDING SEPARABILITY AND INTEGRALITY
General recognition theory
General Recognition Theory (Ashby and Townsend, 1986) is a multivariate generalization of signal detection theory (Green and Swets, 1966). In this framework, each stimulus is represented by a distribution, often a bivariate or multivariate normal distribution, capturing the mean location of the stimulus in a multidimensional perceptual space as well as the perceptual variability associated with that stimulus. A theory of categorization decisions is made possible in this framework by assuming that a decision boundary is established in the category space (Ashby and Gott, 1988) and integrating the perceptual distribution in each category region. This value provides the probability with which a particular categorization decision is made given a particular stimulus.

GRT provides a theoretical unification of differing ideas about *perceptual independence*, *perceptual separability* and *decisional separability*. For example, the category space shown in **Figure 2A** (GRT PS + DS) shows the isoprobability contours for nine two-dimensional stimuli. The isoprobability contours can be thought to represent a top view of a slice through the bivariate normal distributions representing each stimulus. Note that the distributions are circular representing the idea that there is no statistical correlation between the perceived values of the dimensions. This absence of correlation is termed *perceptual independence* and is a construct which refers to a single stimulus.

By contrast, separability and integrality are constructs which refer to collections of stimuli. To explain, *perceptual separability* occurs when the mean locations, and variability, of the stimuli are aligned along a dimension making it possible to represent the collection of the stimuli by the same marginal distribution along that dimension. Note that perceptual separability can occur with or without perceptual independence. A violation of perceptual

FIGURE 2 | (A) Isoprobability contours when perceptual separability and decisional separability hold. **(B)** Isoprobability contours when there is a violation of perceptual separability due to mean shift integrality but decisional separability holds. **(C)** Isoprobability contours with mean shift integrality and an optimal decision bounds.
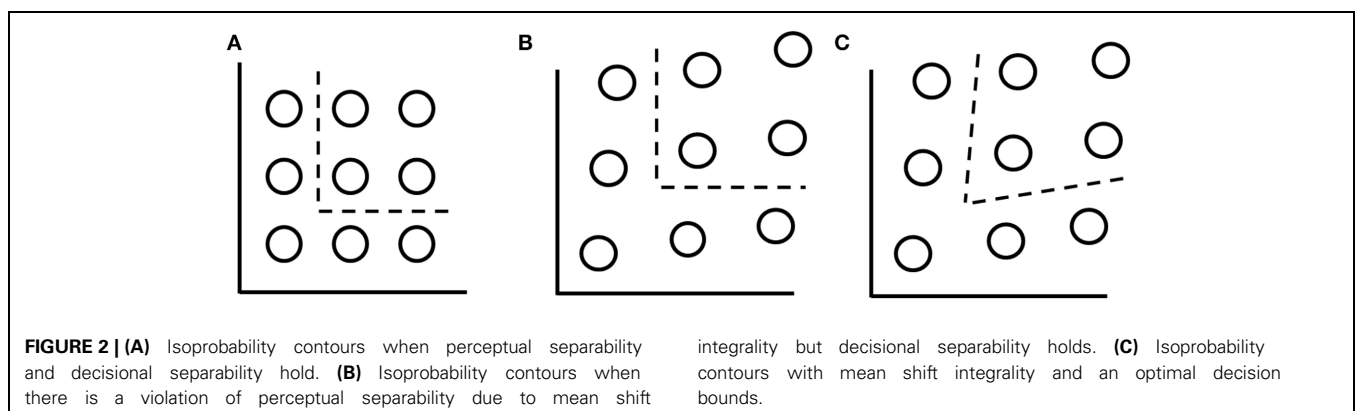
separability occurs if the perceptual effect of one dimension is affected by the level of another dimensions. Although there are many ways in which this can occur, two of these violations are through varying the means of the distributions, termed *mean shift integrality*, or by altering the variances between the stimuli, termed *variance shift integrality* (Ashby and Maddox, 1994). **Figures 2B,C** illustrate mean shift integrality. In contrast to variation of the stimulus characteristics, *decisional separability* refers to the alignment of decision bounds with the dimensional axes of the stimuli. When decisional separability holds, the decision bound is orthogonal to the dimensional axis to which it applies. By contrast, violations of decisional separability occur when the boundaries are not orthogonal. For instance, in **Figure 2C**, the placement of the decision boundaries at an optimal orientation with respect to the stimuli represents a violation of decisional separability.

These constructs are important and useful because they provide a quantitative framework which can be used to predict some of the different empirical operations which differentiate performance with integral and separable dimensions; though predicting the response time effects in, for instance, Garner's (1974) classic experiments, requires auxiliary assumptions about how RTs are generated. For instance, Maddox (1992) adopted the RT-distance hypothesis which posits that RTs are a monotonically decreasing function of the distance of a stimulus from the decision boundary (Ashby and Maddox, 1991). Within this framework, facilitation for integral dimensioned stimuli when there is correlated variation between dimensions can then be explained by assuming optimal decision boundaries. By contrast, interference effects due to irrelevant dimensional variation can be explained by an increase in perceptual variability.

Nosofsky and Palmeri (1997) examined these predictions by examining the full RT distributions from a replication of Garner's (1974) conditions. These authors argued that if perceptual variability increases with irrelevant variation, then under the RT-distance hypothesis the fastest RTs from the filtration condition should be faster than in those in a control condition (with no irrelevant variation). That is the increase in perceptual variability would mean that some proportion of the RTs would be generated when the perception of the stimulus was further from the decision boundary than in a control condition. Nosofsky and Palmeri's results, however, showed that RTs were slower overall with irrelevant variation at all quantiles of the RT distribution. This result argues against the RT-distance hypothesis (see also Nosofsky and Little, 2010). However, coupling the GRT framework with other mechanisms for generating response times, such as sequential sampling models, does not make this prediction since the integrated distribution can be thought to provide a "drift rate" which represents the evidence for which a stimulus belongs to each category (cf., Ashby, 2000; Fifić et al., 2010). Furthermore, new theoretical insight can be gained by combining GRT with mental architecture approaches to understanding when stimulus dimensions are processed independently and when they are pooled together into a single process.

In summary, in the present work, we utilize the representational assumptions defined in GRT but couple these with processing-based assumptions that allow us to predict RTs for each item in the task. This is a novel departure from GRT because it allows a theoretical definition of integrality which is not based on the representation of the stimulus dimensions but on how those dimensions are processed. In the following section, we present coactivity (i.e., the pooling of information from all stimulus dimensions into a common processing channel) as a plausible theoretical definition of how integral dimensions are processed.

Coactivity as a theoretical definition of integrality

A novel, theoretically-driven definition of integrality can be achieved by directly contrasting the information processing of multidimensional stimuli. In particular, by using factorial experiments and analyzing full RT distributions, one can differentiate between processing which analyzes each of the dimensions independently (i.e., either in serial or in parallel) and processing which pools the dimensions together into a single processing channel (hereafter, termed *coactive* processing; Townsend and Nozawa, 1995; Townsend and Wenger, 2004). Independent channel processing and coactive processing provide a novel theoretical distinction between separability and integrality that coheres with the traditional definitions of these concepts that emphasize analytic vs. non-analytic or holistic processing.

Using a combination of non-parametric analyses and parametric response time models, Little et al. (2013 see also Fifić et al., 2008; Fifić and Townsend, 2010; Little et al., 2011) demonstrated that integral dimensions of brightness and saturation are pooled into a single, coactive processing channel, but separable dimensions, such as brightness and size, are processed independently and in multiple channels. In this paper, we test whether the arbitrarily-defined face morph dimensions also demonstrate coactivity. Before turning to our experimental results, we first briefly introduce our methodology, the logical-rule models framework, which allows identification of independent channel and coactive processing, and in turn, we describe how our experiment implements this methodology.

Logical-rule model framework

The logical rule-based models (Fifić et al., 2010) synthesize the representational assumptions of GRT and decision-bound theory (Ashby and Townsend, 1986; Ashby and Gott, 1988), along with sequential sampling (e.g., random walk models; Ratcliff, 1978; Townsend and Ashby, 1983; Busemeyer, 1985; Luce, 1986; Link, 1992; Ratcliff and Rouder, 1998) and mental architecture frameworks (e.g., serial vs. parallel; Sternberg, 1969; Kantowitz, 1974; Townsend, 1984; Schweickert, 1992). The models are best explained with reference to the stimulus space shown in **Figure 1**. In this space, nine face-morph stimuli are created by orthogonally combining two dimensions, each varying in three levels.

The four stimuli in the upper right quadrant, which are assigned to the *target* category, Category A, factorially combine an easy or *high discriminability* (H) boundary decision and a difficult or *low discriminability* (L) boundary decision across two dimensions; hence, the four target category stimuli are referred to as LL, LH, HL, and HH. The target category is defined by a conjunctive rule; that is, a stimulus must have a

value on dimension X greater than the vertical category boundary *and* a value on dimension Y greater than the horizontal boundary to belong to the target category. Because the stimuli in the target category must satisfy both rules, the dimensions of these stimuli must be processed *exhaustively* (i.e., both dimensions must be processed before a target category decision can be made).

Like GRT, the logical rule-based models (Fifić et al., 2010) assume that the perception of each stimulus dimension is represented by a normal distribution of perceptual effects. In order to make a decision, evidence is sampled from these distributions and used to drive a random walk process (see **Figure 3**). More specifically, following decision-bound theory (Ashby and Townsend, 1986; Ashby and Gott, 1988), observers are assumed to establish a decision boundary (represented by the dashed line in **Figure 3**) to separate Category A and Category B. In order to make a category decision the observer samples from the stimulus distribution using a random walk process. A sample from Category A, for example, will lead to a step toward the criterion +A. This process

of evidence accumulation continues until a criterion is reached. The logical-rule models assume that the closer a stimulus is to a decision boundary in space, the more difficult it is to classify, and therefore the larger the RT.

The possible combinations of separate random-walk processes can be described using three mental architectures (i.e., serial, parallel, and coactive). For serial and parallel processes, two separate random walks occur, each driven by samples from each separate dimension. These independent random walks can occur in a serial or parallel fashion. In the case of a self-terminating stopping rule, the dimension that finishes first determines the final categorization decision and RT. In the case of an exhaustive stopping rule, however, final categorization decisions and RTs are determined by the output of both random walks.

In contrast to serial and parallel processing, coactive processing assumes that a single random walk model is driven by samples from a joint bivariate normal distribution on both dimensions X and Y. At each time step, a sample is drawn from the bivariate distribution representing the particular stimulus. If the sample falls
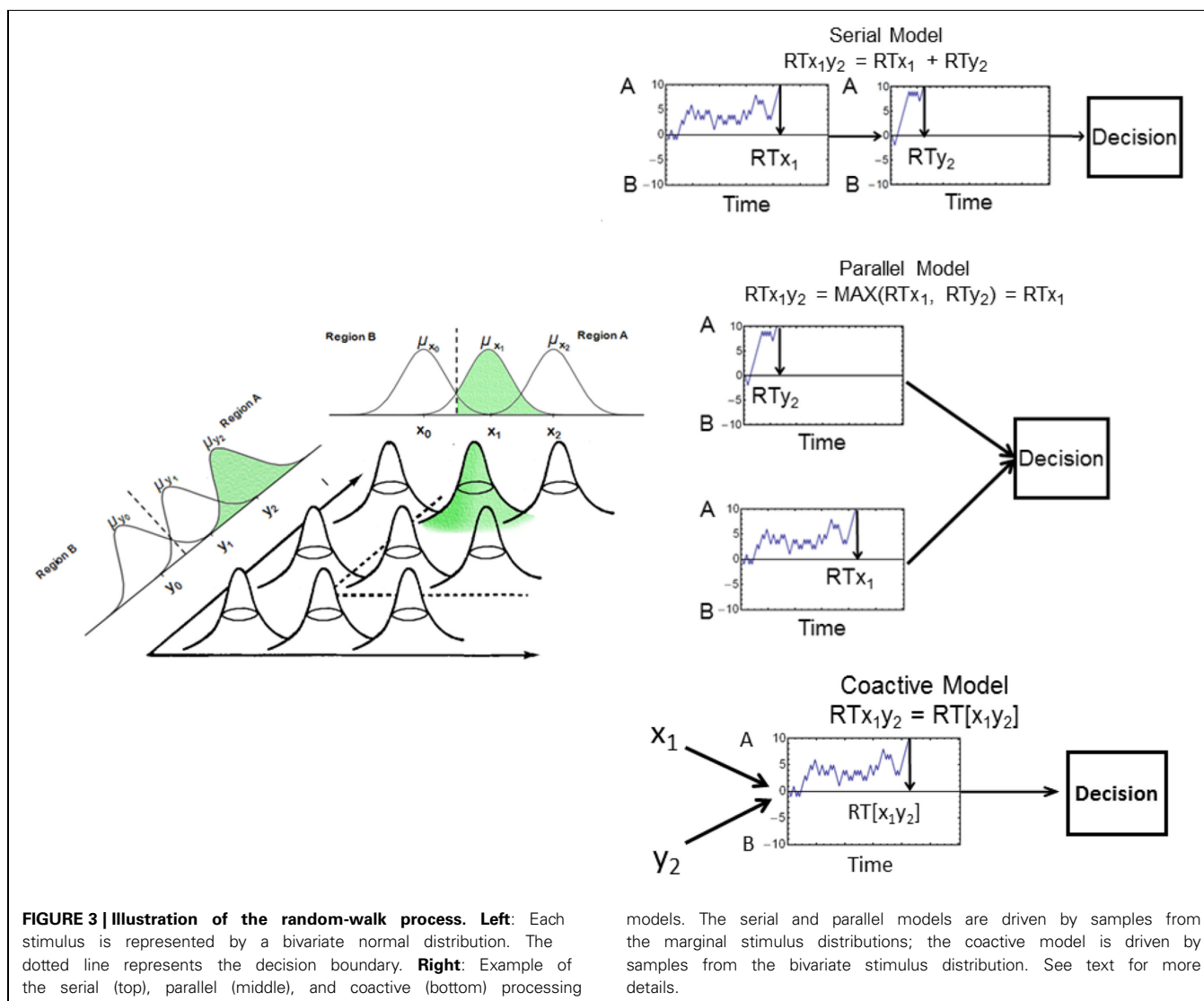
FIGURE 3 | Illustration of the random-walk process. Left: Each stimulus is represented by a bivariate normal distribution. The dotted line represents the decision boundary. **Right**: Example of the serial (top), parallel (middle), and coactive (bottom) processing

models. The serial and parallel models are driven by samples from the marginal stimulus distributions; the coactive model is driven by samples from the bivariate stimulus distribution. See text for more details.

in the Category A region, the model will take a step toward the decision criterion +A. However, if the sample falls in the Category B region, the random walk will take a step toward the decision criterion −B. This single, pooled random-walk process continues until one of the criteria is reached.

ANALYSIS OF MODEL PREDICTIONS

As described by Fifić et al. (2010), the *double factorial* combination of the dimensional values in the target category allows us to leverage several non-parametric measures known as *Systems Factorial Technology* (SFT; Townsend and Nozawa, 1995; Townsend and Wenger, 2004) to qualitatively differentiate the candidate models. For example, the mean interaction contrast (MIC) and survivor interaction contrasts (SIC) can be used to differentiate serial, parallel, and coactive information processing architectures. These non-parametric analyses require correct stochastic ordering (i.e., *stochastic dominance*) for items in the target category. To explain, the RT for the HH face is expected to be faster than RT for the LL face since the former is further away from the category boundary than the latter. In order for the qualitative predictions to provide meaningful diagnostic information, the RTs for the HL and LH faces should be between the HH and LL faces. This ordering is reflective of the effective selective influence (Townsend and Nozawa, 1995; Heathcote et al., 2010; see also Schweickert et al., 2000; Dzhafarov, 2003; Dzhafarov et al., 2004; Dzhafarov and Gluhovsky, 2006) of each of the dimensions on the RT. Under the condition of selective influence, the MIC and SIC provide an empirically-observable, non-parametric measure which speaks directly to theoretical questions about the processing architecture and the underlying stopping rule.

Piloting of the experimental stimuli revealed that most participants demonstrated a violation of stochastic dominance, even after extended categorization training. Consequently, the current experiments will not report the SFT analyses to differentiate between information processing architectures. Instead, we will only fit RT distributions to the logical-rule models, and utilize model comparison to differentiate between mental architectures. (Further information about these analyses is available from the authors upon request).

PROCESSING DIFFERENCES FOR SEPARABLE AND INTEGRAL-DIMENSIONED STIMULI

To date, a number of different dimensions and stimulus manipulations have been analyzed using this logical-rules framework. Across experiments, the largest differences in processing have been observed between separable-dimensioned and integral-dimensioned stimuli. For instance, when the stimulus dimensions were separable and located in spatially-separated locations (Fifić et al., 2010; Little et al., 2011) processing of the dimensions was best explained by a serial and self-terminating model. When separable dimensions were spatially overlapped (Little et al., 2011; Experiment 2), processing was best described as a trial-by-trial mixture of serial and parallel processing. By contrast, when stimulus dimensions were integral (i.e., Munsell colors varying in brightness and saturation; Fifić et al., 2008; Little et al., 2013), processing conformed to the predictions of the coactive model.

To highlight the large effects of separability and integrality on processing, it is worthwhile noting that several manipulations had very little effect on processing (Fifić et al., 2010; Little et al., 2011). For instance, with separable dimensions, processing was serial regardless of whether observers were given the rule that defined the categories upfront, whether the rule had to be learned via trial-by-trial feedback, whether observers were instructed to focus on responding quickly or on responding accurately, and whether the dimensions were spatially separated or part of a single object (cf. Fifić et al., 2010; Little et al., 2011).

RELATIONSHIP TO GRT's DEFINITIONS OF SEPARABILITY AND INTEGRALITY

In previous studies, the application of the logical rule models has always assumed perceptual independence, perceptual separability, and decisional separability. In those studies, the full RT distributions from the entire collection of stimuli from both categories could be accounted for by varying only the architecture used to determine how the information from each dimension was integrated over time. Little et al. (2013) tested whether allowing mean shift integrality and diagonal decision boundaries would allow, for instance, a parallel model to mimic a coactive model when fitting the integral dimensioned data. In that analysis, mean shift integrality was introduced by shifting the means of the stimuli so that they lied on a tilted parallelogram rather than a square grid. Even with this systematic violation of perceptual separability, neither a serial model nor a parallel model could mimic the coactive model's predictions.

Nonetheless, it is reasonable that less systematic shifts in stimulus location might require allowing for violations of perceptual separability and decisional separability. In the following, we analyze the RT distributions from individual categorization responses using the face morph stimuli shown in **Figure 1**. In analyzing this data, we fit several models which allow for differences in processing architecture (serial, parallel, and coactive), stopping rule (self-terminating vs. exhaustive) as well as violations of perceptual and decisional separability. To limit the scope of the project, in addition to the categorization data, we also collected similarity ratings for each pair of stimuli which we use to derive an MDS solution that can inform whether perceptual separability holds or is violated. For example, by constraining the MDS solution to lie on a grid (e.g., Borg and Groenen, 2005) we enforce perceptual separability, but by allowing the mean locations of the stimuli to vary, we capture any violations of perceptual separability.[1] The MDS solutions also act as a further independent empirical assessment of stimulus integrality since we can also test whether the scaling solution is better fit using a city-block or Euclidean metric (Attneave, 1950; Torgerson, 1958; Shepard, 1964, 1987; Nosofsky, 1992). Our approach therefore combines three major theoretical

---

[1] We do not examine variance shift integrality (or other violations of perceptual separability) in this paper because when coupled with the decision boundary, the effect of changing the mean or changing the variance of a perceptual distribution in the logical rule models is to change the probability that the random walk takes a step up or down toward the +A or −B boundary. We considered it unlikely that we would be able to differentiate these two accounts using the present design and instead leave that for future research.

approaches to understanding separability and integrality: GRT, MDS and the logical-rule modeling framework.

Finally, we also assumed that the decision boundaries might be either orthogonal to the decision axes or rotated to capture the optimal discrimination between stimuli from the target and contrast categories. Consequently, for each of the mental architectures, we tested three different sets of the assumptions about the perceptual representation:

(1) By assuming perceptual separability (represented by using stimulus coordinates found using a constrained MDS solution) and decisional separability (orthogonal decision bounds).
(2) By assuming violations of perceptual separability (by using an unconstrained MDS solution) and decisional separability (orthogonal decision bounds).
(3) By assuming violations of both perceptual and decisional separability (represented by using stimulus coordinates found using an unconstrained MDS solution and by allowing optimal decision boundaries).

EXPERIMENT 1

We examined a set of purportedly integral stimuli created from arbitrary morph dimensions. By using the conjunctive category design shown in **Figure 1**, we test whether the morph stimuli are processed in a coactive fashion or whether the morph dimensions are better described by an independent channel processing model (i.e., parallel or serial processing). We utilized these face morphs in both an upright and inverted orientation to extend the generalizability of our basic procedure. There is a possibility that upright faces are processed holistically, whereas inverted faces are not (Yin, 1969). However, there is a dimensionality to these face morphs which is relevant for categorizing both the upright and inverted faces (i.e., unlike for, say, recognizing upright vs. inverted faces in daily life), and consequently, we do not *a priori* expect a difference between them.

METHOD
Participants
Eight participants from the University of Melbourne community with normal or corrected-to-normal vision were randomly assigned into the *upright* condition and the *inverted* condition with four in each condition (labeled U1–U4 and I1-I4 for the upright and inverted conditions, respectively). Participants received $12 for each session plus an extra $3 bonus for accurate performance (over 90% accuracy) during categorization sessions. All procedures were approved by the University of Melbourne Human Ethics Advisory Group.

Apparatus and stimuli
A category space was created using a field morphing technique (Steyvers, 1999), to morph four base faces together into a two-dimensional array (i.e., each dimension was a systematic blend from one face to a second face; **Figure 1**), creating a 3 × 3 matrix of faces, that are composed of factorial proportions of each of the four base faces. The base faces used in this study were identical to base faces used in Goldstone and Steyvers (2001, Experiment

1; Kayser, 1984). Dimension X was formed using the morph between faces C and D and Dimension Y was formed using the morph between faces A and B (see **Figure 1**). Each face in the stimulus space can be defined by a factorial combination of values on Dimension X and Dimension Y. Stimuli in the inverted condition were rotated 180°, but were otherwise identical. The stimuli were presented at a monitor resolution of 1280 × 1024 and subtended a visual angle of approximately 10°. RTs for categorization sessions were collected using a calibrated response time box (Li et al., 2010).

PROCEDURE
Categorization
Each participant completed a series of 1-h sessions on consecutive or near consecutive days for five sessions. At the beginning of each session, participants were shown experimental instructions, including example stimuli relevant to their condition (i.e., upright or inverted faces).

Each session consisted of 819 trials (9 practice trials and 810 experimental trials, divided into 9 blocks of 90 trials). Although each stimulus was presented 10 times during each block, presentation of stimuli was randomized. In between each block, participants were instructed to take a short break and were given feedback on their percentage accuracy. Participants advanced to the next block by pressing any button on the RT box. During each trial a fixation cross was presented for 1170 ms. After 1070 ms a warning tone was presented for 700 ms. A face was then presented and the participant was required to decide whether the face belonged to Category A or Category B. Faces were presented until a response was made. Feedback was provided only after incorrect responses; feedback "too slow" was provided for RTs greater than 5000 ms.

Similarity ratings
We ran a similarity rating study using Amazon Mechanical Turk to obtain similarity ratings for the faces shown in **Figure 1**. In two conditions, participants rated the similarity of the stimuli in either the upright or inverted condition of Experiment 1. A single Human Intelligence Task (HIT) was created on Amazon Mechanical Turk with 40 assignments. We restricted access to the HIT by requiring users to have at least a 90% acceptance rate (i.e., 90% of a user's completed HITs were accepted by the requester), having completed at least 1000 approved HITs, and were located in the United States. Participants were paid $2.00 USD to complete the task, which took approximately 25 min to complete. Allocation of participants to conditions was random; this resulted in 20 participants in upright condition and 20 participants in the inverted condition.

On each trial, a pair of stimuli was presented in the upper-left and upper-right of the screen. Subjects rated the similarity of each pair from 1, "least similar" to 8 "most similar." Subjects were instructed to try to use the full range of ratings, and were given examples of high, medium, and low similarity pairs using a different set of upright faces before commencing the task. For each condition, there were 36 unique pairings of the 9 stimuli. Each pair was presented six times for each subject; the order of presentation was completely randomized as

was the left-right presentation of each face. The experiment was self-paced.

RESULTS

For the categorization task, any trials with RTs less than 200 ms or greater than 3 SDs above the mean were removed from the analysis. No trials were removed using this method. The first session was considered practice and discarded from these analyses. Mean RTs and error rates for each participant are reported in **Table 1**. In the upright condition, error rates across items were low; only three items showed error rates above 10% (LH and E_X for U2, and LH for U3). As expected the greater difficulty in processing of inverted faces resulted in higher error rates for all four participants in the inverted condition. Participants I1 and I2

showed high error rates across all items ($>20\%$), with very poor accuracy for items HL, LH, LL, and E_X and E_Y. Overall error rates for participants I3 and I4 were comparatively lower (12 and 16% respectively). Similar to I1 and I2, items HL, LL, E_X, and E_Y were poorest for I3 and I4. All four participants showed high error rates for item LL. This is unsurprising since LL lies adjacent to both decision boundaries.

COMPUTATIONAL MODELING

Multidimensional scaling of similarity ratings

We first sought to identify participants who utilized the entire rating scale as instructed; consequently, we computed the multinomial likelihood of the counts of each rating value 1 to 8 (i.e., across all pairs) assuming that responses were (a) generated

Table 1 | Mean RTs and error rates for each stimulus.

Subject	HH	HL	LH	LL	Ex	Ix	Ey	Iy	R
EXPERIMENT 1									
Mean RTs									
U1	846.65	1045.90	1092.50	1083.70	749.29	757.55	1034.40	834.62	739.03
U2	726.59	824.60	928.07	762.62	779.13	781.31	946.26	787.91	707.82
U3	643.29	812.05	766.35	741.77	626.60	627.55	708.08	604.06	525.17
U4	504.02	570.28	582.97	535.11	492.51	497.92	520.67	524.22	450.07
I1	820.75	840.61	888.08	949.38	910.41	815.71	858.14	769.13	746.85
I2	764.86	851.34	878.78	924.04	1036.00	830.62	752.88	740.32	685.86
I3	1362.50	1656.30	1459.90	1847.00	1534.70	1653.50	1624.00	1528.20	1551.00
I4	978.05	1424.90	1198.40	1394.00	1292.90	1152.80	1341.50	1376.10	966.11
Error rates									
U1	0.03	0.08	0.06	0.23	0.10	0.01	0.16	0.00	0.00
U2	0.03	0.18	0.14	0.20	0.12	0.04	0.12	0.18	0.01
U3	0.00	0.00	0.04	0.03	0.01	0.01	0.01	0.00	0.00
U4	0.02	0.03	0.12	0.08	0.04	0.01	0.09	0.07	0.00
I1	0.05	0.06	0.21	0.17	0.03	0.00	0.10	0.30	0.03
I2	0.01	0.02	0.03	0.10	0.11	0.01	0.06	0.02	0.00
I3	0.01	0.15	0.07	0.08	0.11	0.09	0.09	0.03	0.00
I4	0.00	0.03	0.05	0.08	0.04	0.02	0.09	0.01	0.01
EXPERIMENT 2									
Mean RTs									
U5	738.05	769.17	790.8	853.55	792.6	696.99	821.56	701.51	642.98
U6	842.46	890.03	962.3	957.15	867.32	830.68	866.29	906.27	777.98
U7	577.65	641.28	711.48	782.32	653.74	635.23	677.29	637.93	549.97
U8	1105.4	1514	1662.4	1713.8	1302.1	1097	1311.1	1643	983.33
I5	861.98	1000.6	1021.1	1056.7	787.89	717.57	940.36	1048.7	813.29
I6	752.23	916.65	918.9	1202.3	998.28	737.51	1020.9	821.31	664.97
I7	917.49	1410	1183.3	1696.5	1082.7	1188.7	1235.6	1121.1	951.46
I8	829.3	1038.3	1006.8	1140.5	965.36	1038.9	1016.6	921.6	910.21
Error rates									
U5	0.01	0.02	0.08	0.09	0.01	0.00	0.03	0.02	0.00
U6	0.00	0.03	0.12	0.01	0.02	0.01	0.11	0.02	0.00
U7	0.01	0.08	0.12	0.03	0.02	0.10	0.07	0.03	0.00
U8	0.00	0.04	0.05	0.01	0.01	0.02	0.02	0.02	0.00
I5	0.09	0.24	0.25	0.58	0.47	0.14	0.34	0.10	0.04
I6	0.07	0.19	0.26	0.60	0.47	0.13	0.14	0.06	0.01
I7	0.01	0.12	0.08	0.29	0.13	0.18	0.18	0.07	0.03
I8	0.00	0.29	0.06	0.40	0.19	0.09	0.14	0.21	0.02

uniformly for each rating value, (b) assuming that responses were sampled primarily from only one rating value and (c) assuming that responses were sampled primarily from only two rating values. That is, each of these assumptions was used to generate a prior probability of selecting each of the response options [e.g., (a) with equal probability for each response option, (b) with most of the probability on one response option, or (c) with most of the probability spread across two response options]. Using these prior probability distributions and a multinomial likelihood, we computed the posterior probability for each hypothesis given the observed distribution of counts across rating values, using Bayes' rule. We then removed any observer with a posterior probability less than 0.5 for the uniformly distributed rating hypothesis. This resulted in the removal of two participants from the upright condition and six participants from inverted condition

We computed the averaged similarity rating for each pair of stimuli and found the two-dimensional scaling solutions for each condition. This was done by fitting the averaged ratings using a model which assumed a negative linear relationship between the predicted similarity ratings and the Euclidean distance between the estimated coordinates. To find the best fitting coordinates, we minimized the sum-of-squared deviations between the predicted and observer ratings from 100 starting points chosen to span the coordinate space. There were 20 parameters in total (the nine coordinate values, and the slope and the intercept of the negative linear distance-to-similarity function) used to fit the 36 similarity ratings. The estimated two-dimensional-scaling solution accounted for 97 and 99% of the variance in the averaged ratings for the upright and inverted conditions, respectively. To display the scaling solutions, we first performed a Procrustes rotation (Borg and Groenen, 2005) to the ideal coordinate values (see **Figure 1**). The rotated scaling solutions for the upright and inverted condition are shown in **Figure 4**. In general, both the inverted and upright scaling solutions conformed to the ideal category space outlined in **Figure 1**. In the upright condition, the scaling solution showed a pattern whereby the interior stimuli are positioned further from the (presumed location of) the orthogonal boundary compared to the exterior stimuli. In the inverted condition, the overall shape of the scaling solution is best described by a parallelogram. In particular, both the interior and exterior stimuli of the A-B and C-D morph dimensions appear to "slope" away from the orthogonal boundary.

For each condition, we also fitted a scaling solution that constrained each of the nine co-ordinates to a 3 × 3 grid. This model only had six free parameters and allowed only the distance between values on the A,B and C,D morph dimensions to vary. This constrained scaling solution accounted for 85 and 79% of the variance in the averaged ratings for the upright and inverted conditions, respectively. As explained above, the constrained and unconstrained scaling solutions allow for the examination of whether changing the perceptual representation affects the model fitting.

Finally, we fitted additional scaling solutions that assumed city-block distance instead of Euclidean distance between the estimated coordinates. The unconstrained model accounted for 94 and 98% of the variance in the averaged ratings for the
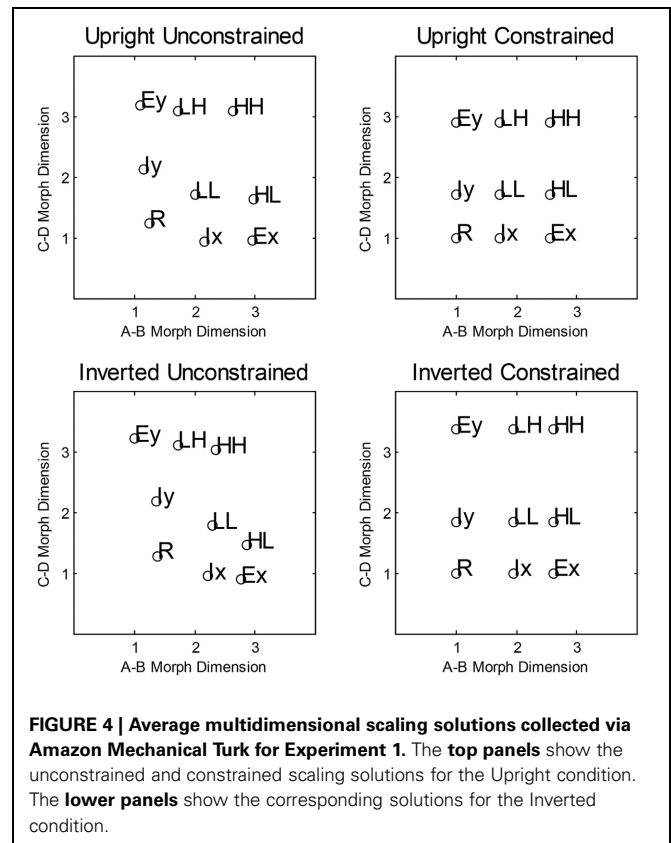
FIGURE 4 | Average multidimensional scaling solutions collected via Amazon Mechanical Turk for Experiment 1. The **top panels** show the unconstrained and constrained scaling solutions for the Upright condition. The **lower panels** show the corresponding solutions for the Inverted condition.

upright and inverted conditions, respectively. In contrast, the constrained model accounted for 77 and 73% of the variance in the upright and inverted conditions. As illustrated in **Table 2**, the models assuming city-block distance provided worse fitting scaling solutions than the models assuming Euclidean distance. Consequently, better fitting scaling solutions with a Euclidean distance metric suggests that these face morph dimensions are integral dimensions.

Model fitting

Having established the coordinate values from the scaling analysis, we then estimated, for each model, the variances of the perceptual distributions, the decision boundaries, and the random walk parameters. For simplicity, we assumed equal variance across all levels of a given dimension, but allowed for differences in the variances between dimensions. As illustrated in **Figure 4**, the unconstrained scaling solution for both conditions deviates greatly from the ideal 3 × 3 grid layout. Given that the logical-rule models (Fifić et al., 2010) utilize the representational assumptions of GRT (Ashby and Townsend, 1986; Ashby and Gott, 1988), we can use the GRT framework to fit models that vary in the assumption of the perceptual representation of the stimuli.

We fitted three sets of models, each set containing the five possible logical-rule models, which accounted for violations of perceptual and/or decisional separability. The first set of models allowed violations of perceptual separability but maintained the assumption of decisional separability; we label this set of models

Table 2 | Summary of the fits of the scaling models for Experiment 1 and 2.

Condition	Model	Cityblock			Euclidean		
		SSD	BIC	R^2	SSD	BIC	R^2
EXPERIMENT 1							
Upright	Full	4.73	−27.80	0.94	2.51	**−50.56**	0.97
	Constrained	17.25	−31.40	0.77	11.45	−46.15	0.85
Inverted	Full	1.00	−66.97	0.98	0.59	**−85.72**	0.99
	Constrained	12.86	−24.99	0.73	9.88	−34.48	0.79
EXPERIMENT 2							
U5	Full	1.06	10.26	0.82	1.21	15.01	0.79
	Constrained	2.92	−3.43	0.50	2.82	**−4.71**	0.52
U6	Full	2.76	−15.76	0.91	3.17	−10.76	0.90
	Constrained	12.23	−12.29	0.61	9.27	**−22.28**	0.70
U7	Full	6.52	−28.17	0.94	4.44	−41.98	0.96
	Constrained	21.38	−35.57	0.80	10.10	**−62.56**	0.90
U8	Full	4.88	−31.62	0.94	3.27	−45.99	0.96
	Constrained	21.04	−29.18	0.76	12.07	**−49.20**	0.86
I5	Full	1.37	**−82.87**	0.99	1.55	−78.31	0.98
	Constrained	11.38	−56.72	0.89	9.65	−62.65	0.90
I6	Full	4.57	−27.17	0.94	2.93	−43.22	0.96
	Constrained	13.85	−37.43	0.81	9.90	**−49.53**	0.86
I7	Full	11.32	−21.13	0.92	8.92	−29.73	0.94
	Constrained	35.59	−30.07	0.76	28.51	**−38.04**	0.81
I8	Full	3.94	−16.67	0.91	2.93	−27.30	0.94
	Constrained	11.83	−27.29	0.74	8.26	**−40.20**	0.82

SSD, Sum of Squared deviations; BIC, Bayesian Information Criterion.
The best model for each observer is shown in bold.

MSI and DS for *mean shift integrality* and *decisional separability*. In effect, these models were fitted using the unconstrained scaling solutions and assumed orthogonal decision bounds. The second set of models assumed both perceptual and decisional separability (hereafter, PS and DS). These models were fitted using the constrained scaling solutions. The third family of models assumed both violations of perceptual and decisional separability (hereafter, MSI and OP, because the boundaries were rotated to an optimal orientation). A diagonal decision boundary was estimated using the unconstrained scaling solution. We freely estimated for each participant and each model perceptual variances, σ_X and σ_Y, and decision boundaries, D_X and D_Y, for Dimensions X and Y, respectively. For the optimal decision bound models, the slope (in degrees) of the decision boundaries along the X and Y dimensions was calculated prior to model fitting. The intercepts of these bounds (called *Offset1* and *Offset2*) were estimated as free parameters and they replaced parameters D_X and D_Y from the previously described models. For the random walk components of the models, we freely estimated response criteria $+A$ and $-B$. We also assumed an additional non-decision time (i.e., time associated with encoding and movement time) was generated from a log-normal distribution \with location, μ_r, and scale, σ_r and added to the decision time generated from the random walk. We further assumed that each step in the random walk was scaled to milliseconds by a multiplicative scaling constant, k. Hence, each of the logical rules models has nine free parameters.

The sole exception is the serial self-terminating model for which we also estimated the probability that dimension X was processed before dimension Y, p_X.

We fitted the models simultaneously to the correct-RT distributions and the error rates for each item by using quantile-based maximum likelihood estimation (Heathcote et al., 2002). For each item, correct RT predictions were generated for the 10, 30, 50, 70, and 90% quantiles. We did not attempt to fit the error-RT distributions since error rates were generally low. The fit of the models to the data was given using the multinomial log-likelihood function:

$$\ln L = \sum_{i=1}^{n} \ln(N_i!) - \sum_{i=1}^{n} \sum_{j=1}^{m+1} \ln(f_{ij}!) + \sum_{i=1}^{n} \sum_{j=1}^{m+1} f_{ij} \cdot \ln(p_{ij})$$

where N_i is the total number of times each item i ($i = 1, n$) was presented, f_{ij} is the frequency with which item i had a correct RT in the jth bin ($j = 1, m$) or was an error response ($m + 1$), and p_{ij} is the predicted probability that each item i had a correct RT in the jth bin or was an error. We compared each model's log-likelihood adjusted for model complexity using the Bayesian information criterion (BIC; Schwarz, 1978). The complexity penalty in the BIC is based on the number of free parameters and the size of the sample as follows:

$$\text{BIC} = -2 \ln L + n_p \ln(M),$$

where n_p is the number of free parameters and M is the total number of observations in the sample. Models with smaller BIC values are preferred. Predictions were generated by simulating 10,000 RTs for each item; details of the simulation method for each model are given in Fifić et al. (2010, pp. 311–317; numerical methods for generating model predictions are given in Little, 2012). The model fits for each subject in the upright and inverted conditions are shown in **Table 3** and the parameters of the best fitting model are shown in **Table 4**.

Upright condition

Table 3 shows the best fitting model (serial, parallel, or coactive) for each participant within each set of models. Inspection of **Table 3** shows that the coactive model was the best fitting model for all participants in the models assuming MSI and DS. When both PS and DS was assumed, the parallel self-terminating was the best fitting model for three of four participants (U1, U2, and U4); the serial exhaustive model best fits U3 within this set of models. However, when MSI and OP were assumed, the parallel self-terminating model provided the best fit for all four observers in the upright condition.

Overall, there was a consistency of the best fitting model (parallel self-terminating or coactive) within each set of models. That is, we can rule out serial processing and, for the most part, any exhaustive processing, which accords with previous findings regarding integral dimensioned stimuli (Little et al., 2013) and stimuli with dimensions in the same spatial location (Little et al., 2011). However, in considering the best fitting model for each individual participant across all stimulus sets, there were marked individual differences. For instance, the parallel self-terminating model, was the best fitting model for participants U1 BIC = 546.51) and U2 (BIC = 821.08), and the coactive model was the best fitting model for U3 (BIC = 753.72) and U4 (BIC = 505.62). The assumption of PS also varied between these participants. The best fitting model assumes MSI and DS for U3 and U4, but the best fitting models assume PS and DS for U1, and MSI and OP for U2. The predictions of the best fitting parameters are plotted against individual RT distributions in **Figure 5**.

Inverted condition

For the inverted condition, the coactive model was the best fitting model for all participants in the two sets of models that assume perceptual integrality (regardless of decisional separability or integrality). For the set of models that assume both PS and

Table 3 | Model Fits to subjects in Experiment 1 (model with the lowest BIC in each set is bolded; best overall model is bolded an italics).

Subject	Coactive		Parallel exhaustive		Parallel Self-terminating		Serial exhaustive		Serial self-terminating	
	-lnL	BIC	-lnL	BIC	-lnL	BIC	-lnL	BIC	-lnL	BIC
PS AND DS										
U1	290.07	616.04	471.95	979.8	**255.26**	**546.41**	380.75	797.41	308.04	655.97
U2	437.07	910.04	448.34	932.58	**419.45**	**874.79**	489.27	1014.4	443.56	927
U3	462.32	960.55	533.06	1102	452.12	940.13	**427.55**	**891.01**	441.19	922.27
U4	314.04	663.98	442.85	921.6	**264.83**	**565.55**	384.25	804.39	341.48	722.84
I1	*231.62*	*499.14*	376.03	787.96	267.22	570.33	514.06	1064	280.39	600.67
I2	*215.65*	*467.2*	400.75	837.4	256.6	549.09	531.55	1099	253.04	545.97
I3	306.93	649.76	288.92	613.75	*258.91*	*553.72*	366.95	769.8	258.42	556.73
I4	**289.2**	**614.29**	413.1	862.1	236.11	508.11	542.26	1120.4	268.18	576.24
MSI AND DS										
U1	**277.82**	**591.54**	418.40	872.69	307.71	651.31	349.33	734.56	309.52	658.93
U2	**417.36**	**870.61**	486.76	1009.40	419.94	875.78	470.07	976.04	425.09	890.06
U3	*358.91*	*753.72*	579.67	1195.20	444.50	924.90	462.10	960.09	446.36	932.61
U4	*234.86*	*505.62*	387.66	811.23	263.47	562.84	396.34	828.59	344.65	729.19
I1	**291.25**	**618.41**	536.66	1109.20	435.17	906.25	639.92	1315.70	457.87	955.64
I2	**296.89**	**629.68**	616.82	1269.50	481.39	998.68	716.87	1469.60	498.94	1037.80
I3	**325.64**	**687.19**	515.23	1066.40	494.15	1024.20	561.77	1159.40	504.63	1049.20
I4	*208.42*	*452.73*	558.00	1151.90	424.44	884.77	608.31	1252.50	503.77	1047.40
MSI AND OP										
U1	277.46	590.82	380.43	796.76	**263.48**	**562.86**	302.37	640.64	894.47	1828.84
U2	413.13	862.16	472.74	981.38	*392.59*	*821.08*	459.87	955.64	1026.32	2092.53
U3	373.81	783.52	541.82	1119.54	**366.98**	**769.86**	369.01	773.92	1115.98	2271.84
U4	275.87	587.64	419.63	875.16	**274.94**	**585.78**	358.00	751.90	1011.85	2063.59
I1	**331.61**	**699.12**	456.14	948.18	427.00	889.90	569.69	1175.28	586.19	1212.28
I2	**444.62**	**925.14**	610.14	1256.18	605.07	1246.04	726.03	1487.96	791.77	1623.42
I3	**466.31**	**968.52**	588.70	1213.30	585.75	1207.40	727.57	1491.04	786.34	1612.56
I4	**307.18**	**650.26**	645.95	1327.80	631.75	1299.40	817.53	1670.96	808.54	1656.97

Table 4 | Parameters for the best fitting model for subjects in Experiment 1 and 2.

Subject	Set	Model	-lnL	BIC	Dx	Dy	σx	σy	+A	+B	μ$_r$	σ$_r$	k
EXPERIMENT 1													
U1	PS and DS	ParallelST	255.26	546.41	1.35	1.59	0.76	0.19	5	3	6.38	0.36	38.71
U2	MSI and OP	ParallelST	392.59	821.08	1.34	1.73	0.46	0.25	4	3	6.15	0.23	45.07
U3	MSI and DS	Coactive	358.91	753.72	1.40	1.27	0.19	0.24	3	2	5.58	0.38	99.09
U4	MSI and DS	Coactive	234.86	505.62	1.43	1.28	0.13	0.13	3	2	5.81	0.16	50.40
I1	PS and DS	Coactive	231.62	499.14	1.18	0.50	4.17	2.98	4	3	6.13	0.14	37.60
I2	PS and DS	Coactive	215.65	467.20	1.44	0.50	1.61	6.12	3	3	6.04	0.16	53.86
I3	PS and DS	ParallelST	258.91	553.72	1.47	1.51	2.27	2.78	6	7	6.87	0.49	15.68
I4	MSI and DS	Coactive	208.42	452.73	0.76	1.08	1.83	7.87	7	7	6.34	0.17	18.36
EXPERIMENT 2													
U5	PS and DS	Coactive	289.90	615.70	2.20	1.51	4.25	2.98	9	5	6.31	0.17	7.48
U6	PS and DS	ParallelST	253.97	543.83	2.50	2.48	2.02	0.95	3	3	6.09	0.14	56.98
U7	MSI and DS	ParallelST	244.73	525.37	1.68	1.57	1.89	1.57	8	8	5.97	0.02	9.48
U8	MSI and DS	ParallelST	338.97	713.84	1.66	1.43	1.53	1.24	6	6	5.69	0.07	48.81
I5	MSI and OP	ParallelST	304.50	644.89	1.66	1.59	2.88	2.89	5	6	6.19	0.09	18.44
I6	MSI and DS	Coactive	290.12	616.13	1.46	1.16	1.35	0.46	5	4	5.96	0.14	47.22
I7	PS and DS	ParallelST	261.63	559.17	2.50	2.50	1.98	1.37	5	5	5.93	0.06	46.72
I8	PS and DS	ParallelST	299.58	635.06	2.50	2.38	1.52	0.99	6	5	6.24	0.18	26.94

For U2 and I5, Dx and Dy refer to Offset1 and Offest2, the slope of decision bound for dimension X and Y. The value of Offset1 and Offset2 are -2.21 and 97.45°, respectively, for U2, and -1.21 and 86.79° for I5.

DS, the coactive model was the best fitting model for participant I1, I2, and I4 but the parallel self-terminating model was the best model for I3.

Examining the best model across all model sets, participants I1 (BIC = 499.14) and I2 (BIC = 467.20) demonstrated coactive processing under the assumption of PS and DS. Under the same assumptions, the parallel self-terminating model was the best model for I3 (BIC = 553.72). Finally, I4 (BIC = 452.73) demonstrated coactive processing under the assumptions of MSI and DS. The predictions of the best fitting parameters are plotted against individual RT distributions in **Figure 6**.
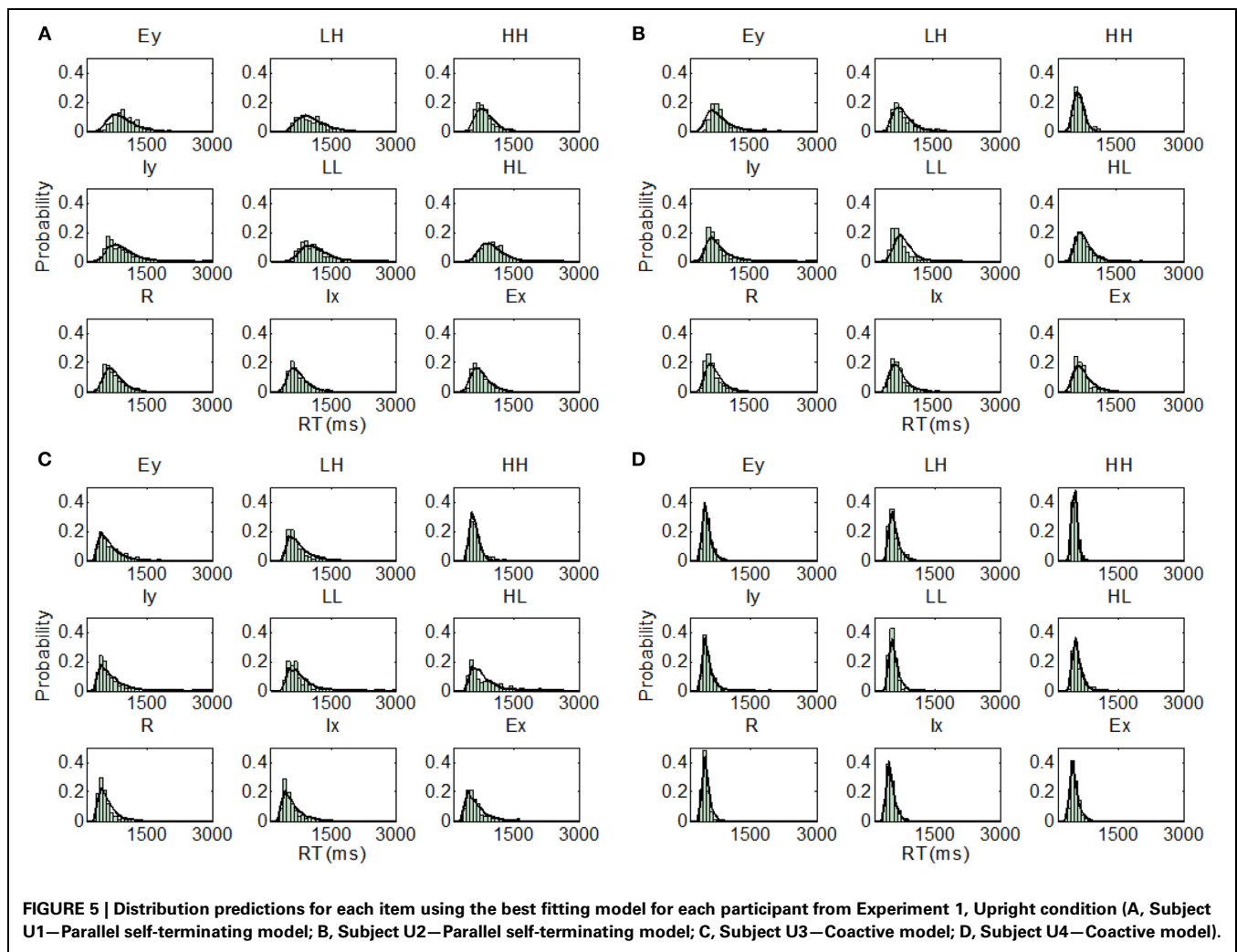
In each of the logical rule models there are two key components which determine the types of predictions that are generated. The first component is the architecture of the model. The second component is the psychological representation of the stimuli, which can vary based on the nature of perceived similarity between each of the stimuli. For the current set of stimuli, we fitted a series of models by varying the assumption of perceptual and decisional separability. It is clear that changing these assumptions affects the best model for each participant. A benefit of the parametric approach taken here is that we are able to test these different assumptions in a systematic fashion.

DISCUSSION

Experiment 1 highlighted two important findings. First, there were individual differences in the processing of the face morph dimensions. In the general, participants in the upright and inverted conditions were best explained by either the coactive or parallel self-terminating models. Specifically, two of four participants processed the face morphs coactively in the upright condition, and three of four participants showed coactivity in the inverted condition.

Second, the best fitting model for each participant varied with changes in the perceptual representation of the stimuli. In the upright condition for example, the coactive model provided the best fit for all participants when the perceptual representation was not assumed to conform to a 3 × 3 grid-layout (see **Figure 1**) and when an orthogonal decision boundary was utilized. However, a parallel self-terminating model best fitted these participants when the model assumed an optimal (diagonal) category boundary. This highlights the necessity of accounting for not only architecture, but also the perceptual representation of the stimuli.

A potential caveat on this interpretation is that the scaling solution was obtained from averaged similarity ratings of online participants. Given the individual differences in processing architecture, it is highly possible that there are also individual differences in the psychological representation of the face morphs shown in **Figure 1**. For example, averaging the similarity data might result in greater symmetry than is observed in any of the individual participants (Ashby et al., 1994); furthermore, the results from the average data may exhibit properties which are not found in any of the individual participants. Consequently, using a single scaling solution for the computational modeling of individual participant data may mask individual differences in the MDS, and possibly also, in processing architecture. A better method would be to fit an MDS model such as INDSCAL, which allows for differential dimension weightings for each observer (Carroll and Chang, 1970). However, this would have still necessitated using an MDS solution collected from observers different from those who completed our categorization task. As an alternative, we conducted a second experiment in which in which RT distributions and scaling solutions were obtained for each participant. For this experiment, we also varied

FIGURE 5 | Distribution predictions for each item using the best fitting model for each participant from Experiment 1, Upright condition (A, Subject U1—Parallel self-terminating model; B, Subject U2—Parallel self-terminating model; C, Subject U3—Coactive model; D, Subject U4—Coactive model).

the stimulus parameters to further increase the generality of our results.

EXPERIMENT 2

Experiment 2 replicated the upright and inverted conditions of Experiment 1 with two important alterations. First, a different stimulus space was created by swapping the positions of the two of the base faces from the set used in Experiment 1. The result of this change in base faces is that all of the stimuli except for E_Y, LL, and E_X are different in Experiment 2 than in Experiment 1 (though similar because they are comprised of the same four base faces). Second, each participant completed a session of similarity ratings following their categorization sessions. Thus, participant-specific scaling solutions were used in the computational modeling.

METHOD
Participants

Eight participants from the University of Melbourne community with normal or corrected-to-normal vision were randomly assigned into the *upright* condition and the *inverted* condition with four in each condition (labeled U5–U8 and I5-I8 for the upright and inverted conditions, respectively). Participants

received $12 for each session plus an extra $3 bonus for accurate performance (over 90% accuracy) during categorization sessions.

Apparatus and stimuli

The apparatus was identical to Experiment 1. The base faces used to create the stimulus space were also identical to those used in Experiment 1, however, the positions of base faces A and C were swapped. This led to a morph sequence between faces A and D, and B and C. This resulted in a different stimulus space, which was nonetheless similar as it comprised the same base faces (see **Figure 7**). The stimuli were presented at four degrees of visual angle.

Procedure

The procedure was identical to the categorization sessions of Experiment 1. Each participant completed five 1-h sessions on consecutive or near consecutive days, and only the final four sessions of categorization were used for analysis. In order to improve overall performance accuracy, participants were first shown the entire stimulus space with decision boundaries removed and were instructed take some time to study these faces to improve their performance during the experiment.

FIGURE 6 | Distribution predictions for each item using the best fitting model for each participant from Experiment 1, Inverted condition (A, Subject I1—Coactive model; B, Subject I2—Coactive model; C, Subject I3—Parallel self-terminating model; D, Subject I4—Coactive model).

FIGURE 7 | Comparison of stimulus spaces for Experiment 1 and 2.
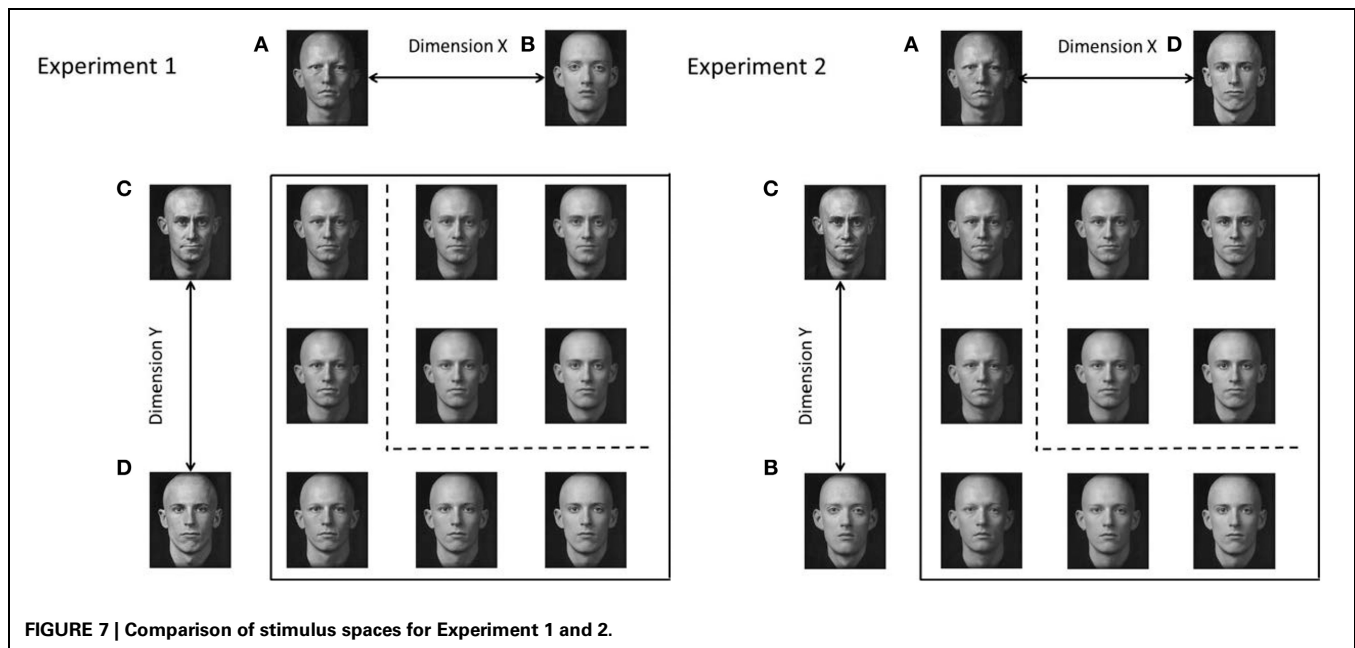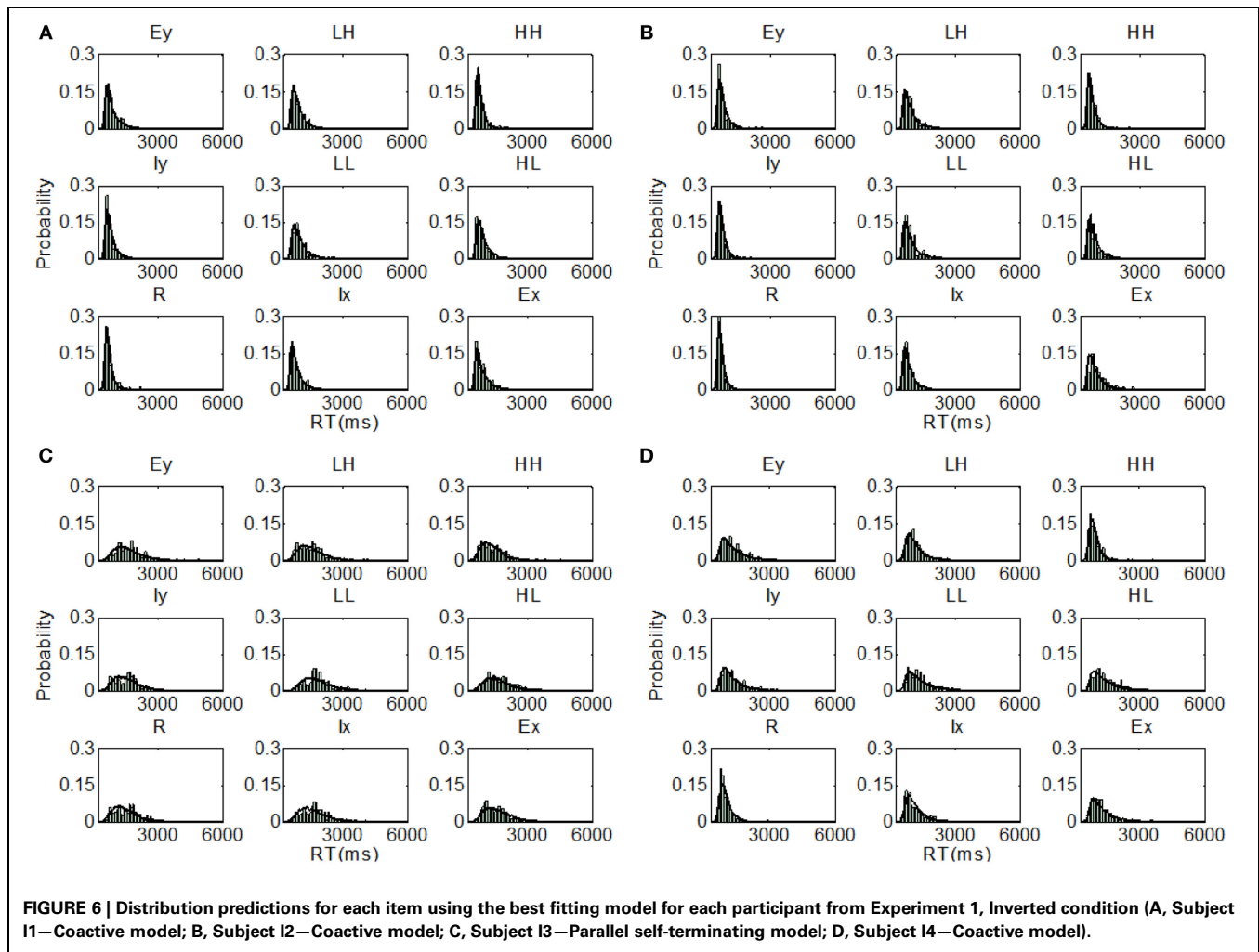
After completing the categorization sessions, participants were asked to return for a subsequent 1 h session in which they rated the similarity of the morphed faces used in the categorization task. There were 36 unique combinations of these stimuli, which were presented to participants 20 times each. On each of the 720 trials, a fixation cross was presented for 500 ms, then one of the combinations of faces was presented (i.e., two faces appeared on the screen, one face in the center of the upper right quadrant and the other in the center of the upper left quadrant of the monitor) and participants were then asked to rate the faces on the number pad using a scale of 1–8, where 1 was least similar and 8 was most similar. The presentation order of each unique pair was counterbalanced across the 20 repetitions. Comparisons were randomized for each participant. Participants in the *upright* condition made similarity judgments for upright faces, and participants in the *inverted* condition made similarity judgments for inverted faces.

RESULTS

For the categorization task, any trials with RTs less than 200 ms or greater than 3 SDs above the mean were removed from the analysis. This resulted in the removal of less than 1% of trials. The mean RTs and error rates are shown in **Table 1**, respectively. Overall, the error rates for the upright and inverted conditions were lower in Experiment 2 compared to Experiment 1, with comparable error rates between the upright and inverted conditions in Experiment 2. This shows that accuracy was approximately equal between conditions for this experiment. As seen in Experiment 1, error rates for stimulus LL in Experiment 2 were generally higher than the remaining eight stimuli.

Multidimensional scaling of similarity ratings

The scaling solutions for participants in the upright and inverted conditions are presented in **Figure 8**. Overall, scaling solutions for each participant in the upright condition adhered to the general layout presented in **Figure 1**. However, participants U5 and U6 demonstrated greater deviations from the grid-layout than U7 and U8. Moreover, the unconstrained solutions revealed violations of perceptual separability for all four participants, as values on the A–D morph dimension changes with each level of the B,C morph dimension. A similar pattern of results was observed for participants in the inverted condition. Participants I6–I8 showed a perceptual representation in which items LL and I_X were lower on the B,C dimension than the corresponding items at that level (i.e., I_Y and HL, and R and E_X). Participant I5, however, showed a pattern in which the items were more dispersed along the B,C dimension than the A–D dimension.

Similar to Experiment 1, unconstrained and constrained models assuming city-block and Euclidian distance between the estimated coordinates were fitted for each participant. A summary of the two sets of scaling solutions is provided in **Table 2**. For the constrained scaling solutions, models that assumed a Euclidean distance metric provided better fits of the scaling solution. A similar pattern of results was observed for the unconstrained scaling solutions. The only exception was that best fitting unconstrained

solutions for subjects U5, U6, and I5 assumed city-block distance metric. Taken across all observers, the pattern suggests that these face morphs are consistent with integrality in that most observer's scaling solutions are better fit by assuming a Euclidean metric. The unconstrained model fit better but was typically less preferred based on BIC due to the larger number of parameters. Hence, based on the MDS modeling along we would conclude that for seven of our observers, there was no violation of perceptual separability. Nevertheless, we continued to utilize the unconstrained solution when fitting the different architectures to capture the assumption of MSI. As before, we also fit each of the models assuming either PS or MSI and assuming either DS or optimal category boundaries.

COMPUTATIONAL MODELING

The model fits for each subject in the upright and inverted conditions are shown in **Table 5** and the parameters of the best fitting model are shown in **Table 3**.

Upright condition

Inspection of **Table 5** reveals that the parallel self-terminating model was the best fitting model for three participants in the upright condition. For the set of models assuming MSI and DS, the parallel model was the best model for U7 and U8, but the coactive and serial models were the best models for U5 and U6 respectively. The parallel self-terminating model was the best fitting model for U6–U8, when assuming both MSI and OP; the coactive model was the best model for U5. For the models assuming both PS and DS, the coactive model was the best model for U5, U7, and U8, but the parallel model was the best model for U6.

Individually, participant U5 demonstrated coactive processing under all three different assumptions of perceptual representation, but the model that assumes PS and DS was the overall best fitting model (BIC = 615.70). The parallel self-terminating model best fitted U6 (BIC = 543.83) with the same assumptions of perceptual representation. The parallel self-terminating model best fitted U7 (BIC = 525.37) and U8 (BIC = 713.84) under the assumption of MSI and DS. The predictions of the best fitting models are plotted against individual RT distributions in **Figure 9**.

Inverted condition

The model fits of the inverted condition present a clear picture. The parallel self-terminating model best fitted the data for participants I5, I7, and I8 under all three different assumptions of perceptual representations. Participant I5 (BIC = 644.89) was best fitted with the assumption of MSI and OP, but participants I7 (BIC = 559.17) and I8 (BIC = 635.06) were best fitted with the assumption of PS and DS. For participant I6, the coactive model with the assumption of MSI and DS was the overall best fitting model (BIC = 616.13). The predictions of the best fitting parameters are plotted against individual RT distributions in **Figure 10**.

DISCUSSION

In sum, parallel self-terminating processing was observed for three of the four participants in both the upright and inverted

FIGURE 8 | (A,B) shows the individual multidimensional scaling solutions for the Upright and Inverted conditions in Experiment 2, respectively.
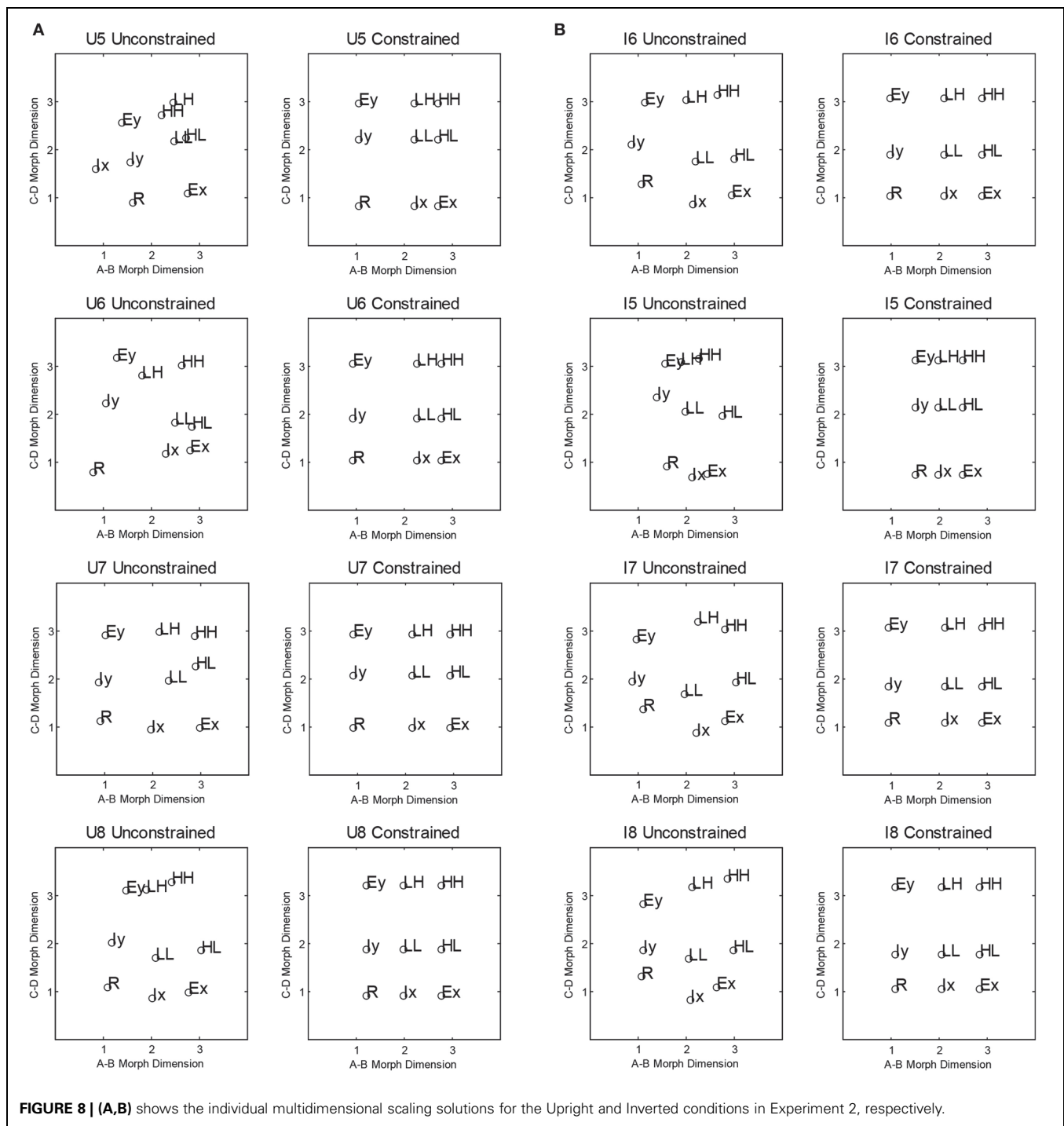
conditions of Experiment 2. This is in contrast to Experiment 1 in which a majority of participants demonstrated coactive processing of upright and inverted face morphs dimensions. Taken together with Experiment 1, and given the small number of observers, our conclusion is that there are individual differences in the manner in which the face morph dimensions are processed. Regardless of whether the morphs are presented in an upright or inverted fashion, processing may be coactive or parallel depending on the individual observer. Similar to Experiment

1, Experiment 2 showed that changing the assumption of the underlying perceptual representations affects the best fitting model.

GENERAL DISCUSSION

In this paper, we examined processing of purportedly integral, arbitrary morph dimensions, comparing both upright and inverted face morphs. Our primary finding was that some individuals process the dimensions in a parallel self-terminating

Table 5 | Model fits to subjects in Experiment 2 (model with the lowest BIC in each set is bolded; best overall model is bolded an italics).

Subject	Coactive		Parallel exhaustive		Parallel self-terminating		Serial exhaustive		Serial self-terminating	
	-lnL	BIC	-lnL	BIC	-lnL	BIC	-lnL	BIC	-lnL	BIC
PS AND DS										
U5	*289.90*	*615.70*	492.62	1021.14	361.32	758.54	517.10	1070.09	409.43	858.75
U6	272.71	581.32	376.66	789.23	*253.97*	*543.83*	454.94	945.78	354.61	749.11
U7	**264.74**	**565.39**	495.81	1027.51	282.97	601.85	525.06	1086.02	402.50	844.89
U8	**366.41**	**768.72**	502.23	1040.37	368.20	772.31	491.54	1018.98	458.31	956.52
I5	472.66	981.22	657.95	1351.79	**444.75**	**925.40**	744.24	1524.38	492.54	1024.97
I6	**290.34**	**616.58**	737.07	1510.05	365.19	766.28	716.51	1468.91	533.73	1107.35
I7	412.70	861.30	518.87	1073.64	*261.63*	*559.17*	514.43	1064.76	376.34	792.58
I8	372.39	780.67	368.31	772.52	*299.58*	*635.06*	410.63	857.17	353.91	747.72
MSI AND DS										
U5	**367.95**	**771.80**	552.57	1141.04	455.15	946.19	591.10	1210.13	447.10	934.10
U6	408.12	852.14	361.11	758.12	364.58	765.06	433.45	894.82	**340.58**	**721.05**
U7	294.90	625.71	531.11	1098.12	*244.73*	*525.37*	519.29	1066.50	286.60	613.10
U8	515.94	1067.78	538.03	1111.97	*338.97*	*713.84*	508.53	1044.99	406.77	853.43
I5	439.36	914.62	553.70	1143.29	**320.56**	**677.02**	602.31	1232.55	328.97	697.84
I6	*290.12*	*616.13*	737.84	1511.59	350.37	736.65	693.59	1415.10	468.98	977.84
I7	568.12	1172.14	610.76	1257.43	**379.92**	**795.74**	650.95	1329.82	474.17	988.22
I8	446.98	929.87	393.25	822.40	**351.66**	**739.22**	467.70	963.32	366.24	772.36
MSI AND OP										
U5	**409.46**	**854.82**	731.59	1499.09	1055.75	2147.40	807.25	1650.41	482.32	1004.54
U6	496.00	1027.90	490.14	1016.19	**477.51**	**990.92**	570.76	1177.42	870.17	1780.23
U7	395.85	827.60	488.78	1013.45	**309.05**	**654.01**	550.07	1136.03	1000.76	2041.40
U8	507.92	1051.75	590.52	1216.93	**356.71**	**749.31**	568.15	1172.21	1288.88	2617.65
I5	455.77	947.44	512.42	1060.74	*304.50*	*644.89*	596.91	1229.73	1202.71	2445.32
I6	720.69	1477.27	860.90	1757.70	**662.16**	**1360.23**	865.15	1766.20	1588.40	3216.69
I7	860.90	1757.70	803.03	1641.97	**801.66**	**1639.22**	837.51	1710.93	1108.51	2256.90
I8	833.26	1702.42	784.05	1603.99	**759.12**	**1554.14**	827.49	1690.87	1385.56	2811.00

fashion and others process the dimensions coactively for both upright and inverted face morphs.

A strength of the present study is the comparison of the model fits under different assumptions of the underlying perceptual representation. The scaling solutions from both experiments reveal deviations from the 3×3 grid-layout outline in **Figure 1**. Experiment 1 showed that the preferred model varied based on the underlying representational assumption. For example, the coactive model was the best fitting model in the upright condition for all participants when perceptual integrality and decisional separability were assumed; however, once the model assumed either optimal responding or mean shift integrality, the parallel model was superior in terms of BIC. A clear benefit of the parametric approach taken here that we are able to tease apart differences in representation from differences in architecture.
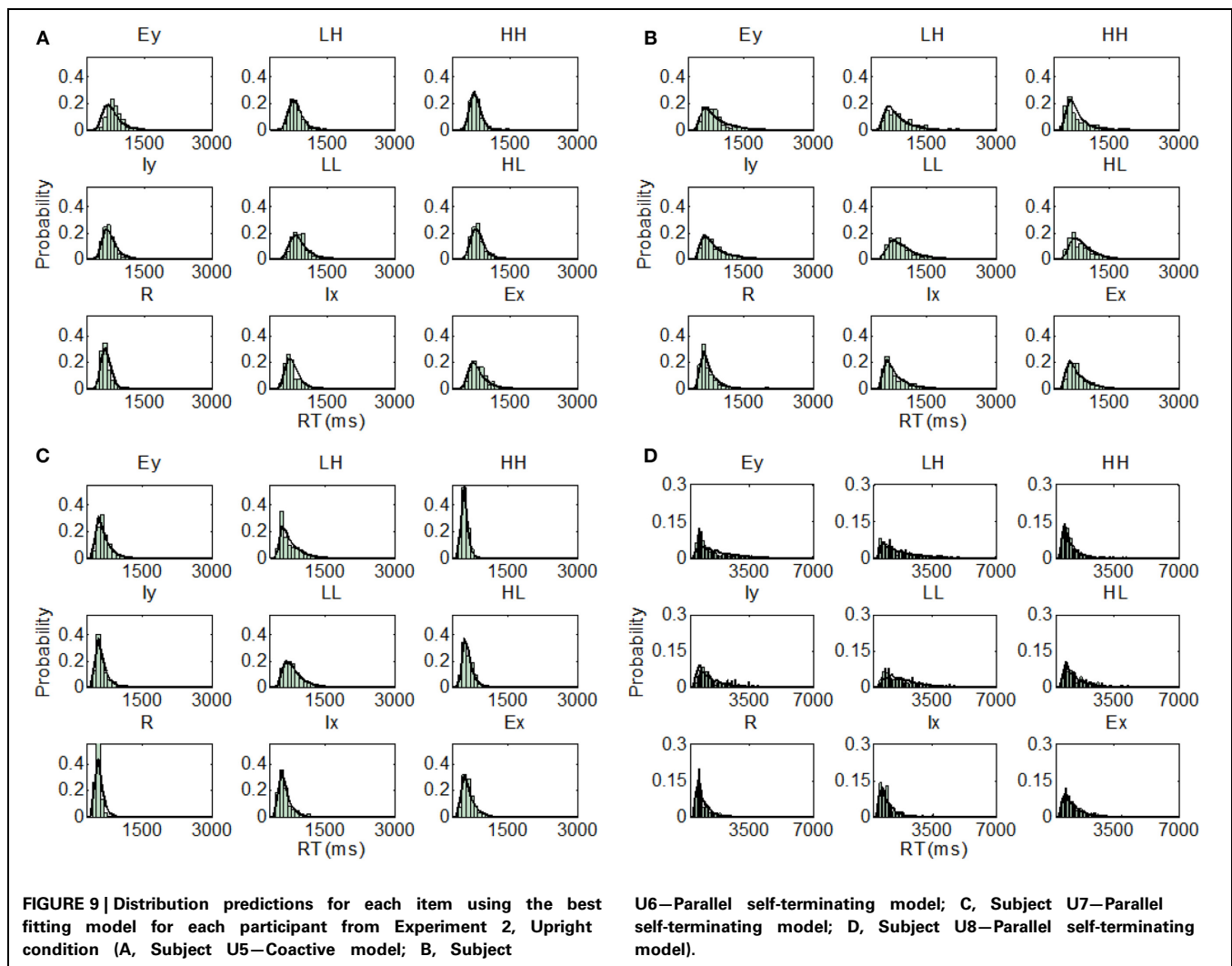
Overall, more participants used a coactive strategy in Experiment 1 compared to Experiment 2. There are two possible reasons for this difference. Firstly, participants may have perceived the face morphs differently since the visual angle and the face morph dimensions were altered between experiments (i.e., the position of two base faces were swapped). Secondly, model fitting for Experiment 1 utilized the averaged scaling solution of independent participants, but model fitting for Experiment 2 utilized individual scaling solutions after categorization training. In general, there is high variability in the perceptual representation of these face morphs between individuals and thus the average scaling solution may not have adequately represented the perceptual representation of each participant in Experiment 1.

IMPLICATIONS FOR PREVIOUS RESEARCH

The finding of individual differences in processing face morph stimuli implies that previous studies employing these stimuli on the assumption that they are processed in an integral fashion need to be interpreted with caution. On the one hand, the stimuli clearly satisfy one of the empirical operational definitions of integrality in that for most observers, the best fitting scaling metric was Euclidean. On the other hand, only half of the observers required assuming a violation of perceptual separability. Furthermore, only half of the observers were best fit by a coactive processing architecture, and of those, only two observers from Experiment 2, where individual scaling solutions were used, were found to be coactive. Consequently, the evidence that the face morph stimuli provide consistent and converging evidence of coactive processing is rather weak.

In their study of perceptual differentiation, Goldstone and Steyvers (2001) found that the face morph dimensions were

FIGURE 9 | Distribution predictions for each item using the best fitting model for each participant from Experiment 2, Upright condition (A, Subject U5—Coactive model; B, Subject U6—Parallel self-terminating model; C, Subject U7—Parallel self-terminating model; D, Subject U8—Parallel self-terminating model).
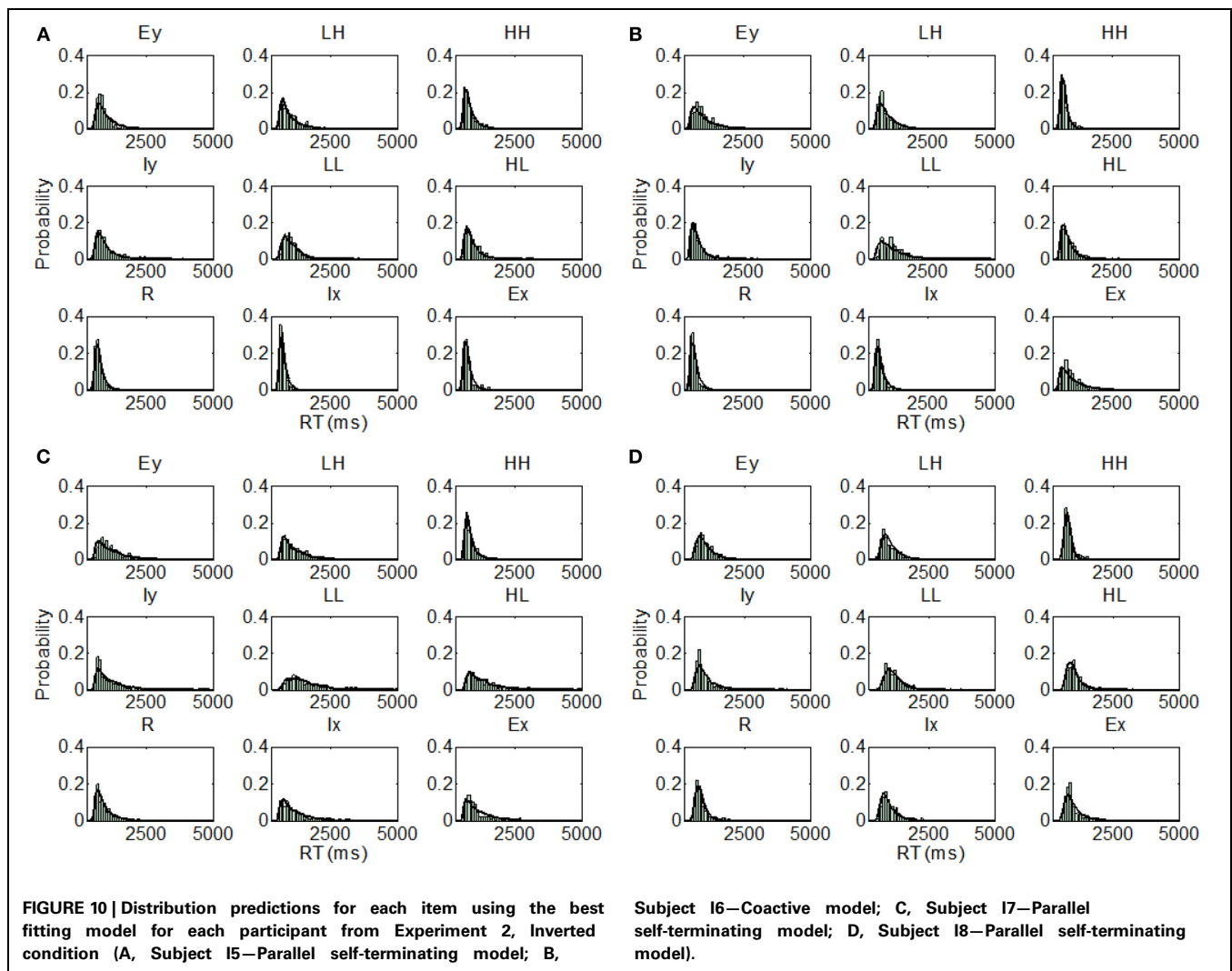
independently analyzable after training on a boundary orthogonal to the stimulus dimensions. Goldstone and Steyvers acknowledge the possibility that because of the grid-like arrangement of the stimuli, participants may have realized that there was a consistent dimensional structure. Indeed, in their Experiment 3, they utilized a stimulus space which did not have a grid-like structure (i.e., the face morphs were arranged in a circle), yet they still found evidence for differentiation. Consequently, it would seem prudent to limit our conclusions of individual differences to the case in which the face morphs are aligned to a grid making potentially making the dimensional structure particularly identifiable.

An alternative interpretation of our result would be to assume that differentiation is not precluded by training a category boundary on both stimulus dimensions, and that our observation that some observers processed the dimensions independent (in a parallel, self-terminating fashion) is evidence of that differentiation. In support of this idea, the MDS solution from Experiment 1, which was the only data collected prior to category learning (concerns about averaging notwithstanding; Ashby et al., 1994), is best fit by a Euclidean distance metric suggesting integrality. However,

we note that a Euclidean metric was also found for most of our observers in Experiment 2 *after* extensive category learning. It is clear from the present results that individuals differ with regard to how they represent and process the face morphs used in the present study. Whether this results from a difference in the time course of differentiation and learning (i.e., across sessions) is left for future research. Nonetheless, we note that the MDS solutions found in Experiment 2 were found using data collected after extensive category learning. These solutions all indicate that a constrained solution (i.e., which exhibits perceptual separability as defined by GRT) provides a better account of the similarity data. This result is in line with the hypothesis that the stimulus dimensions were differentiated after category learning.

Finally, a further caveat on the implications of the present research is that we tested a relatively small number of individuals. This is a consequence of the experimental design which necessitates collecting large numbers of observations from each observer. Nevertheless, we can clearly rule out a large number of models including all serial models and all exhaustive models. This leaves coactivity and parallel self-termination as the remaining candidate processing models for the present face morph stimuli. That

FIGURE 10 | Distribution predictions for each item using the best fitting model for each participant from Experiment 2, Inverted condition (A, Subject I5—Parallel self-terminating model; B, Subject I6—Coactive model; C, Subject I7—Parallel self-terminating model; D, Subject I8—Parallel self-terminating model).

we found, essentially, the same sorts of individual differences in both experiments suggests that the individual differences are real and not due to small idiosyncratic differences between subjects.

IMPLICATIONS OF THEORETICAL NOTIONS OF INTEGRALITY

Here we have shown that stimuli which were previously thought to be integral on the basis of one empirical test of integrality, do not necessarily meet all other tests of integrality (cf. Cheng and Pachella, 1984). The face morph dimensions used in this experiment had been previously shown to result in an interference effect when variation on an irrelevant dimension was introduced suggesting integrality. In the current study, the scaling solutions demonstrated clear violations of perceptual separability (Ashby and Townsend, 1986; Ashby and Maddox, 1991; Maddox, 1992; Maddox and Ashby, 1996) and the Euclidean metric was preferred for most observers, but for observers in Experiment 2 a constrained solution was preferred after taking the complexity of the solution into account. Taken in conjunction with the RT data, however, there was a good deal of variation in whether perceptual separability was violated or not. Little et al.'s (2013) experiments using Munsell color stimuli suggest a theoretical definition of

integrality in terms of coactive processing. For the present stimuli, however, we also did not find consistently coactive processing suggesting that the face morphs used here do have some identifiable structure which can be processed in an independent fashion.

Yet, one may question why additional theoretical definitions of integrality are necessary. GRT offers a theoretical definition of perceptual representation, which rigorously defines violations of perceptual independence, perceptual separability and decisional separability, so is there any need to posit coactivity as a theoretical representation for integrality? As a background consideration, it is worthwhile to note that GRT does not predict RTs without additional mechanisms, and aside from the logical rule models presented here, only the distance-from-boundary hypothesis has been applied to explain some of the empirically observable definitions of integrality (Ashby and Maddox, 1991; Maddox, 1992). Though, as previously discussed, the distance-from-boundary hypothesis makes untenable predictions for the speed of the fastest RTs when perceptual variability is increased (Nosofsky and Palmeri, 1997). Consequently, we feel that GRT provides a representational-level theory of integrality, but does

not extend adequately to understanding how integral dimensions are processed. Though we highlight recent advances in developing a non-parametric dynamic GRT, which extends the concepts defined within GRT to a class of parallel processing models (Townsend et al., 2012); these models have not yet been applied to differentiating separable and integral dimensioned stimuli. By contrast, the logical rule models approach are a process-level theory of integrality but one which offers a way to simultaneously consider both the perceptual representation and the underlying processing architecture.

There are two somewhat orthogonal ideas that might be considered when addressing the question of whether aligning integrality with coactivity is necessary. The first is that defining integrality as coactivity might confound integrality at the perceptual and decisional stages. For instance, one could imagine that perceptual separable dimensions might be pooled together at a decisional stage. While this is a conceptually possible, we do not consider this to be very plausible in the present case. This hypothesis would capture ideas present in many two-stage, salience-based models of visual search (Neisser, 1967; Wolfe et al., 1989; Wolfe, 1994; Found and Muller, 1996) that an initially independent parallel stage selects out information for later processing by an apparently coactive system. In the present case, however, the stimuli are presented until a response is generated; consequently, the early system is likely completely saturated. In this case, the GRT representations likely do not capture the early, salience-based perceptual qualities of the stimulus dimensions, but rather capture something like the relative similarity between each of the stimuli (Ashby and Perrin, 1988). Under extended display conditions, representations of the dimensions that are independent and driven by the marginal representation of the dimensions are not likely to exhibit patterns of effects which are the empirical hallmarks of integrality. The present approach allows one to test these assumptions parametrically by varying both the representation and the architecture thereby separating perceptual and decisional separability from the architecture used to generate the RTs.

A second issue arising from consideration of the mechanisms used to generate the RTs is that to the extent that integrality is aligned with the notion of holism and to what extent coactivity captures what is typically meant by that latter concept. For instance, in a task similar to the task used here, Fifić and Townsend (2010) examined the processing of secondary holistic features (e.g., the distance between the eyes or the between the lips and the nose) which are thought to be part of the underlying configural advantage underlying face perception. In that study, under conditions conducive to holistic processing, observers were found to demonstrate coactivation. Strong definitions of holistic processing seem commensurate with the theoretical notions implied by coactivity; the same is true when ideas of holistic processing are applied to dimensional integrality.

Fifić and Townsend's (2010) finding of coactivation using faces with secondary-level facial feature differences stands in contrast to the relative lack of consistent coactivation in the present experiments using face morphs. One possibility is that, like Fifić and Townsend's study, coactivation would develop over time with repeated presentation of the stimuli as the individual morph dimensions are unitized into a holistic representation. Although

this is possible, it is the opposite of the direction of perceptual learning assumed by Goldstone and Steyver's (2001) in which the face morph dimensions became more separable with training. A key difference in that study was that the training only utilized discriminations along a single dimension, whereas here, both dimensions are relevant. Nonetheless, we find mixed evidence of coactivation when both dimensions are relevant. This also did not vary based on whether faces were presented in an upright or inverted fashion. We tentatively suggest that the morph dimensions we use here do not contain the sort of individual identification information that seems to drive superior face identification performance but instead contain dimensional structure which can be utilized by some observers. This clearly renders overarching inferences based on averaged data problematic. We argue that without factorial manipulations to tease apart how dimensional information is integrated for each observer, general conclusions may be misleading.

Finally, although the logical rules framework that we adopt here combines many existing approaches to studying integrality and separability, it is worth considering whether some deeper theoretical insight can be used to understand the variety of converging operations. Three converging operations are worth considering: the MDS metric (Attneave, 1950; Torgenson, 1958; Shepard, 1964, 1987; Nosofsky, 1992), the efficiency of selective attention (Nosofsky, 1987), and Garner's (1974) facilitation and interference results. When coupled with our modeling results, the finding that a Euclidean metric persists after extensive category learning suggests the distance metric is an unreliable indicator of integrality (Grau and Kemler-Nelson, 1988). This suggests that a target for future research is to determine how different processing architectures predict the types of proximity measures which are used to derive the scaling solutions. For instance, one question of interest is whether serial and parallel processing models, which compare dimensions independently necessarily always lead to solutions with city-block distance metrics. A second question is whether coactivity always leads to solutions with Euclidean distance metrics. At present, these relations are intuitive, but the strength of this relationship is unclear.

With regard to the efficiency of selective attention, in the logical rules models, there are at least two possible ways by which selective attention might influence processing. One mechanism is to increase the processing rate of attended dimensions and decrease the rate of less attended dimensions (see for example, Nosofsky and Palmeri, 1997; Ashby and Perrin, 1988). A second possibility is that selective attention might be linked to selective, fixed-order serial processing. That is, dimensions which are learned to be relevant for categorization or are more salient might be selected to be processed before (or to the exclusion of, in a self-terminating model) less relevant or less salient dimensions. In support of this idea, Lamberts (1995; 1998; 2000; see also Cohen and Nosofsky, 2003) showed that for separable-dimensioned stimuli, attributes vary in their temporal order within the decision process with more salient dimensions processed before other dimensions. These results are consistent with the logical rules account of serial processing of separable dimensioned stimuli (see Fifić et al., 2010), and by contrast, support the idea that integral dimensions should be process coactively (Little et al., 2013).

As noted in the introduction, Garner's (1974; Garner and Felfoldy, 1970) tasks do not allow one to differentiate between different processing architectures. The reason for this is that these tasks involve categorization using a single relevant dimension. Under these conditions, there is no difference in the processing rate predicted using the joint bivariate distributional representation and the marginal distribution representation. Likewise, there is only one processing channel (i.e., the relevant dimension). Hence, separable dimensions, which show no facilitation (i.e., with correlated variation) or interference (i.e., with irrelevant variation) might be processed either in a serial, a parallel, or a coactive fashion. On the other hand, the signature integral result of facilitation and interference could indicate either coactivity or some form of parallel processing. To explain, we consider coactivity to be likely for integral dimensions, but a change in architecture alone cannot predict Garner's results for integral dimensions. As discussed in Little et al. (2013, p. 817), other representational mechanisms would need to vary to predict facilitation and interference. For instance, one might expect optimal responding (i.e., a diagonal decision bound; Maddox, 1992) with correlated variation or an increase in perceptual variability (Maddox, 1992; Nosofsky and Palmeri, 1997) with irrelevant variation. However, the latter interference result could also be predicted via other mechanism; for instance, parallel processing with increased response caution could cause a slowing of RTs with irrelevant dimensional variation. We offer the method we employ in the present paper, which combines both processing and representational assumptions, as a framework for addressing these complex issues.

ACKNOWLEDGMENTS

REFERENCES

Ashby, F. G. (2000). A stochastic version of general recognition theory. *J. Math. Psychol.* 44, 310–329. doi: 10.1006/jmps.1998.1249

Ashby, F. G., and Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 33–53. doi: 10.1037/0278-7393.14.1.33

Ashby, F. G., and Maddox, W. T. (1991). "A response time theory of perceptual independence," in *Mathematical Psychology: Current Development*, eds J. C. Falmagne and J. P. Doignon (New York, NY: Springer-Verlag), 389–413.

Ashby, F. G., and Maddox, W. T. (1994). A response time theory of separability and integrality in speeded classification. *J. Math. Psychol.* 38, 423–466. doi: 10.1006/jmps.1994.1032

Ashby, F. G., Maddox, W. T., and Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychol. Sci.* 5, 144–152. doi: 10.1111/j.1467-9280.1994.tb00651.x

Ashby, F. G., and Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychol. Rev.* 95, 124–150. doi: 10.1037/0033-295X.95.1.124

Ashby, F. G., and Townsend, J. T. (1986). Varieties of perceptual independence. *Psychol. Rev.* 93, 154–179. doi: 10.1037/0033-295X.93.2.154

Attneave, F. (1950). Dimensions of similarity. *Am. J. Psychol.* 63, 546–554. doi: 10.2307/1418869

Biederman, I., and Checkosky, S. F. (1970). Processing redundant information. *J. Exp. Psychol.* 83, 486–490. doi: 10.1037/h0028841

Blaha, L. M., Busey, T. A., and Townsend, J. T. (2009). "An LDA approach to the neural correlates of configural learning," in *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, eds N. A. Taatgetn and H. van Rijn (Austin, TX: Cognitive Science Society), 2540–2545.

Borg, I., and Groenen, P. J. F. (2005). *Modern Multidimensional Scaling*. New York, NY: Springer Science+Business Media, Inc.

Burns, B., and Shepp, B. E. (1988). Dimensional interactions and the structure of psychological space: the representation of hue, saturation, and brightness. *Percept. Psychophys.* 43, 494–507. doi: 10.3758/BF03207885

Busemeyer, J. R. (1985). Decision making under uncertainty: a comparison of simple scalability, fixed-sample, and sequential-sampling models. *J. Exp. Psychol. Learn. Mem. Cogn.* 11, 538–564. doi: 10.1037/0278-7393.11.3.538

Carroll, J. D., and Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika* 35, 283–319. doi: 10.1007/BF02310791

Cheng, P. W., and Pachella, R. G. (1984). A psychophysical approach to dimensional separability. *Cogn. Psychol.* 16, 279–304. doi: 10.1016/0010-0285(84)90011-2

Cohen, A. L., and Nosofsky, R. M. (2003). An extension of the exemplar based random-walk model to separable-dimension stimuli. *J. Math. Psychol.* 47, 150–165. doi: 10.1016/S0022-2496(02)00031-7

Dzhafarov, E. N. (2003). Selective influence through conditional independence. *Psychometrika* 68, 7–25. doi: 10.1007/BF02296650

Dzhafarov, E. N., and Gluhovsky, I. (2006). Notes on selective influence, probabilistic causality, and probabilistic dimensionality. *J. Math. Psychol.* 50, 390–401. doi: 10.1016/j.jmp.2006.03.003

Dzhafarov, E. N., Schweickert, R., and Sung, K. (2004). Mental architectures with selectively influenced by subadditive effects on visual quality. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 551–560. doi: 10.1016/j.jmp.2003.12.003

Egeth, H. (1967). Selective attention. *Psychol. Bull.* 67, 41–57. doi: 10.1037/h0024088

Fifić, M., Nosofsky, R. M., and Townsend, J. T. (2008). Information processing architectures in multidimensional classification: a validation test of the systems-factorial technology. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 356–375. doi: 10.1037/0096-1523.34.2.356

Fifić, M., Little, D. R., and Nosofsky, R. M. (2010). Logical-rule models of classification response times: a synthesis of mental-architecture, random walk, and decision-bound approaches. *Psychol. Rev.* 117, 309–348. doi: 10.1037/a0018526

Fifić, M., and Townsend, J. T. (2010). Information-processing alternatives to holistic perception: identifying the mechanisms of secondary-level holism within a categorization paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 1290–1313. doi: 10.1037/a0020123

Foard, C. F., and Kemler, D. G. (1984). Holistic and analytic model of processing: the multiple determinants of perceptual analysis. *J. Exp. Psychol. Gen.* 113, 94–111. doi: 10.1037/0096-3445.113.1.94

Folstein, J. R., Gauthier, I., and Palmeri, T. J. (2012). How category learning affects object representations: not all morph spaces stretch alike. *J. Exp. Psychol. Learn. Mem.* 38, 807–820. doi: 10.1037/a0025836

Folstein, J. R., Palmeri, T. J., and Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cereb. Cortex* 23, 814–823. doi: 10.1093/cercor/bhs067

Found, A., and Muller, H. J. (1996). Search for unknown feature targets on more than one dimension: investigating a "dimension-weighting" account. *Percept. Psychophys.* 58, 88–101. doi: 10.3758/BF03205479

Garner, W. R. (1969). Speed of discrimination with redundant stimulus attributes. *Percept. Psychophys.* 6, 221–224. doi: 10.3758/BF03207021

Garner, W. R. (1974). *The Processing of Information and Structure.* Potomac, MD: Lawrence Erlbaum.

Garner, W. R. (1978). Selective attention to attributes and to stimuli. *J. Exp. Psychol. Gen.* 107, 287–308. doi: 10.1037/0096-3445.107.3.287

Garner, W. R., and Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cogn. Psychol.* 1, 225–241. doi: 10.1016/0010-0285(70)90016-2

Goldstone, R. L. (1998). Perceptual learning. *Annu. Rev. Psychol.* 49, 585–612. doi: 10.1146/annurev.psych.49.1.585

Goldstone, R. L. (2000). Unitization during category learning. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 86–112. doi: 10.1037/0096-1523.26.1.86

Goldstone, R. L., Gerganov, A., Landy, D., and Roberts, M. E. (2008). "Learning to see and conceive," in *The New Cognitive Sciences (Part of the Vienna Series in Theoretical Biology)*, eds L. Tommasi, M. Peterson, and L. Nadel (Cambridge, MA: MIT Press), 163–188.

Goldstone, R. L., Son, J. Y., and Byrge, L. (2011). Early perceptual learning. *Infancy* 16, 45–51. doi: 10.1111/j.1532-7078.2010.00054.x

Goldstone, R. L., and Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *J. Exp. Psychol. Gen.* 130, 116–139. doi: 10.1037/0096-3445.130.1.116

Goldstone, R. L., Steyvers, M., Spencer-Smith, J., and Kersten, A. (2000). "Interactions between perceptual and conceptual learning," in *Cognitive Dynamics: Conceptual Change in Humans and Machines*, eds E. Dlettnch and A. B. Markman (Mahwah, NJ: Lawrence Erlbaum and Associates), 191–228.

Gottwald, R. L., and Garner, W. R. (1975). Effects of focusing strategy on speeded classification with grouping, filtering, and condensation tasks. *Percept. Psychophys.* 11, 179–182. doi: 10.3758/BF03210371

Grau, J. W., and Kemler-Nelson, D. G. (1988). The distinction between integral and separable dimensions: evidence for the integrality of pitch and loudness. *J. Exp. Psychol. Gen.* 117, 347–370. doi: 10.1037/0096-3445.117.4.347

Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics.* New York, NY: Wiley.

Gureckis, T. M., and Goldstone, R. L. (2008). "The effect of the internal structure of categories on perception" in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, eds B. C. Love, K. McRae, and V. M. Sloutsky (Austin, TX: Cognitive Science Society), 843.

Handel, S., and Imai, S. (1972). The free classification of analyzable and unanalyzable stimuli. *Percep. Psychophys.* 12, 108–116. doi: 10.3758/BF032 12854

Heathcote, A., Brown, S., and Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychon. Bull. Rev.* 9, 394–401. doi: 10.3758/BF03196299

Heathcote, A., Brown, S., Wagenmakers, E. J., and Eidels, A. (2010). Distribution-free tests of stochastic dominance for small samples. *J. Math. Psychol.* 54, 454–463. doi: 10.1016/j.jmp.2010.06.005

Hendrickson, A. T., Kachergis, G., Gureckis, T. M., and Goldstone, R. L. (2010). "Is categorial perception really verbally-mediated perception?" in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, eds S. Ohlsson and R. Catrambone (Austin, TX: Cognitive Science Society), x–x.

Hockema, S. A., Blair, M. R., and Goldstone, R. L. (2005). "Differentiation for novel dimensions," in *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (Hillsdale, NJ: Lawrence Erlbaum Associates), 953–958.

Imai, S., and Garner, W. R. (1968). Structure in perceptual classification. *Psychn. Monogr. Suppl.* 2, 154–172.

Jones, M., and Goldstone, R. L. (2013). The structure of integral dimensions: contrasting topological and Cartesian respresenations. *J. Exp. Psychol. Hum. Perform.* 39, 111–132. doi: 10.1037/a0029059

Kantowitz, B. H. (1974). *Human Information Processing: Tutorials in Performance and Cognition*. Hillside, NJ: Erlbaum.

Kayser, A. (1984). *Heads.* New York, NY: Abbeville Press Inc.

Lamberts, K. (1995). Categorization under time pressure. *J. Exp. Psychol. Gen.* 124, 161–180. doi: 10.1037/0096-3445.124.2.161

Lamberts, K. (1998). The time course of categorization. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 695–711. doi: 10.1037/0278-7393.24.3.695

Lamberts, K. (2000). Information accumulation theory of categorization. *Psychol. Rev.* 107, 227–260. doi: 10.1037/0033-295X.107.2.227

Levy, R. M., and Haggbloom, S. J. (1971). Test of a multidimensional discrimination model of stimulus identification. *Psychon. Sci.* 25, 203–204. doi: 10.3758/BF03329092

Li, X., Liang, Z., Kleiner, M., and Lu, Z. L. (2010). RTbox: a device for highly accurate response time measurements. *Behav. Res. Methods* 42, 212–225. doi: 10.3758/BRM.42.1.212

Link, S. W. (1992). *The Wave Theory of Difference and Similarity.* Hillsdale, NJ: Erlbaum.

Little, D. R. (2012). Numerical predictions for serial, parallel and coactive logical rule-based models of response times. *Behav. Res. Methods* 44, 1148–1156. doi: 10.3758/s13428-012-0202-4

Little, D. R., Nosofsky, R. M., and Denton, S. E. (2011). Response-time tests of logical-rule models of categorisation. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1–27. doi: 10.1037/a0021330

Little, D. R., Nosofsky, R. M., Donkin, C., and Denton, S. E. (2013). Logical rules and the classification of integral-dimension stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 801–820. doi: 10.1037/a0029667

Lockhead, G. R. (1966). Effeccts of dimensional redundancy on visual discrimination. *J. Exp. Psychol.* 72, 95–104. doi: 10.1037/h0023319

Lockhead, G. R. (1972). Processing dimensional stimuli: a note. *Psychol. Rev.* 79, 410–419. doi: 10.1037/h0033129

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. New York, NY: Oxford University Press.

Maddox, W. T. (1992). "Perceptual and decisional separability," in *Multidimensional Models of Perception and Cognition*, ed F. G. Ashby (Hillsdale, NJ: Lawrence Erlbaum Associates), 147–180.

Maddox, W. T., and Ashby, F. G. (1996). Perceptual separability, decisional separability and the identification-speed classification relationship. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 795–817. doi: 10.1037/0096-1523.2 2.4.795

Melara, R. D., and Marks, L. E. (1990). Hard and soft interacting dimensions: differential effects of dual context on classification. *Percept. Psychophys.* 47, 307–325. doi: 10.3758/BF03210870

Neisser, U. (1967). *Cognitive Psychology.* New York, NY: Appleton-Century-Crofts.

Nosofsky, R., and Little, D. R. (2010). Classification response times in probabilistic rule-based category structures: contrasting exemplar-retrieval and decision-bound. *Mem. Cognit.* 38, 916–927. doi: 10.3758/MC.38.7.916

Nosofsky, R. M. (1986). Attention, similarity, and the identification categorization relationship. *J. Exp. Psychol. Gen.* 115, 39–57. doi: 10.1037/0096-3445.1 15.1.39

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 87–108. doi: 10.1037/0278-7393.13.1.87

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Ann. Rev. Psychol.* 43, 25–53. doi: 10.1146/annurev.ps.43.020192.000325

Nosofsky, R. M., Little, D. R., Donkin, C., and Fifić, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychol. Rev.* 118, 280–315. doi: 10.1037/a0022494

Nosofsky, R. M., Little, D. R., and James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 333–338. doi: 10.1073/pnas.1111304109

Nosofsky, R. M., and Palmeri, T. J. (1996). Learning to classify integral-dimension stimuli. *Psychon. Bull. Rev.* 3, 222–226. doi: 10.3758/BF03212422

Nosofsky, R. M., and Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychol. Rev.* 104, 266–300. doi: 10.1037/0033-295X.104.2.266

Op de Beeck, H., Wagemans, J., and Vogels, R. (2003). The effect of category learning on the representation of shape: dimensions can be biased by not differentiated. *J. Exp. Psychol. Gen.* 132, 491–511. doi: 10.1037/0096-3445.13 2.4.491

Pomerantz, J. R., and Sager, L. C. (1975). Asymmetric integrality with dimensions of visual pattern. *Percept. Psychophys.* 18, 460–466. doi: 10.3758/BF03204121

Posner, M. L. (1964). Information reduction in the analysis of sequential tasks. *Psychol. Rev.* 71, 491–504. doi: 10.1037/h0041120

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108. doi: 10.1037/0033-295X.85.2.59

Ratcliff, R., and Rouder, J. N. (1998). Modeling response times for two choice decisions. *Psychol. Sci.* 9, 347–356. doi: 10.1111/1467-9280.00067

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136

Schweickert, R. (1992). "Information, time, and the structure of mental events: a twenty-five year review," in *Attention and Performance: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience—A Silver Jubilee,* Vol. 14, eds D. E. Meyer and S. Komblum (Cambridge, MA: MIT Press), 535–566.

Schweickert, R., Giorgini, M., and Dzhafarov, E. N. (2000). Selective influence and response time cumulative distribution functions in serial-parallel task networks. *J. Math. Psychol,* 44, 504–535. doi: 10.1006/jmps.1999.1268

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *J. Math. Psychol.* 1, 54–87. doi: 10.1016/0022-2496(64)90017-3

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science* 237, 1317–1323. doi: 10.1126/science.3629243

Shepard, R. N. (1991). "Integrality versus separability of stimulus dimensions: from an early convergence of evidence to a proposed theoretical basis," in *The Perception of Structure*, eds G. R. Lockhead and J. R. Pomerantz (Washington, DC: APA), 53–71.

Shepard, R. N., and Chang, J.-J. (1963). Stimulus generalization in the learning of classifications. *J. Exp. Psychol.* 65, 94–102. doi: 10.1037/h0043732

Shepard, R. N., Hovland, C. I., and Jenkins, H. M. (1963). Learning and memorization of classifications. *Psychol. Monogr. Gen. Appl.* 75, 517.

Shepp, B. E. (1989). "On perceiving objects: holistic versus featural properties," in *Object Perception: Structure and Process*, eds B. E. Shepp and S. Ballesteros (Hillsdale, NJ: Lawrence Erlbaum Associations, Inc), 203–234.

Smith, L. B. (1989). "From global similarity to kinds of similarity: the construction of dimensions in development," in S*imilarity and Analogical Reasoning*, eds S. Vosniadou and A. Ortony (Cambridge: Cambridge University Press), 146–178.

Smith, L. B., and Kemler, D. G. (1978). Levels of experienced dimensionality in children and adults. *Cogn. Psychol.* 10, 502–532. doi: 10.1016/0010-0285(78)90009-9

Sternberg, S. (1969). Memory scanning: mental processes revealed by reaction-time experiments. *Am. Sci.* 4, 421–457.

Steyvers, M. (1999). Morphing techniques for generating and manipulating face images. *Behav. Res. Methods Instrum. Comput.* 31, 359–369. doi: 10.3758/BF03207733

Torgenson, W. S. (1958). *Theory and Methods of Scaling.* New York, NY: John Wiley and Sons Inc.

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *J. Mathmat. Psychol.* 28, 363–400. doi: 10.1016/0022-2496(84)90007-5

Townsend, J. T., and Ashby, F. G. (1983). *The Stochastic Modeling of Elementary Mental Organization.* Cambridge, England: Cambridge University Press.

Townsend, J. T., Houpt, J. W., and Silbert, N. H. (2012). General recognition theory extended to include response times: predictions for a class of parallel systems. *J. Math. Psychol.* 56, 476–494. doi: 10.1016/j.jmp.2012.09.001

Townsend, J. T., and Nozawa, G. (1995). Spatio-temporal properties of elemental perception: an investigation of parallel, serial and coac-tive theories. *J. Math. Psychol.* 39, 321–360. doi: 10.1006/jmps.1995.1033

Townsend, J. T., and Wenger, M. J. (2004). A theory of interactive parallel processing: new capacity measures and predictions for a response time inequality series. *Psychol. Rev.* 111, 1003–1035. doi: 10.1037/0033-295X.111.4.1003

Wolfe, J. M. (1994). Guided Search 2.0: a revised model of visual search. *Psychon. Bull. Rev.* 1, 202–238. doi: 10.3758/BF03200774

Wolfe, J. M., Cave, K. R., and Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 419–433. doi: 10.1037/0096-1523.15.3.419

Yin, R. K. (1969). Looking at upside-down faces. *J. Exp. Psychol.* 81, 141–145. doi: 10.1037/h0027474

Individual differences in working memory capacity and workload capacity

Ju-Chi Yu , Ting-Yun Chang and Cheng-Ta Yang*

Department of Psychology, National Cheng Kung University, Tainan, Taiwan

We investigated the relationship between working memory capacity (WMC) and workload capacity (WLC). Each participant performed an operation span (OSPAN) task to measure his/her WMC and three redundant-target detection tasks to measure his/her WLC. WLC was computed non-parametrically (Experiments 1 and 2) and parametrically (Experiment 2). Both levels of analyses showed that participants high in WMC had larger WLC than those low in WMC only when redundant information came from visual and auditory modalities, suggesting that high-WMC participants had superior processing capacity in dealing with redundant visual and auditory information. This difference was eliminated when multiple processes required processing for only a single working memory subsystem in a color-shape detection task and a double-dot detection task. These results highlighted the role of executive control in integrating and binding information from the two working memory subsystems for perceptual decision making.

Keywords: executive function, linear ballistic accumulator model, systems factorial technology, working memory capacity, workload capacity

INTRODUCTION

The present study aimed to investigate the relationship between two capacity measures: working memory capacity (WMC) in the literature of working memory (Baddeley and Hitch, 1974; Barrett et al., 2004) and workload capacity (WLC) in the literature of perceptual decision making (Townsend and Ashby, 1978; Townsend and Nozawa, 1995; Townsend and Eidels, 2011). Although both measures assess an individual's information processing capacity, it was unclear whether the two capacity measures assess a unitary, central capacity of an information processing system. We used a non-parametric approach (systems factorial technology, SFT) (Townsend and Nozawa, 1995) and a parametric approach (linear ballistic accumulator model, LBA) (Brown and Heathcote, 2008; Eidels et al., 2010) to assess WLC in different task contexts and examined individual differences in WLC and WMC. We will briefly introduce the concepts of the two capacity measures.

Working memory refers to aspects of on-line cognition, such as monitoring, processing, and maintenance of information. A key component of Baddeley and Hitch's (1974) model of working memory, also known as the "short-term storage" of information (Henderson, 2013), is the central executive system, which is a modality-free function that supervises two slave systems of working memory: the phonological loop and the visuospatial sketchpad. The central executive system plays an important role in integrating information from the two subsystems for manipulation and operation. Following Baddeley and Hitch (1974), many theories regarding the construct of the central executive system have been proposed—for example, the *supervisory attention system* (SAS) in Norman and Shallice (1986) and the *executive control* in Posner and Digirolamo (2000). WMC is an index that denotes the capability of attention control in central executive of a working memory system and researchers typically use a

counting span task (Case et al., 1982), an operation span task (OSPAN task) (Turner and Engle, 1989) and a reading span task (Daneman and Carpenter, 1980) to measure one's WMC. Measures of WMC are strongly related to general fluid intelligence (Conway et al., 2003) and show considerable construct validity insofar as they predict performance on a wide range of tasks that require domain-general controlled attention. WMC is different from the traditional concept of short-term memory capacity, which is thought to reflect primarily domain-specific storage. One of the most widely supported theories, particularly when applied to individual differences in working memory, is the attention control theory of working memory (Engle and Kane, 2004). Individuals with high WMC have greater attention control in integrating information from different domain-specific subsystems (Rosen and Engle, 1997; Engle et al., 1999; Barrett et al., 2004; Engle and Kane, 2004). These results have been supported by computational modeling research (Anderson, 2013) and neurobiological research (Miller and Cohen, 2001).

At approximately the same time, another capacity measure, WLC was developed (Townsend and Ashby, 1978; Wenger and Gibson, 2004; Townsend and Eidels, 2011). WLC is also known as perceptual capacity. In contrast to WMC, which measures an individual's capacity to maintain and process information, WLC measures the efficiency of perceptual processing as workload (i.e., the number of channels or signals to be processed) increases. If the processing rate of an individual channel does not change as the workload increases, the system is described as unlimited-capacity processing. If the individual-channel processing speed slows down with an increasing workload, the system is described as limited-capacity processing, and if processing speeds up, the system is described as supercapacity processing. WLC is commonly measured with a redundant-target detection task

(Miller, 1982; Townsend and Nozawa, 1995) where participants are required to monitor two sources of information. Participants have to make a positive response when they detect the presence of both of the targets (redundant-target condition) or either target (single-target condition); otherwise, they have to make a negative response when they detect neither target (no-target condition). WLC can be assessed by comparing the reaction time distributions between the redundant-target and single-target conditions. For more theoretical derivations, please see Townsend and Nozawa (1995) and Wenger and Gibson (2004). Previous studies have widely applied the measure of WLC to study how people process multiple sources of information and how this measure is related to different aspects of human cognition. For example, in a double-dot detection task, participants were of limited-capacity in processing redundant spatially-independent visual information (Townsend and Nozawa, 1995; Eidels et al., 2010), which was against the prediction from the unlimited-capacity, independent, parallel (UCIP) model. In a redundant color-shape detection task, participants were of unlimited-capacity in processing separable perceptual dimensions when inter-stimulus contingency information was removed (Mordkoff and Yantis, 1991, 1993). In a visual search task, participants were of supercapcity in searching for a feature singleton defined by luminance and/or orientation (Zehetleitner et al., 2009). In a visual-auditory detection task, participants were of supercapacity in processing multisensory information (Miller, 1982), which was known as an effect of "multisensory integration" (Hugenschmidt et al., 2010; Altieri and Townsend, 2011).

In addition to WLC, there are two other important characteristics to describe information processing in a system, including the processing architecture (serial vs. parallel vs. coactive) that denotes the order of multiple-signal processing and the decisional stopping rule (self-terminating vs. exhaustive) that denotes the amount of information required for a decision. Although WLC and the processing architecture are two independent measures of information processing (Townsend and Nozawa, 1995), WLC may constrain the order of multiple-signal processing. For example, a standard serial model is assumed to involve limited-capacity processing (Townsend and Ashby, 1983); an independent parallel model usually involves unlimited-capacity processing, which is known as the UCIP (Houpt and Townsend, 2012); and a coactive model is assumed to involve supercapacity processing (Wenger and Townsend, 2001). On the other hand, a recent simulation study (Eidels et al., 2011) demonstrated that a parallel model with supercapacity processing suggests the existence of facilitatory between-channel crosstalk during the stage of information accumulation, whereas a parallel model with limited-capacity processing suggests an inhibitory interaction between channels.

Both WMC and WLC represent a system's capacity to process information, but they are different constructs in nature. The processing capacity in a working memory system describes the capacity of domain-general controlled attention to maintain and process information and, especially, integrate information from the two subsystems. In contrast, WLC represents a system's capacity of multiple-signal processing and is referred to as the variation of the processing efficiency of an individual channel as a function of workload. The relationship between WMC and WLC remains unclear, however, and to our knowledge, no prior studies have investigated the relationship between the two constructs, except for a recent study conducted by Heathcote et al. (2014).

The present study examined the relationship between WMC and WLC. To measure WMC, participants were asked to performed an OSPAN task, in which they had to remember a few words while solving an arithmetic equation at the same time (Turner and Engle, 1989). In addition, they performed three different redundant-target detection tasks to measure their WLC. Modalities that the participants had to supervise in three redundant-target detection tasks were well defined; redundant information may come from a single visual modality (two visual features, two distinct spatial positions) or two different modalities (i.e., visual and auditory modalities). The reasons why we chose these tasks were as follows: (1) These redundant-target detection tasks have been widely used to study multiple-signal processing in the previous literature (Miller, 1982; Mullin et al., 1988; Townsend and Nozawa, 1995; Eidels et al., 2010), but less is known about the individual variation of the perceptual processing capacity in different tasks. (2) Relating WLC and WMC in different task contexts enables us to examine whether it requires a unitary, central capacity of information processing to process multiple signals that come from the same or different modalities. If both WMC and WLC assess the central processing capacity, we expect WLC to be positively related to WMC, regardless of whether redundant information is from the same modality. These results can shed light on the nature of the working memory system and the role of executive control in processing multiple signals for perceptual decision making.

EXPERIMENT 1

In Experiment 1, an OSPAN task was conducted to measure the participants' WMC and three redundant-target detection tasks—i.e., a color-shape detection task, a double-dot detection task, and a visual-auditory detection task— were conducted to measure their WLC. We expect that participants high in WMC would have larger WLC in multiple-signal processing.

METHOD
Participants
Fifty-seven (29 males and 28 females) undergraduates with a mean age of 20.63 years ($SD = 2.72$) at National Cheng Kung University volunteered in this experiment. All the participants had normal or corrected-to-normal vision and hearing. They signed a written informed consent prior to the experiment and received NTD 120 per hour after they completed the experiment.

Equipment
All the stimuli were presented on a 19-inch CRT monitor (CTX) with a refresh rate of 85 Hz and a display resolution of 1024 × 768 pixels. The viewing distance was 60 cm. Auditory stimuli were presented via a Philips Shm6500 headphone. The experiment was programmed with E-prime 1.1 (Schneider et al., 2012).

Stimuli, design, and procedure
Each participant performed three redundant-target detection tasks to measure his/her WLC and an OSPAN task to measure the capacity of a dynamic working memory system that involved

both the storage and processing of information. Each task last for approximate 1 hour and four tasks were conducted on different days.
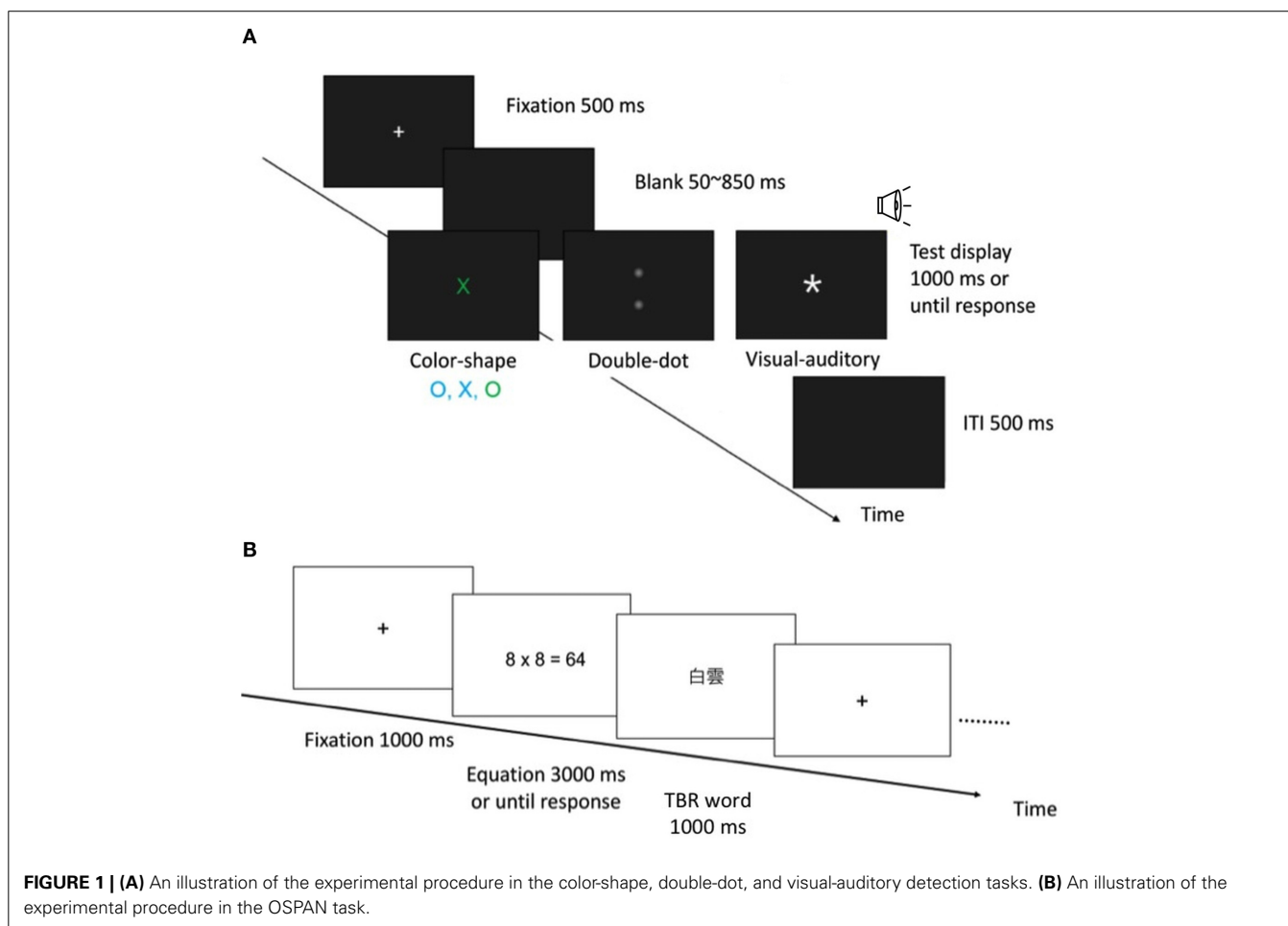
In the color-shape detection task, a test display consisted of a letter that was either an O or an X in shape, either green or cyan in color, and 1° (horizontal) × 1° (vertical) in size. The target color was defined as green, and the target shape was defined as X. In the redundant-target condition, the test stimuli consisted of both the target color and target shape (green X). In the single-target condition, the test stimuli consisted of either the target color or target shape (green O, cyan X). In the no-target condition, the test stimuli consisted of neither the target color nor target shape (cyan O). Each condition was equally probable and randomly intermixed within a block. After the participants practiced for 40 trials, they performed 12 blocks of 80 test trials.

Each trial began with a 500 ms fixation point (see **Figure 1A** for an illustration). Following a uniformly distributed random foreperiod ranging from 50 to 850 ms, a test stimulus was presented until participants responded or 1000 ms elapsed. Participants had to make a go/no-go response as quickly as possible when they detected either target feature (green or X). If *either* or *both* target features were detected, participants were required to press the "/" button (go response); if *neither* target feature was detected, they had to hold their response and wait for the next trial (no-go response). The inter-trial interval (ITI) was 500 ms.

In the double-dot detection task, the design and procedure were the same as those used in the color-shape detection task, except for the test stimuli. A 1° × 1° light dot (*luminance* = 0.031 cd/m^2) was presented 6° above and/or below the fixation point[1]. There were three types of test trials: redundant-target condition (both locations contained a light dot), single-target condition (either the top or bottom location contained a light dot), and no-target condition (neither location contained a light dot). Participants had to detect the presence of either or both dots as quickly as possible; otherwise, they had to hold their response and wait for the next trial (see **Figure 1A**).

In the visual-auditory detection task, the design and procedure were also the same as the former two tasks, except for the test stimuli, which consisted of a star sign (1° × 1°, *luminance* = 29.4 cd/m^2) and/or a 750-Hz pure tone (47.5 db). There were three types of test trials: redundant-target condition (both visual and

---

[1]One might argue that the distance between the two dots was too far, such that participants may adopt a serial processing strategy with reduced WLC. However, in Yang et al. (2014), we used a similar display setting and we found that participants adopted parallel processing when they did not have any prior information about the target location. In addition, in one of our unpublished studies, we used an eye tracker to record the participants' eye movements when they were detecting the double dots and no eye movements were found even when the distance between the two dots was 16 degree.

FIGURE 1 | (A) An illustration of the experimental procedure in the color-shape, double-dot, and visual-auditory detection tasks. **(B)** An illustration of the experimental procedure in the OSPAN task.

auditory signals were presented), single-target condition (either visual or auditory signal was presented), and no-target condition (neither visual nor auditory signal was presented). Participants had to detect the presence of either or both the visual and auditory targets as quickly as possible; otherwise, they had to hold their response (see **Figure 1A**).

In the OSPAN task, participants first saw an arithmetic equation, for example, $8 \times 8 = 64$, then they had to indicate whether the presented answer was correct, and finally saw a to-be-remembered (TBR) two-character Chinese word for later recall (see **Figure 1B** for an illustration). In each trial, there were two to six such processing-and-storage presentations. After the presentations, participants were required to write down the TBR words in correct serial order. There were a total of 15 trials that consisted of 5 presentation conditions (2/3/4/5/6) and three trials per condition. All the trials were randomly presented.

Data analysis

Reaction time data of the correct responses in the redundant-target detection tasks was analyzed to estimate WLC. According to SFT (Townsend and Nozawa, 1995; Townsend and Eidels, 2011), the capacity coefficient is expressed as follows:

$$C(t) = \frac{\log\left[S_{1,2}(t)\right]}{\log\left[S_1(t) \times S_2(t)\right]} \quad (1)$$

for $t > 0$, $S_1(t)$, $S_2(t)$, and $S_{1,2}(t)$ represent the survivor function, the complement of the cumulative probability function $[1-F(t)]$, of the two single-target conditions and a redundant-target condition, respectively. The capacity coefficient provides a comparison of the amount of work that is completed by the system while processing redundant targets and the summed amount of work that is completed by each single target processed individually at the same amount of time. A value of $C(t) = 1$ suggests unlimited-capacity processing: the processing efficiency of an individual channel is *not* affected by the change in workload. $C(t) > 1$ suggests supercapacity processing: increasing the to-be-processed signals speeds up the processing time of an individual channel. $C(t) < 1$ indicates limited-capacity processing: increasing the workload slows down the processing time of an individual channel.

To assess WMC for each participant, we first computed the recall score for each trial, which was defined as the number of TBR words fully recalled in correct serial order. WMC was computed by summing the recall scores of all the trials. The recall score ranges from 0 to 60.

Result

The number of correct answers on the processing component of the OSPAN task (i.e., solving the arithmetic equation) was analyzed. Four participants' data were excluded from further analysis because their processing accuracy was below 0.7. Under this criterion, the mean processing accuracy was 0.85 with a standard deviation of 0.06. We then computed the total number of items recalled from the storage component of the OSPAN task (i.e., recall score). The mean recall score was 36.38 with a standard deviation of 10.49.

We then conducted an extreme-group approach to investigate the relationship between WMC and WLC. This approach has been widely used to analyze continuous variables (Preacher et al., 2005). We selected the subject for further analysis on the basis of the extreme WMC scores (i.e., recall scores in the OSPAN task) to emphasize the differences in WLC between the high-WMC and low-WMC groups. The high-WMC group included the participants with the top 30% of recall scores ($M = 47.33$, $SD = 4.45$, $N = 18$), and the low-WMC group included the participants with the bottom 30% of recall scores ($M = 24.44$, $SD = 5.49$, $N = 18$). The recall scores of the two groups were significantly different [$t_{(34)} = 13.75$, $p < 0.0001$].

To do further analysis, we then excluded the trials with reaction time less than 150 ms in the redundant-target detection tasks. This criterion was selected because simple reaction times are generally slower than 150 ms. The mean performance of the redundant-target detection tasks for each group was summarized in **Table 1**. Accuracies were very high across conditions for both groups of participants except for the performance in the no-target condition of the color-shape detection task (0.89), suggesting a potential response bias in detecting color and/or shape. A Two-Way (high-WMC/low-WMC group × redundant-target/single-target condition) analysis of variance (ANOVA) was conducted to analyze the accuracy and correct reaction time data of the three tasks. We found that all the effects were not significant for the accuracy data of all the tasks. For reaction time data, there were significant main effects of group [CS^2: $F_{(1, 68)} = 9.00$, $p < 0.005$; DD: $F_{(1, 68)} = 7.27$, $p < 0.01$; VA: $F_{(1, 68)} = 13.31$, $p < 0.001$] and condition [CS: $F_{(1, 68)} = 33.50$, $p < 0.001$; DD: $F_{(1, 68)} = 8.33$, $p < 0.01$; VA: $F_{(1, 68)} = 47.09$, $p < 0.001$]. The interaction effects were not significant ($ps > 0.5$), suggesting that the redundancy gain (RG), which is defined by the difference in mean reaction times between the single-target and redundant-target conditions, was consistently found for both groups in all the tasks.

$C(t)$s of the three redundant-target detection tasks were computed individually and were plotted by group. **Figure 2** showed the results of $C(t)$ as a function of reaction time for each group and for each task[3]. From visual inspection, all the results, except for those in the double-dot detection task, showed unlimited-capacity to supercapacity processing. Specifically, in the color-shape detection task, we did not observe any difference in $C(t)$ between the high-WMC and low-WMC groups. In this task, both groups of participants had unlimited-capacity (most of the participants had $C(t)$ equal to 1) to supercapacity (a few participants had $C(t)$ greater than 1 at the faster reaction times). In the double-dot detection task, most participants had limited-capacity processing with $C(t)$ less than 1. Lastly, in the visual-auditory detection task, both groups of participants had unlimited-capacity (a few participants had $C(t)$ equal to 1) to

---

[2]CS, DD, and VA are the abbreviations of the color-shape, double-dot, and visual-auditory detection tasks, respectively.

[3]We thank Dr. James T. Townsend for providing us the guideline to draw the figure of $C(t)$. He suggested re-scaling the figure to emphasize the value of 1 because the inference of processing capacity is made based on the comparison between the value of $C(t)$ and the value of 1.

Table 1 | Mean performance for both groups of participants in each task in Experiment 1.

Task	Group	Mean accuracy			Mean reaction time (ms)		
		RT	ST	NT	RT	ST	RG
CS	High	1.00	1.00	0.89	374.90	421.76	46.86
	Low	1.00	1.00	0.89	396.02	446.95	50.93
DD	High	1.00	1.00	0.99	370.34	393.59	23.25
	Low	1.00	1.00	0.99	391.13	419.69	28.55
VA	High	1.00	1.00	0.99	310.42	373.99	63.57
	Low	1.00	1.00	0.99	338.19	413.03	74.83

"CS," "DD," and "VA" are the abbreviations of the color-shape, double-dot, and visual-auditory detection tasks, respectively. "High" and "Low" denote the high-WMC and low-WMC group. "RT," "ST," and "NT" represent the redundant-target, single-target, and no-target conditions, respectively. RG is the abbreviation of redundancy gain and is defined as the difference in mean reaction times between the single-target and redundant-target conditions. Note that the mean reaction time of the no-target condition is not shown because any response in this condition is incorrect for a go/no-go version of the redundant-target detection task.

supercapacity (most of the participants had $C(t)$ greater than 1 at the faster reaction times). Specifically, more high-WMC participants had $C(t)$ greater than 1 at the faster reaction times than low-WMC partipipants, suggesting that high-WMC participants processed redundant visual and auditory information more efficiently.

To verify these observations, we adopted a non-parametric bootstrapping method to simulate 1000 samples for each condition and to construct the 95% confidence interval for $C(t)$ individually (Van Zandt, 2000). If the 95% confidence interval for $C(t)$ exceeds 1 at some times t, we conclude that the participant adopts supercapacity processing to process multiple signals. If the 95% confidence interval for $C(t)$ includes 1 for all times t, we conclude that the participant adopts unlimited-capacity processing. Otherwise, we conclude that the participant adopts limited-capacity processing. **Table 2** presents the classification results of the inferences based on the simulated data for each group in each task[4].

Based on the classification results, we then did two levels of analyses. First, we computed the odds ratios between the supercapacity/limited-capacity of the high-WMC group and the supercapacity/limited-capacity of the low-WMC group in different tasks. If the odds ratio equals 1, it suggests that high-WMC and low-WMC groups are classified into different WLC categories similarly, and that they have similar WLC in processing multiple signals. Otherwise, we can conclude that they have different WLC in processing multiple signals. Results showed that the odds ratio in the color-shape detection task was 1.17, suggesting that two groups of participants did not differ from each other in their WLC. In the double-dot detection task, the corrected

odds ratio was 1[5], suggesting that both the high-WMC and low-WMC groups processed multiple signals with limited capacity. In the visual-auditory detection task, the odds ratio was 7.63, suggesting that more high-WMC participants adopted supercapacity processing than low-WMC group participants did.

Second, we fitted the classification data of each task with a multinomial loglinear model which can describe the log expected frequency of each WLC category of different groups (Agresti, 1996). The model consists of a log equation with separate parameters for each WLC category of different groups. We chose the limited-capacity category of the low-WMC group as the baseline category for dummy coding. The intercept describes the log expected frequency of being classified into the baseline category and the estimated parameters for the other category describe the log expected frequency of being classified into the other WLC categories. The Wald test was conducted to examine whether each estimated parameter was significantly different from the frequency of the baseline category. The estimated proportion of being classified into one category and the baseline category can be computed. Results showed that in the color-shape detection task, all the estimated parameters were not significant ($ps > 0.2$), suggesting that the frequencies of being classified into different WLC categories for the high-WMC and low-WMC groups were comparable. That is, both groups had similar WLC in processing multiple signals. In the double-dot detection task, due to all participants being classified as limited-capacity, no further analysis was required. In the visual-auditory detection task, the estimated parameter of the high-WMC group being classified into the supercapacity category was significant [$\chi_1^2 = 6.63, p < 0.05$], and the estimated proportion between this category and the baseline category was 7. In addition, the estimated parameter of the low-WMC group being classified into the supercapacity category was marginal significant [$\chi_1^2 = 3.7, p = 0.054$], and the estimated proportion between this category and the baseline category was 4.5. Although for both groups, there were more participants classified into supercapacity category compared to the baseline category, the estimated proportion between the frequency of supercapacity category and that of the baseline category was larger for the high-WMC group than for the low-WMC group, verifying that high-WMC group had larger WLC than the low-WMC group in processing redundant visual and auditory signals.
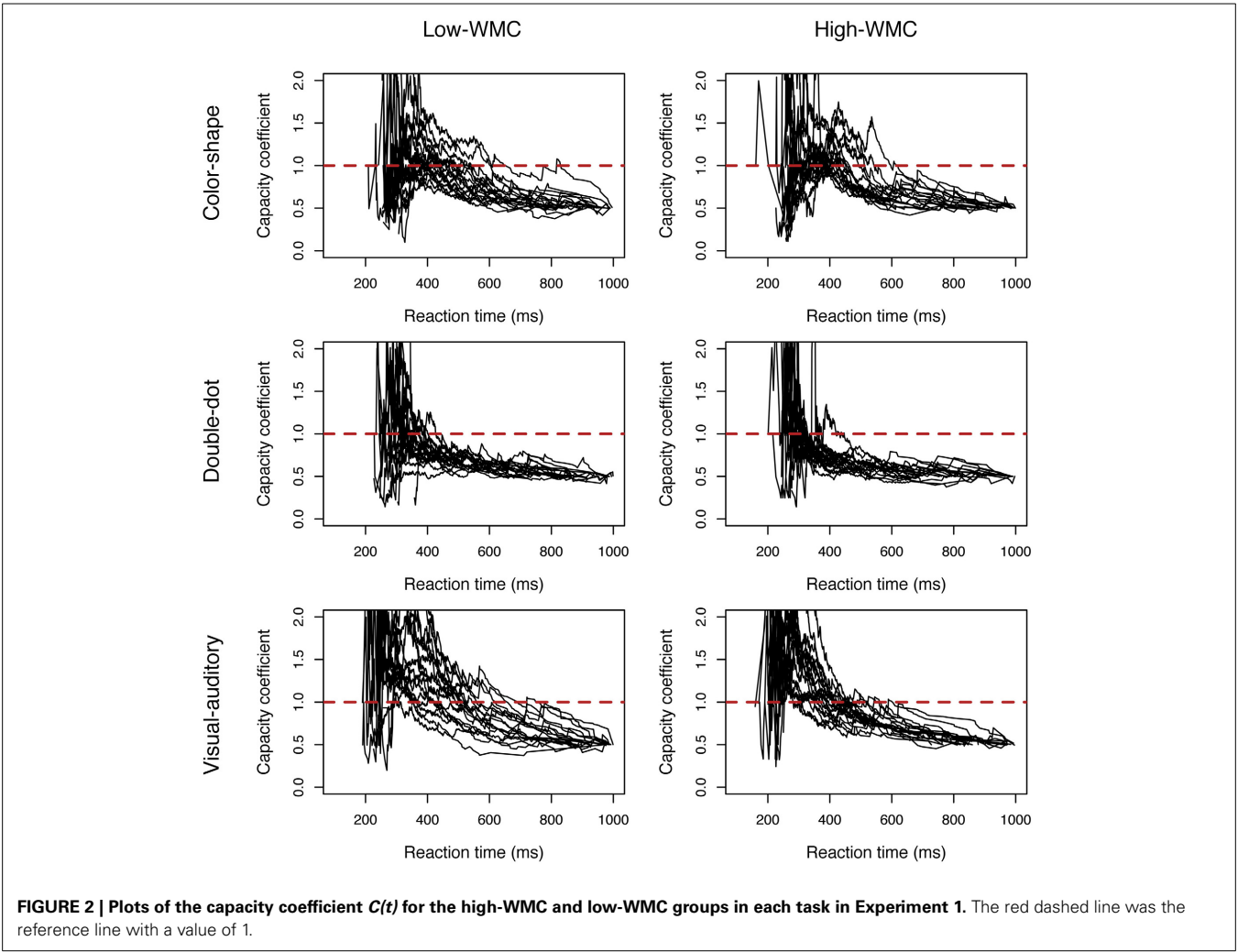
These results suggested that performance on the OSPAN task can predict the capacity of processing redundant information from different modules; however, WMC cannot predict the capacity for processing redundant featural information of an object and the capacity for processing visual information from two spatial locations.

EXPERIMENT 2
The results of Experiment 1 showed that high-WMC and low-WMC participants differed in their WLC when performing a visual-auditory detection task, but not in the other tasks.

---

[4]**Figure 2** presents the estimated $C(t)$ for each group. We do not plot the confidence interval individually in **Figure 2** due to information complexity. We summarize the inferences based on the bootstrapping results in **Table 2**.

[5]Because the value of some cells is zero, we followed Haldane's correction (Haldane, 1956) to compute the corrected odds ratio by adding 0.5 to each cell.

FIGURE 2 | Plots of the capacity coefficient *C(t)* for the high-WMC and low-WMC groups in each task in Experiment 1. The red dashed line was the reference line with a value of 1.

Table 2 | The WLC classification results of the inferences based on the simulated data for both groups in each task.

Task	Group	Supercapacity	Unlimited-capacity	Limited-capacity
CS	High	3	9	6
	Low	3	8	7
DD	High	0	0	18
	Low	0	0	18
VA	High	14	4	0
	Low	9	7	2

"CS," "DD," and "VA" are the abbreviations of the color-shape, double-dot, and visual-auditory detection tasks, respectively. "High" and "Low" denote the high-WMC and low-WMC group. The table shows the number of participants who were classified as supercapacity, unlimited-capacity, or limited-capacity for both groups in each task.

However, there were a few limitations in Experiment 1. For example, the results drawn from the non-parametric approach (SFT) can only provide a discrete distinction between the high-WMC and low-WMC groups. We were curious about whether there

is a linear relationship between WLC and WMC. Second, with the SFT, we only analyzed the correct reaction times; thus, the incorrect responses were not taken into consideration. Third, we observed a potential response bias in the color-shape detection task in Experiment 1; however, we did not collect reaction time data for the no-go response. Therefore, Experiment 2 was motivated to use a yes/no version of redundant-target detection task and adopt a parametric approach LBA (Brown and Heathcote, 2008) to estimate WLC, the LBA-based capacity, for each participant. With this approach, we incorporated both correct and incorrect reaction times and both target-present and target-absent trials into analyses. The estimated LBA-based capacity can be used to correlate with WMC to test whether there is a linear relationship between WMC and WLC. We aimed to provide converging evidence to support the relationship between WLC and WMC found in Experiment 1.

METHOD
Participants
Participants included 131 undergraduates at National Cheng Kung University who had not participated in the first experiment. Three participants were not considered in this study because they

could not participate in the OSPAN task. There were 53 males and 75 females with an average age of 19 and a standard deviation of 1.33. All the participants had normal or corrected-to-normal vision and hearing. They signed a written informed consent prior to the experiment and received NTD 120 per hour for their participation.

Stimuli, design, and procedure

The stimuli, design, and procedure were the same as those used in Experiment 1, except that a yes/no response was required. We adopted a yes/no task instead of a go/no-go task because we needed to collect the reaction times of the no-target condition to estimate the drift rates (rates of the information accumulation) and estimate the parametric measure of WLC (see details in the following **Data analysis** Section). Participants were instructed to press the "/" button when *either* or *both* target features were detected and press the "z" button when *neither* target feature was detected.

Data analysis

To estimate the parametric measure of WLC, we adopted the LBA model to analyze the reaction time data of the redundant-target detection tasks. Take the color-shape detection task for an example. Two target features, color (C) and shape (S), require four independent, parallel accumulators that collect evidence: (1) target color is present (i.e., green), (2) target color is absent (i.e., cyan), (3) target shape is present (i.e., X), and (4) target shape is absent (i.e., O). We denoted these accumulators C, $\sim$C, S, and $\sim$S, respectively. Each accumulator collects evidence from a starting point, which is uniformly distributed and ranges from 0 to A. A decision is made when the amount of accumulated evidence collected by one of the accumulators reaches the threshold b. The information accumulation rate (drift rate) of an accumulator is drawn from a normal distribution with a mean of v and a standard deviation of s. The reaction time can be separated into two components: (1) decision time: the time taken for an accumulator to reach the threshold, and (2) non-decision time (t_0), also called base time, i.e., the time taken for sensory preparation and motor execution. There are a total of five parameters used to describe an accumulator: $\theta = (b, A, v, s, t_0)$.

In the redundant-target detection task, participants were required to make a yes/no response. A *"YES"* response, indicating that either or both target features are present, is made if either C or S reaches the threshold while $\sim$C, $\sim$S, or both have not reached the threshold. Hence, the overall likelihood of a positive response at time t is the sum of the likelihoods of the two events (i.e., C reaches the threshold and S has not, and vice versa.):

$$L(YES, \ t) = [1 - F_{\sim C}(t) \cdot F_{\sim S}(t)] \cdot$$
$$\left[f_C(t) \cdot S_S(t) + f_S(t) \cdot S_C(t) \right] \quad (2)$$

where $S_i(t)$, $f_i(t)$, and $F_i(t)$ represent the survivor function, probability density function, and the cumulative distribution function of the accumulator i at time t, respectively. A *"NO"* response (neither the target color nor the target shape is present) is made if both $\sim$C and $\sim$S reach the threshold and both C and S

have not reached the threshold. Hence, the overall likelihood of a negative response is the sum of the likelihood of the two events (i.e., $\sim$C reaches threshold after $\sim$S reaches the threshold, and vice versa):

$$L(NO, \ t) = S_C(t) \cdot S_S(t) \cdot$$
$$\left[f_{\sim C}(t) \cdot F_{\sim S}(t) + f_{\sim S}(t) \cdot F_{\sim C}(t) \right] \quad (3)$$

Given a set of parameters for each condition, Equations (2) and (3) were used to evaluate the likelihood of all the correct and incorrect reaction time data. We adopted an optimization algorithm to find a set of parameters that maximized the likelihood separately for each participant. In accordance with Eidels et al. (2010), a total of eleven free parameters were used (i.e., A, b_T, b_{NT}, t_{0RT}, t_{0ST}, t_{0NT}, v_{RT}, v_{ST}, v_{NT}, $v_{\sim T}$, $v_{\sim NT}$). Because the stimulus encoding of base time may decrease with two targets versus one target due to perceptual factors, we estimated separate base time parameters of t_{0RT}, t_{0ST}, and t_{0NT} for the redundant-target, single-target, and no-target conditions, respectively. Due to the unequal number of trials between target-present (i.e., redundant-target and single-target condition) and target-absent conditions (i.e., no-target condition), participants might be biased toward making a positive response. Therefore, we estimated separate threshold parameters b_T and b_{NT} for the target-present and target-absent conditions, respectively. We estimated a single value A for the starting point across all responses and conditions. The standard deviation of the drift rate (s) was fixed at 1 in the double-dot and visual-auditory detection tasks and at 0.25 in the color-shape detection task in order to obtain the best fit for our models (see Eidels et al., 2010). We assumed five free drift rate parameters, although there could be up to 16. These five parameters were three drift rate parameters when the targets were present (v_{RT}, v_{ST}, v_{NT}) and two drift rate parameters ($v_{\sim T}$, $v_{\sim NT}$) when the targets were absent. The drift rate parameters were summarized in **Table 3**. We chose only five parameters because we assumed that drift rates were equivalent for processing C and S and for processing $\sim$C and $\sim$S. This assumption may not be true; however, when we incorporated these parameters into further analysis, we can draw the same conclusion even with a general model that possessed a larger Bayesian information criterion (BIC), indicting a worse fitting than the restricted model.

We used the relative difference between v_{RT} and v_{ST} as a parametric measure of the WLC. The LBA-based capacity can be expressed as follows:

$$v_{diff} = v_{RT} - v_{ST}. \quad (4)$$

If $v_{RT} = v_{ST}$ then unlimited-capacity processing is suggested. If the drift difference is greater or less than 0, a supercapacity processing (when $v_{RT} > v_{ST}$) or limited-capacity processing (when $v_{RT} < v_{ST}$) is suggested.

Result

As in Experiment 1, we estimated the participants' WMC by using the data of the OSPAN task. Ten participants' data were

Table 3 | The simplified set of five drift rate parameters (right-hand side) used in the LBA model and their corresponding drift rates of all accumulators (left-hand side) in the redundant-target detection task.

		Target color	
		Present (C)	**Absent (~C)**
Target shape	Present (S)	$v_{C\|CS} = v_{RT}$	$v_{C\|\sim CS} = v_{\sim T}$
		$v_{S\|CS} = v_{RT}$	$v_{S\|\sim CS} = v_{ST}$
		$v_{\sim C\|CS} = v_{NT}$	$v_{\sim C\|\sim CS} = v_{\sim NT}$
		$v_{\sim S\|CS} = v_{NT}$	$v_{\sim S\|\sim CS} = v_{NT}$
	Absent (~S)	$v_{C\|C\sim S} = v_{ST}$	$v_{C\|\sim C\sim S} = v_{\sim T}$
		$v_{S\|C\sim S} = v_{\sim T}$	$v_{S\|\sim C\sim S} = v_{\sim T}$
		$v_{\sim C\|C\sim S} = v_{NT}$	$v_{\sim C\|\sim C\sim S} = v_{\sim NT}$
		$v_{\sim S\|C\sim S} = v_{\sim NT}$	$v_{\sim S\|\sim C\sim S} = v_{\sim NT}$

Subscripts for the simplified set of five drift rates are described in the **Data analysis** Section of Experiment 2. Subscripts for the full set of sixteen drift rate parameters denote the drift rate for a specific accumulator given any of the four test trials. For instance, $v_{C\|CS}$ represents the drift rate for the accumulator C when both the target color and target shape are present and is mapped to the drift rate for the redundant-target accumulator v_{RT}.

excluded from further analysis because their processing accuracy was below 0.7. Another eleven participant' data were excluded as well because they had relatively slow mean reaction times or low accuracies in the no-target condition when they performed the redundant-target detection tasks. Under these criteria, the mean processing accuracy was 0.86 with a standard deviation of 0.07. The mean recall score was 35.75 with a standard deviation of 10.20. The high-WMC group included the participants with the top 30% of recall scores ($M = 46.86$, $SD = 5.42$, $N = 36$), whereas the low-WMC group included the participants with the bottom 30% of recall scores ($M = 24.69$, $SD = 5.38$, $N = 36$). The difference in recall scores between the high-WMC and low-WMC groups was significant [$t_{(70)} = 17.41$, $p < 0.0001$].

We then excluded the trials with reaction times less than 150 ms in the redundant-target detection tasks for further analyses. The mean performance of the redundant-target detection tasks for each group was summarized in **Table 4**. Accuracies were very high across conditions for both groups of participants except for the no-target conditions of the color-shape detection task (High: 0.88; Low: 0.89), suggesting a potential response bias in detecting color and/or shape. We will limit the remainder of our analyses to the reaction time. A two-way (high-WMC/low-WMC group × redundant-target/single-target condition) ANOVA was conducted to analyze the accuracy and correct reaction time data of the three tasks. For accuracy data, there were significant main effects of condition [CS: $F_{(1, 140)} = 148.09$, $p < 0.001$; DD: $F_{(1, 140)} = 16.77$, $p < 0.001$; VA: $F_{(1, 140)} = 187.81$, $p < 0.001$], showing lower accuracy in the no-target conditions than in the other two conditions. These results were different from what we found in Experiment 1 where accuracy in most conditions reached the ceiling. For reaction time data, there were significant main effects of group in the color-shape and double-dot detection task [CS: $F_{(1, 140)} = 12.76$, $p < 0.001$; DD: $F_{(1, 140)} = 5.14$, $p <$

Table 4 | Mean performance for both groups of participants in each task in Experiment 2.

| Task	Group	Mean accuracy			Mean reaction time (ms)			
		RT	ST	NT	RT	ST	NT	RG
CS	High	1.00	0.97	0.88	346.72	390.01	454.67	43.29
	Low	1.00	0.98	0.89	369.78	418.83	491.96	49.05
DD	High	0.99	0.98	0.93	400.13	420.46	574.43	20.33
	Low	0.99	0.98	0.93	417.35	440.78	595.11	23.43
VA	High	1.00	0.97	0.93	323.27	382.87	539.70	59.60
	Low	1.00	0.97	0.91	329.31	391.29	556.98	61.99

"CS," "DD," and "VA" are the abbreviations of the color-shape, double-dot, and visual-auditory detection tasks, respectively. "High" and "Low" denote the high-WMC and low-WMC groups. "RT," "ST," and "NT" represent the redundant-target, single-target, and no-target conditions, respectively. RG is the abbreviation of redundancy gain and is defined as the difference in mean reaction times between the single-target and redundant-target conditions.

0.05; VA: $F_{(1, 140)} = 0.82$, $p = 0.37$] and condition in all the tasks [CS: $F_{(1, 140)} = 40.42$, $p < 0.001$; DD: $F_{(1, 140)} = 6.98$, $p < 0.01$; VA: $F_{(1, 140)} = 58.05$, $p < 0.001$]. The interaction effects were not significant in all the tasks for both groups ($ps > 0.2$), suggesting that the RG was consistently found for both groups in all the tasks.

$C(t)$s of the three redundant-target detection tasks were computed individually and were plotted by group. **Figure 3** showed the results of $C(t)$ as a function of reaction time for both groups in each task. The results in Experiment 2 were comparable to those in Experiment 1. In the color-shape and double-dot detection task, no difference in $C(t)$ between the high-WMC and low-WMC groups was observed. Both groups of participants had unlimited-capacity in processing color and shape with $C(t)$ equal to 1 for all times t; however, we found a few participants had $C(t)$ greater than 1 at the faster RTs. In the double-dot detection task, most participants had limited-capacity processing with $C(t)$ less than 1. Finally, in the visual-auditory detection task, both groups of participants had $C(t)$ greater than 1 at the faster RTs, suggesting supercapacity processing. In addition, more high-WMC participants showed this pattern than low-WMC participants did, suggesting that high-WMC participants processed redundant visual and auditory information more efficiently. The results of Experiment 1 can be generalized to a yes/no task.

To verify these observations, we adopted the non-parametric bootstrapping method as Experiment 1 to construct the 95% confidence interval for $C(t)$ of all the tasks and for each participant. **Table 5** presents the classification results of the inferences based on the simulated data for both groups in each task.

We then computed the odds ratios between the supercapacity/limited-capacity of the high-WMC and the supercapacity/limited-capacity of the low-WMC group in the three tasks. Results showed that the odds ratios were 1.42 in the color-shape detection task and 0.97 in the double-dot detection task, suggesting that the two groups were classified into different WLC categories similarly. In the visual-auditory detection task, the odds ratio was 3.55, suggesting that more
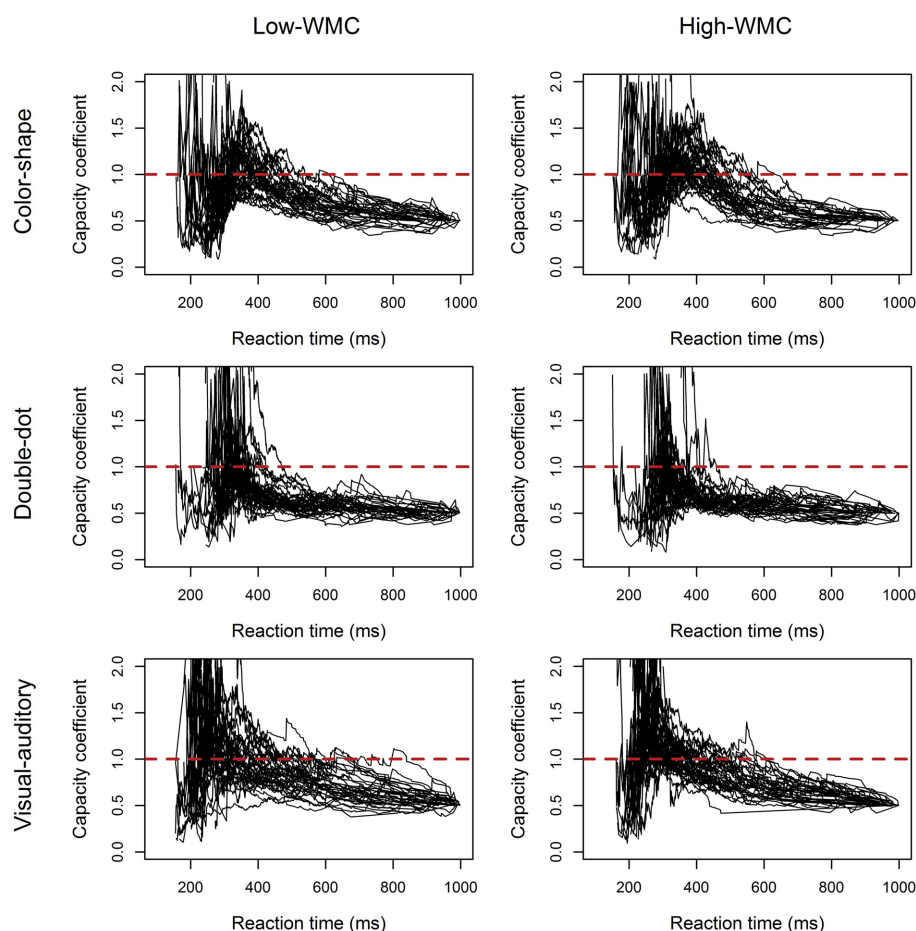
FIGURE 3 | Plots of the capacity coefficient *C(t)* for the high-WMC and low-WMC groups in each task in Experiment 2. The red dashed line was the reference line with a value of 1.

Table 5 | The WLC classification results of the inferences based on the simulated data for both groups in each task.

Task	Group	Supercapacity	Unlimited-capacity	Limited-capacity
CS	High	2	26	7
	Low	2	24	10
DD	High	2	1	33
	Low	2	2	32
VA	High	13	21	1
	Low	11	21	3

"CS," "DD," and "VA" are the abbreviations of the color-shape, double-dot, and visual-auditory detection tasks, respectively. "High" and "Low" denote the high-WMC and low-WMC group. The table shows the number of participants who were classified as supercapacity, unlimited-capacity, or limited-capacity for both groups in each task.

participants adopted supercapacity processing in the high-WMC group than in the low-WMC group.

The results analyzed with the multinomial loglinear model also supported our observations. Results showed that in the color-shape detection task, the estimated parameters of the high-WMC group being classified into the supercapacity category and unlimited-capacity category were significant (supercapacity: $\chi_1^2 = 4.32$, $p < 0.05$; unlimited-capacity: $\chi_1^2 = 6.59$, $p < 0.05$), and the estimated proportions between these categories and the baseline category were 0.2 and 2.6. Also, the estimated parameters of the low-WMC group showed a similar pattern of results (supercapacity: $\chi_1^2 = 4.32$, $p < 0.05$; unlimited-capacity: $\chi_1^2 = 5.41$, $p < 0.05$), and the estimated proportions between these categories and the baseline category were 0.2 and 2.4. These results suggested that the two groups were classified into different WLC categories similarly. In the double-dot detection task, the results of the estimated parameters were significant for both the high-WMC (supercapacity: $\chi_1^2 = 14.47$, $p < 0.001$; unlimited-capacity: $\chi_1^2 = 11.65$, $p < 0.001$) and low-WMC groups (supercapacity: $\chi_1^2 = 14.47$, $p < 0.001$; unlimited-capacity: $\chi_1^2 = 14.47$, $p < 0.001$). The estimated proportions between high and supercapacity, high and unlimited-capacity, low and supercapacity, and low and unlimited-capacity categories and the baseline category were 0.06, 0.03, 0.06, and 0.06, respectively, suggesting more participants were classified into limited-capacity category for both groups. In the visual-auditory

task, the estimated parameters were significant for both the high-WMC (supercapacity: $\chi_1^2 = 5.24$, $p < 0.05$; unlimited-capacity: $\chi_1^2 = 9.94$, $p < 0.005$) and low-WMC groups (supercapacity: $\chi_1^2 = 3.98$, $p < 0.05$; unlimited-capacity: $\chi_1^2 = 9.94$, $p < 0.005$). The estimated proportions between high and supercapacity, high and unlimited-capacity, low and supercapacity, and low and unlimited-capacity categories and the baseline category were 4.33, 7, 3.67, and 7, respectively. The estimated proportion between the supercapacity category and the baseline category was larger for the high-WMC group than for the low-WMC group, verifying that the high-WMC group had larger WLC than the low-WMC group in processing redundant visual and auditory signals.

However, comparing the results between the two experiments, we found that fewer participants were classified into supercpacity category in Experiment 2 than in Experiment 1. This discrepancy may be due to the nature of the tasks used in the two experiments (go/no-go vs. yes/no tasks). It is worthy to note that our findings were consistent with the previous research (Blurton et al., 2014), in which the race-model inequality was easily violated in a go/no-go task compared to a forced-choice task.

Next, we used the LBA model to analyze the reaction time data and estimated a set of parameters that maximized the likelihood function described in the **Method** Section for each participant. **Table 6** presented the average of 11 estimated parameters for both groups in different tasks. We then used the average of the estimated parameters to simulate data and plotted the model predictions based on the simulated data on top of the empirical histogram (see **Figure 4**). Results showed that the LBA model fitted the participants' reaction time data because the predicted density from the model can capture the empirical density successfully.

We then computed the LBA-based capacity for both groups in each task (see **Figure 5**). Results showed a significant difference in the LBA-based capacity between the high-WMC and low-WMC groups in the visual-auditory detection task [$t_{(70)} = 2.36$, $p < 0.05$]; however, this difference was not observed in the color-shape detection task ($p = 0.35$) and in the double-dot detection task ($p = 0.55$). Finally, we computed the Pearson's product-moment correlation (r) between the recall scores and the LBA-based capacity. A significant positive correlation between the WMC and WLC was found in the visual-auditory detection [$r = 0.25$, $p < 0.01$, 95% CI = (0.06, 0.41)], whereas the correlations in the color-shape detection task [$r = 0.02$, $p = 0.83$, 95%

CI = (−0.17, 0.21)] and double-dot detection task [$r = 0.05$, $p = 0.61$, 95% CI = (−0.14, 0.24)] did not reach the significance level (see **Figure 6**). These results provided converging evidence showing that participants high in WMC had larger WLC only in the visual-auditory detection task.
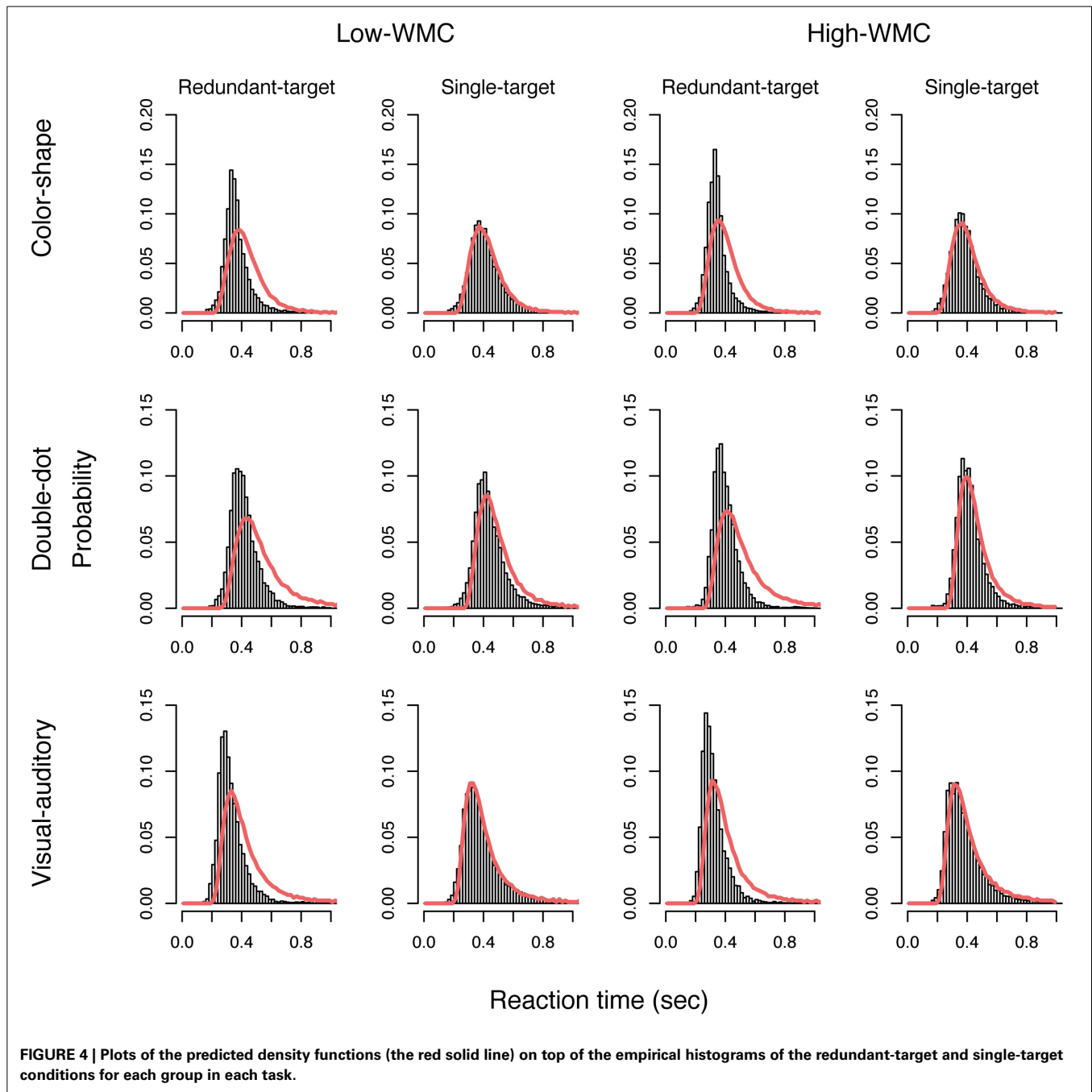
DISCUSSION

We examined the relationship between WMC and WLC, and tested whether the two capacity measures assessed a unitary, central capacity of information processing. We used an OSPAN task to assess WMC and three different redundant-target detection tasks to assess WLC. We conducted an extreme-group approach to split the participants' data according to their WMCs and compared them to their WLCs in both experiments, and computed the Pearson's product-moment correlation to verify the linear relationship between the two capacity measures in Experiment 2. WLC was estimated with the reaction time data of the redundant-target detection tasks both non-parametrically (SFT in Experiments 1 and 2) and parametrically (LBA in Experiment 2). The results from the two experiments showed that participants high in WMC had a larger perceptual processing capacity in detecting multiple signals from different modalities (the visual-auditory detection task); this difference was eliminated when multiple signals came from different object features (the color-shape detection task) and from different spatial locations (the double-dot detection task). These results suggested that the individual differences in WMC can predict the ability to process multiple sources of information in a certain perceptual task and shed light on the functioning of the central executive system of working memory in multiple-signal processing. Further implications on the nature of a working memory system will be discussed.

In the model of working memory (Baddeley and Hitch, 1974), central executive system plays an important role in maintaining, updating, operating, and integrating information between percepts and the two subsystems, which store visuospatial and phonological information, respectively. In previous research on working memory, measures of WMC are strongly correlated to performance in various complex cognitive tasks, such as reading comprehension (McVay and Kane, 2012), logical reasoning (Oberauer et al., 2007), problem solving (Hoffman and Schraw, 2009), and creative thinking (Dietrich, 2004). In addition,

Table 6 | The average values of eleven estimated parameters and the LBA-based capacity (v_{Diff}) for both groups of participants in each task.

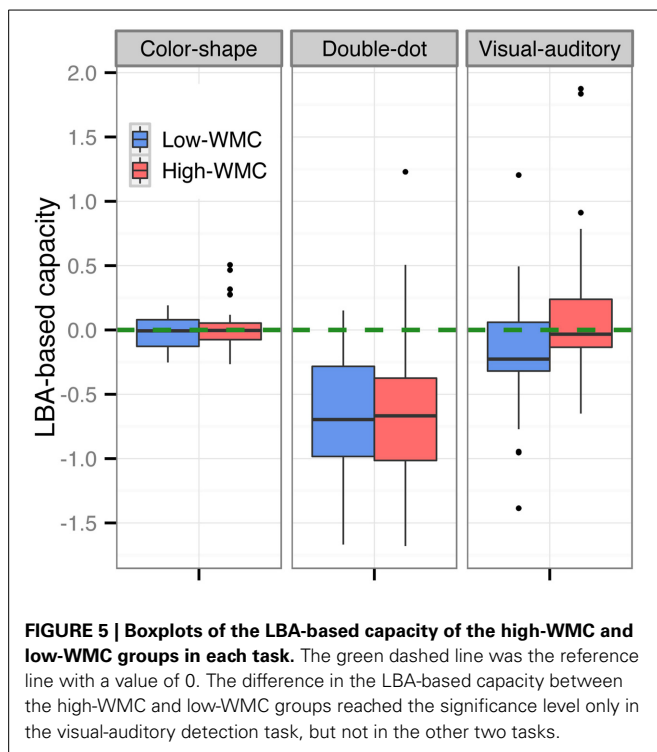| Task | Group | Estimated parameters | | | | | | | | | | |
		A	t_{0RT}	t_{0ST}	t_{0NT}	b_T	b_{NT}	v_{RT}	v_{ST}	v_{NT}	$v_{\sim T}$	$v_{\sim NT}$	v_{Diff}
CS	High	165.64	54.81	51.45	26.79	439.67	532.07	1.08	1.06	0.42	0.67	1.23	0.02
	Low	199.93	60.86	63.15	23.59	476.50	609.03	1.07	1.08	0.38	0.62	1.27	−0.02
DD	High	477.73	156.91	175.91	159.98	1080.78	1031.60	2.67	3.28	−0.85	−0.82	2.47	−0.61
	Low	516.01	140.98	164.25	136.27	1214.55	1104.19	2.72	3.41	−0.29	−0.46	2.44	−0.69
VA	High	269.24	104.98	101.26	89.21	841.62	948.63	2.63	2.52	−0.38	0.09	2.39	0.11
	Low	296.69	90.34	93.20	50.35	930.58	1060.69	2.64	2.81	−0.51	0.28	2.40	−0.18

"CS," "DD," and "VA" are the abbreviations of the color-shape, double-dot, and visual-auditory detection tasks, respectively. "High" and "Low" denote the high-WMC and low-WMC groups.

FIGURE 4 | Plots of the predicted density functions (the red solid line) on top of the empirical histograms of the redundant-target and single-target conditions for each group in each task.

differences in WMC can account for variation in individuals' general intelligence quotient (IQ) (Engle et al., 1999; Kane and Engle, 2002; Conway et al., 2003). Previous researchers suggest that WMC reflects the efficiency of the central executive system in maintaining and processing information (see Barrett et al., 2004 for a review), most notably the ability to selectively maintain task-relevant information (Redick et al., 2007; Lecerf and Roulin, 2009; Minamoto et al., 2010). WMC also reflects individual differences in the ability to focus and maintain attention in binding and integrating multiple sources of information (Barrett et al., 2004), particularly when a salient distractor is likely to capture
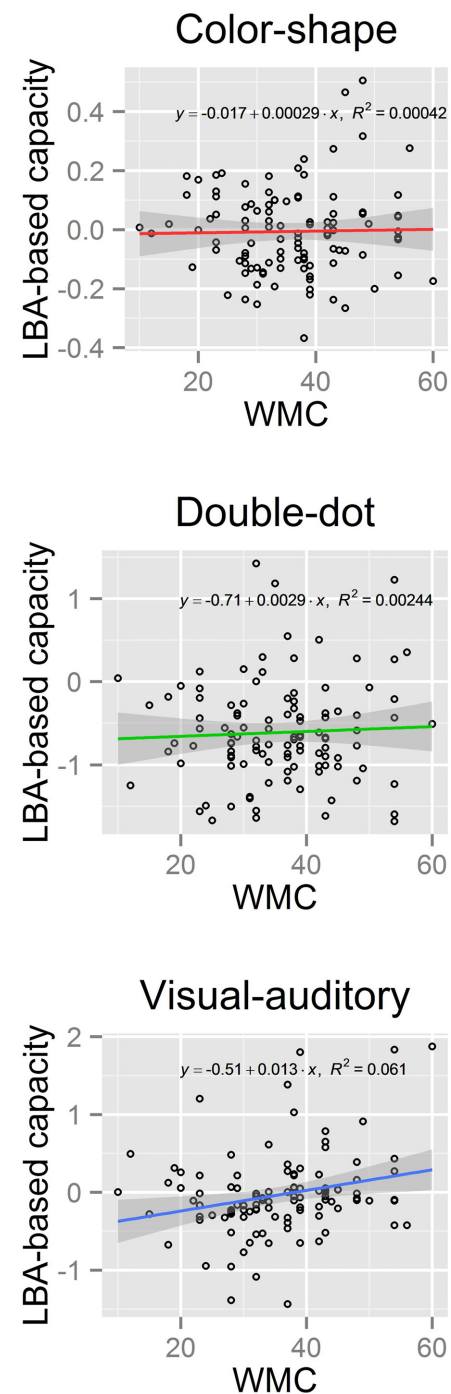
attention; this ability may decrease with age (Palladino and Beni, 1999). Recently, a number of neuroimaging studies have demonstrated the role of the prefrontal cortex in executive function (e.g., Miller and Cohen, 2001; Kane and Engle, 2002).

On the other hand, WLC measures the variation of the processing efficiency of an individual channel as a function of workload (Townsend and Ashby, 1978; Townsend and Nozawa, 1995; Wenger and Gibson, 2004; Townsend and Eidels, 2011). In previous research, WLC has been assessed with a redundant-target detection task (see Townsend and Nozawa, 1995 for a review) in various perceptual domains, such as simple detection (Townsend

FIGURE 5 | Boxplots of the LBA-based capacity of the high-WMC and low-WMC groups in each task. The green dashed line was the reference line with a value of 0. The difference in the LBA-based capacity between the high-WMC and low-WMC groups reached the significance level only in the visual-auditory detection task, but not in the other two tasks.

and Eidels, 2011), visual search (Fifić et al., 2008), memory search (Townsend and Fifić, 2004), face perception (Fifić et al., 2008), categorization (Fifić et al., 2010), multisensory perception (Altieri and Townsend, 2011), and change detection (Yang, 2011; Yang et al., 2011, 2013). WLC is likely to constrain the order of multiple-signal processing. For example, a coactive system is usually of supercapacity (Wenger and Townsend, 2001); an independent parallel system is found to be of unlimited-capacity (Houpt and Townsend, 2012); a standard serial model is of limited-capacity (Townsend and Ashby, 1983). In addition, according to Eidels et al. (2011), multiple processes may interact with each other when a parallel system is of supercapacity or limited-capacity processing. Therefore, it is reasonable to speculate that when participants have a system of larger processing capacity, especially supercapacity, they can process redundant information more efficiently and the multiple processes can be completed in a coactive fashion, or that there would be facilitatory between-channel crosstalk during information accumulation such that the participants can optimize the use of multiple signals in perceptual decision making. In contrast, when participants have a system of limited-capacity processing, they are limited in processing multiple signals such that multiple processes may be completed in a serial fashion. Limited-capacity processing may also indicate that there is an inhibitory interaction during information accumulation, such that processing one channel of information can inhibit the other process, leading to slower individual-channel processing.

Instead of aggregating all the participants' data to do group analysis, a few recent studies inferred individuals' information processing characteristics by examining their reaction time data, focusing mostly on the individual differences in their processing

FIGURE 6 | Scatter plots of the LBA-based capacity and the WMC scores with a regression line (solid colored line) and its 95% confidence interval (band-shaped gray area) of each task. A significant positive correlation between WMC and WLC was found in the visual-auditory detection task, but not in the other two tasks.

strategies and processing capacity. For example, Yang et al. (2011) found individual differences in processing strategies when participants were required to detect a luminance change and an orientation change of a Gabor patch, and the relative decision

difficulty between the two feature-changes were not controlled. One group of participants adopted serial self-terminating processing with limited capacity, and the other group adopted coactive processing with supercapacity. In Yang's (2011) study, when relative saliency existed in detecting an orientation change and a frequency change of a Gabor patch, three participants adopted serial self-terminating processing with limited-capacity to unlimited-capacity processing to detect changes, while one participant adopted parallel self-terminating processing with unlimited-capacity processing. Similarly, in a categorization task, Fifić et al. (2010) found that participants used multiple sources of information differently to make a categorization decision. However, these studies did not explain the causes of individual differences in processing strategies. We speculated that limits in the processing capacity might constrain the information processing strategy. These individual variations in processing capacity can be predicted by ones' capacity of executive attentional control of a working memory system in processing information.

Although WMC and WLC were proposed around the same time, no prior studies, except for a recent one conducted by Heathcote et al. (2014), investigated the relationship between the two capacity measures. Theoretically, the two capacity measures assessed some similar characteristics of information processing. Most notably, controlled attention played an important role in a working memory system (Rosen and Engle, 1997; Engle et al., 1999; Barrett et al., 2004; Engle and Kane, 2004) and in multiple-signal processing at perception, such as feature integration (Treisman and Gelade, 1980), goal-derived visual selection (Bargh, 1982), perceptual organization (Mack et al., 1992), and perceptual learning (Shiffrin and Schneider, 1977). Thus, it was reasonable to hypothesize that these two measures may relate to each other to a certain extent.

Heathcote et al. (2014) adopted a mnemonic redundant-target task to measure WLC. Participants were required to respond if either the auditory or visual target was presented two-back in a trial sequence. This task incorporated the test for working memory into the measurement of WLC, which was different from the OSPAN task used in assessing WMC. They also followed the SFT to estimate WLC by comparing the reaction time data between the redundant-target and single-target conditions. Unfortunately, their preliminary results did not show a clear relationship between the measurements of WLC and WMC. They suggested that these two capacity measures did not assess a unitary, central processing capacity. However, the fact that they could not find a significant correlation may be due to the lack of statistical power.

The present study used three perceptual redundant-target detection tasks instead of the mnemonic redundant-target task used by Heathcote et al. (2014) and tested the relationship between WLC and WMC. We found interesting results. First, we found significant differences in the LBA-based capacity between different perceptual tasks [$F_{(2, 210)} = 47.57$, $p < 0.0001$] (see **Figure 5**), and the results from the non-parametric analyses confirmed this pattern of results (see **Figures 2**, **3**, **Tables 2**, **5**); however, the non-parametric results also showed variations between individuals. Generally, processing capacity was the largest in the visual-auditory task, then in the color-shape detection task, and

smallest in the double-dot detection task. These results were consistent with prior research. For example, a number of studies have demonstrated that processing multisensory information was of supercapacity, which was known as an effect of "multisensory integration" (Hugenschmidt et al., 2010; Altieri and Townsend, 2011). One of the best-known studies conducted by Miller (1982) showed that when participants performed a visual-auditory detection task, the race-model inequality (RMI) was violated, suggesting that participants adopted coactive processing with supercapacity in processing multisensory information. Our study found that the LBA-based capacity was greater than or equal to 0 and that most participants had $C(t)$ greater than 1 at the faster RTs, indicating supercapacity processing of information from different modalities. On the other hand, Mordkoff and Yantis (1991, 1993) have tested the processing for color and shape of an object. In their studies, the race-model inequality was violated when inter-stimulus contingency existed, while it was not violated when there was no inter-stimulus contingency. In the present study, we did not manipulate the inter-stimulus contingency, and we found that the LBA-based capacity was equal to 0, which was consistent with Mordkoff and Yantis's (1991) findings of unlimited-capacity processing without any manipulation of probability information. Lastly, a few studies have demonstrated limited-capacity processing in double-dot detection. The present study also found that the LBA-based capacity was less than 1 in the double-dot detection task, indicating limited-capacity processing. However, when we looked at the non-parametric results (see **Figures 2**, **3**), we found a few participants had $C(t)$ greater than or equal to 1 at the faster RTs, indicating that they may process multiple spatial locations with supercapacity or unlimited-capacity processing. Even though, most participants had $C(t)$ less than 1 for all times t, indicating limited-capacity processing.

Most interestingly, we found differences in WLC between the high-WMC and low-WMC groups. The differences were only found in the visual-auditory detection task, but not in the other two tasks, and this difference was comparable between the two experiments (see **Figures 2**, **3**, **5**). **Figure 6** also shows that there was a significant positive correlation between WMC and WLC only when participants performed a visual-auditory detection task. These results suggested that WMC correlated to WLC only when a system needed to integrate multiple signals from two different subsystems (i.e., visuospatial sketchpad and phonological loop) for manipulation, operation, and decision making. This relationship was not observed when a system integrated multiple signals that only required resources from a single subsystem (i.e., the visuospatial sketchpad in the present study). These results indicated that a domain-general resource was required for the controlled attention to integrate and bind multisensory information for decision making. On the other hand, processing redundant information from a single modality required a domain-specific resource that was not necessarily related to WMC. Nonetheless, future studies should examine the individual differences in information processing of a single subsystem, as we did not test the processing of redundant information that originated from a single auditory modality. Individual differences can be discovered by increasing the sample size to increase the

statistical power and by testing its generalizability in different experimental contexts.

CONCLUSION

We examined the relationship between WMC and WLC. Both the non-parametric and parametric analyses showed that participants high in WMC had larger WLC in processing redundant information from different modalities, suggesting that they processed redundant visual and auditory signals more efficiently and multiple processes were likely to be completed in a coactive fashion. However, the difference was not observed when processing redundant information from a single visual modality. The results highlighted the role of controlled attention in information integration of working memory and multiple-signal processing at perception and further contributed to the understanding of the nature of a working memory system.

AUTHOR CONTRIBUTIONS

Ju-Chi Yu—Data acquisition, data analysis, and drafting the manuscript. Ting-Yun Chang—Programming, data collection, and data analysis. Cheng-Ta Yang—Conception and design, data interpretation, drafting the manuscript, and final approval.

REFERENCES

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. New York, NY: Wiley.

Altieri, N., and Townsend, J. T. (2011). An assessment of behavioral dynamic information processing measures in audiovisual speech perception. *Front. Psychol.* 2:238. doi: 10.3389/fpsyg.2011.00238

Anderson, J. R. (2013). *The Architecture of Cognition*. Cambridge, MA: Harvarad Universtiy Press.

Baddeley, A. D., and Hitch, G. J. (1974). "Working memory," in *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 8, ed G. A. Bower (New York, NY: Academic Press Inc), 47–89.

Bargh, J. A. (1982). *Attention and Automaticity in the Processing of Self-Relevant Information*. American Psychological Association.

Barrett, L. F., Tugade, M. M., and Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychol. Bull.* 130, 553–573. doi: 10.1037/0033-2909.130.4.553

Blurton, S. P., Greenlee, M. W., and Gondan, M. (2014). Multisensory processing of redundant information in go/no-go and choice responses. *Atten. Percept. Psychophys.* 76, 1212–1233. doi: 10.3758/s13414-014-0644-0

Brown, S. D., and Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cogn. Psychol.* 57, 153–178. doi: 10.1016/j.cogpsych.2007.12.002

Case, R., Kurland, D. M., and Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *J. Exp. Child Psychol.* 33, 386–404. doi: 10.1016/0022-0965(82)90054-6

Conway, A. R. A., Kane, M. J., and Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends Cogn. Sci.* 7, 547–552. doi: 10.1016/j.tics.2003.10.005

Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verb. Learn. Verb. Behav.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6

Dietrich, A. (2004). The cognitive neuroscience of creativity. *Psychon. Bull. Rev.* 11, 1011–1026. doi: 10.3758/BF03196731

Eidels, A., Donkin, C., Brown, S. D., and Heathcote, A. (2010). Converging measures of workload capacity. *Psychon. Bull. Rev.* 17, 763–771. doi: 10.3758/PBR.17.6.763

Eidels, A., Houpt, J. W., Altieri, N., Pei, L., and Townsend, J. T. (2011). Nice guys finish fast and bad guys finish last: facilitatory vs. inhibitory interaction in parallel systems. *J. Math. Psychol.* 55, 176–190. doi: 10.1016/j.jmp.2010.11.003

Engle, R. W., and Kane, M. J. (2004). "Executive attention, working memory capacity, and a two-factor theory of cognitive control," in *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 44, ed B. Ross (New York, NY: Elsevier), 145–199.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., and Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *J. Exp. Psychol. Gen.* 128, 309–331. doi: 10.1037/0096-3445.128.3.309

Fifić, M., Townsend, J. T., and Eidels, A. (2008). Studying visual search using systems factorial methodology with target—distractor similarity as the factor. *Percept. Psychophys.* 70, 583–603. doi: 10.3758/PP.70.4.583

Fifić, M., Little, D. R., and Nosofsky, R. M. (2010). Logical-rule models of classification response times: a synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychol. Rev.* 117, 309–348. doi: 10.1037/a0018526

Haldane, J. B. S. (1956). The estimation and significance of the logarithm of a ratio of frequencies. *Ann. Hum. Genet.* 20, 309–311. doi: 10.1111/j.1469-1809.1955.tb01285.x

Heathcote, A., Eidels, A., Houpt, J. W., Coleman, J., Watson, J., and Strayer, D. (2014). "Multi-tasking in working memory," in *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, eds P. Bello, M. Guarini, M. McShane, and B. Scassellati (Quebec City, QC), 601–606.

Henderson, J. (2013). *Memory and Forgetting*. New York, NY: Routledge.

Hoffman, B., and Schraw, G. (2009). The influence of self-efficacy and working memory capacity on problem-solving efficiency. *Learn. Individ. Dif.* 19, 91–100. doi: 10.1016/j.lindif.2008.08.001

Houpt, J. W., and Townsend, J. T. (2012). Statistical measures for workload capacity analysis. *J. Math. Psychol.* 56, 341–355. doi: 10.1016/j.jmp.2012.05.004

Hugenschmidt, C. E., Hayasaka, S., Peiffer, A. M., and Laurienti, P. J. (2010). Applying capacity analyses to psychophysical evaluation of multisensory interactions. *Inf. Fusion* 11, 12–20. doi: 10.1016/j.inffus.2009.04.004

Kane, M. J., and Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychon. Bull. Rev.* 9, 637–671. doi: 10.3758/BF03196323

Lecerf, T., and Roulin, J.-L. (2009). Individual differences in visuospatial working memory capacity and distractor inhibition. *Swiss J. Psychol.* 68, 67–78. doi: 10.1024/1421-0185.68.2.67

Mack, A., Tang, B., Tuma, R., Kahn, S., and Rock, I. (1992). Perceptual organization and attention. *Cogn. Psychol.* 24, 475–501. doi: 10.1016/0010-0285(92)90016-U

McVay, J. C., and Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *J. Exp. Psychol. Gen.* 141, 302–320. doi: 10.1037/a0025250

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cogn. Psychol.* 14, 247–279. doi: 10.1016/0010-0285(82)90010-X

Minamoto, T., Osaka, M., and Osaka, N. (2010). Individual differences in working memory capacity and distractor processing: possible contribution of top–down inhibitory control. *Brain Res.* 1335, 63–73. doi: 10.1016/j.brainres.2010.03.088

Mordkoff, J. T., and Yantis, S. (1991). An interactive race model of divided attention. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 520–538. doi: 10.1037/0096-1523.17.2.520

Mordkoff, J. T., and Yantis, S. (1993). Dividing attention between color and shape: evidence of coactivation. *Percept. Psychophys.* 53, 357–366. doi: 10.3758/BF03206778

Mullin, P. A., Egeth, H. E., and Mordkoff, J. T. (1988). Redundant-target detection and processing capacity: the problem of positional preferences. *Percept. Psychophys.* 43, 607–610. doi: 10.3758/BF03207751

Norman, D. A., and Shallice, T. (1986). "Attention to action: willed and automatic control of behaviour," in *Consciousness and Self-Regulation: Advances in Research and Theory*, eds R. J. Davidson, G. E. Schwartz, and D. Shapiro (New York, NY: Springer US), 1–18.

Oberauer, K., Süß, H.-M., Wilhelm, O., and Sander, N. (2007). "Individual differences in working memory capacity and reasoning ability," in *Variation in Working Memory*, eds A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, and J. N. Towse (New York, NY: Oxford University Press), 49–75.

Palladino, P., and Beni, R. (1999). Working memory in aging: maintenance and suppression. *Aging Clin. Exp. Res.* 11, 301–306. doi: 10.1007/BF03339804

Posner, M. I., and Digirolamo, G. J. (2000). Cognitive neuroscience: origins and promise. *Psychol. Bull.* 126, 873–889. doi: 10.1037/0033-2909.126.6.873

Preacher, K. J., Rucker, D. D., Maccallum, R. C., and Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychol. Methods* 10, 178–192. doi: 10.1037/1082-989X.10.2.178

Redick, T. S., Heitz, R. P., and Engle, R. W. (2007). "Working memory capacity and inhibition: cognitive and social consequences," in *Inhibition in Cognition*, eds D. S. Gorfein and C. M. Macleod (Washington, DC: American Psychological Association), 125–142.

Rosen, V. M., and Engle, R. W. (1997). The role of working memory capacity in retrieval. *J. Exp. Psychol. Gen.* 126, 211–227. doi: 10.1037/0096-3445.126.3.211

Schneider, W., Eschman, A., and Zuccolotto, A. (2012). *E-Prime: User's Guide.* Pittsburg, PA: Psychology Software Tools, Inc.

Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychol. Rev.* 84, 127–190. doi: 10.1037/0033-295X.84.2.127

Townsend, J. T., and Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. *Cogn. Theory* 3, 200–239.

Townsend, J. T., and Ashby, F. G. (1983). *The Stochastic Modeling of Elementary Psychological Processes.* Cambridge: Cambridge University Press.

Townsend, J. T., and Eidels, A. (2011). Workload capacity spaces: a unified methodology for response time measures of efficiency as workload is varied. *Psychon. Bull. Rev.* 18, 659–681. doi: 10.3758/s13423-011-0106-9

Townsend, J. T., and Fifić, M. (2004). Parallel versus serial processing and individual differences in high-speed search in human memory. *Percept. Psychophys.* 66, 953–962. doi: 10.3758/BF03194987

Townsend, J. T., and Nozawa, G. (1995). Spatio-temporal properties of elementary perception: an investigation of parallel, serial, and coactive theories. *J. Math. Psychol.* 39, 321–359. doi: 10.1006/jmps.1995.1033

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5

Turner, M. L., and Engle, R. W. (1989). Is working memory capacity task dependent? *J. Mem. Lang.* 28, 127–154. doi: 10.1016/0749-596X(89)90040-5

Van Zandt, T. (2000). How to fit a response time distribution. *Psychon. Bull. Rev.* 7, 424–465. doi: 10.3758/BF03214357

Wenger, M. J., and Gibson, B. S. (2004). Using hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 708–719. doi: 10.1037/0096-1523.30.4.708

Wenger, M. J., and Townsend, J. T. (2001). "Faces as gestalt stimuli: process characteristics," in *Computational, Geometric, and Process Perspectives on Facial Cognition: Contexts and Challenges*, ed M. J. Wenger and J. T. Townsend (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 229–284.

Yang, C.-T. (2011). Relative saliency in change signals affects perceptual comparison and decision processes in change detection. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1708–1728. doi: 10.1037/a0024257

Yang, C.-T., Chang, T.-Y., and Wu, C.-J. (2013). Relative change probability affects the decision process of detecting multiple feature changes. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 1365–1385. doi: 10.1037/a0030693

Yang, C.-T., Hsu, Y.-F., Huang, H.-Y., and Yeh, Y.-Y. (2011). Relative salience affects the process of detecting changes in orientation and luminance. *Acta Psychol.* 138, 377–389. doi: 10.1016/j.actpsy.2011.09.003

Yang, C.-T., Little, D. R., and Hsu, C.-C. (2014). The influence of cueing on attentional focus in perceptual decision making. *Attent. Percept. Psychophys.* 76, 2256–2275. doi: 10.3758/s13414-014-0709-0

Zehetleitner, M., Krummenacher, J., and Müller, H. J. (2009). The detection of feature singletons defined in two dimensions is based on salience summation, rather than on serial exhaustive or interactive race architectures. *Attent. Percept. Psychophys.* 71, 1739–1759. doi: 10.3758/APP.71.8.1739

Working memory capacity and redundant information processing efficiency

Michael J. Endres[1,2]*, Joseph W. Houpt[3], Chris Donkin[4] and Peter R. Finn[5]

[1] Department of Health, Behavioral Health Administration, Honolulu, HI, USA, [2] Department of Psychology, University of Hawaii, Honolulu, HI, USA, [3] Department of Psychology, Wright State University, Dayton, OH, USA, [4] School of Psychology, University of New South Wales, Sydney, NSW, Australia, [5] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

Working memory capacity (WMC) is typically measured by the amount of task-relevant information an individual can keep in mind while resisting distraction or interference from task-irrelevant information. The current research investigated the extent to which differences in WMC were associated with performance on a novel redundant memory probes (RMP) task that systematically varied the amount of to-be-remembered (targets) and to-be-ignored (distractor) information. The RMP task was designed to both facilitate and inhibit working memory search processes, as evidenced by differences in accuracy, response time, and Linear Ballistic Accumulator (LBA) model estimates of information processing efficiency. Participants ($N = 170$) completed standard intelligence tests and dual-span WMC tasks, along with the RMP task. As expected, accuracy, response-time, and LBA model results indicated memory search and retrieval processes were facilitated under redundant-target conditions, but also inhibited under mixed target/distractor and redundant-distractor conditions. Repeated measures analyses also indicated that, while individuals classified as high ($n = 85$) and low ($n = 85$) WMC did not differ in the magnitude of redundancy effects, groups did differ in the efficiency of memory search and retrieval processes overall. Results suggest that redundant information reliably facilitates and inhibits the efficiency or speed of working memory search, and these effects are independent of more general limits and individual differences in the capacity or space of working memory.

Keywords: working memory capacity, systems factorial technology, linear ballistic accumulator, individual differences, memory retrieval

1. Introduction

Working memory can be described as a multifaceted limited-capacity information processing system, comprising interrelated attention and memory subsystems that govern the controlled processing of goal-relevant information over short periods of time and in light of interference or distraction from goal-irrelevant information (Baddeley and Hitch, 1974; Baddeley, 1986, 2000; Baddeley and Logie, 1999). Complex or dual span tasks have been typically used to measure the processing "capacity" of working memory, quantifying the total "amount" of to-be-remembered information that can be accurately held in mind while resisting distraction from to-be-ignored information (Conway and Engle, 1994; Conway et al., 2005). Researchers have

consistently shown dual span task performance decreases as a function of increases in to-be-remembered and ignored information, supporting the hypothesis that working memory is limited capacity in nature. Although this work has provided strong evidence that working memory capacity is limited, little is yet understood about the effect that redundant information has on working memory processing capacity and efficiency. The current research used an extreme groups approach and a novel redundant memory probes (RMP) task to investigate (a) the extent to which the "efficiency" or "speed" of working memory visual-search processes were affected by redundancies in target and distractor information, and (b) whether such redundancy effects depend on individual differences in "capacity" or "amount" of working memory resources. Here, a simplified linear ballistic accumulator (LBA) model (Brown and Heathcote, 2008; Donkin et al., 2009) of RMP task accuracy and response time was used to characterize working memory efficiency, while working memory capacity was characterized by performance on standard dual span tasks.

The redundant-target paradigm has been commonly used to investigate the efficiency or workload capacity of visual-search processes in divided-attention and short-term memory. In such experiments, participants are presented with stimuli containing 2, 1, or 0 target features. The participant's task is to decide whether or not stimuli contain at least 1 target feature as quickly and as accurately as possible. Redundancy gain effects are demonstrated by decreases in reaction time (RT) performance under redundant-target conditions relative to single-target conditions, indicating increases in the amount of target information facilitates processing efficiency or workload capacity (e.g., Townsend and Eidels, 2011) or potentially statistical facilitation (Raab, 1962). Conversely, increases in RT performance under no-target or distractor conditions relative to all others indicates that increases in the amount of distractor information inhibits processing efficiency or workload capacity (e.g., Townsend and Eidels, 2011), or potentially statistical inhibition (cf. Townsend and Wenger, 2004).

This work has shown redundant target information facilitates speed, and in some cases the accuracy, of visual-search processes while distractor identification is inhibited because it is defined based on the conjunction of multiple properties. Although redundancy effects have been reliably shown in tasks that index divided attention or short-term memory processes, little work has been done to characterize redundancy effects in tasks designed to measure working memory processes. The present research assumed that if working memory governs the interaction between divided attention and short-term memory processes, then tasks that tap both processes index more general working memory resources. Following from this assumption, it was hypothesized that redundant target and distractor information presented during short-term memory search would yield classic redundancy gain and loss effects on decision-making accuracy and RT that can be attributed to facilitation and inhibition of working memory information processing efficiency or workload capacity

Recently, Eidels et al. (2010) used an LBA model to quantify the efficiency and workload capacity of cognitive processes

underlying redundant-target effects in a divided-attention experiment. Results showed that the LBA model was sensitive to the redundancy gain effects observed for choice accuracy and RT, such that model estimates of internal evidence accumulation or drift-rates showed greater efficiency in divided attention under redundant-target conditions relative to single-target conditions. Model simulations of participant drift-rate data also predicted individual differences in workload capacity as indicated by Townsend and colleagues' capacity coefficient (e.g., Townsend and Nozawa, 1995; Townsend and Wenger, 2004; Houpt and Townsend, 2012; Burns et al., 2013; Houpt et al., 2014) which characterized participant's divided attention as super, unlimited, or limited capacity. Crucially, results showed participants with larger differences between redundant-target and single-target drift-rates showed super capacity in divided attention, whereby redundant targets facilitated or increased the workload capacity of target recognition. In contrast, participants with smaller drift-rate differences tended to show limited capacity in divided attention, whereby redundant targets inhibited or decreased the workload capacity of target recognition. In sum, drift-rate efficiency and workload capacity measures showed convergent evidence that suggested individuals can differ in the magnitude of redundancy gain effects on divided attention, whereby some individuals show facilitation in processing efficiency, and others experience inhibition. The present research builds from this work by using the LBA model to (a) investigate redundancy gain and loss effects using a novel working memory experiment, and (b) determine the extent to which such effects differ between individuals classified as having low or high working memory capacity on dual span tasks.

In our current work, we deviate from the (Eidels et al., 2010) approach by using the average of the single conditions processing rates as the baseline for comparison to the dual conditions. The advantage to our approach was that it did not require additional complexity and model development beyond the standard LBA. The disadvantage of our approach compared to the Eidels et al. approach is that the baseline model does not match the traditional unlimited-capacity, independent parallel model baseline (cf. Townsend and Nozawa, 1995; Houpt et al., 2014); instead, our baseline is essentially a fixed-capacity coactive model. A fixed-capacity coactive model predicts the processing rates in the dual conditions will be the sum of one half the processing rates in the single conditions because in that model information regarding target presence or absence is summed across the two sources, but each process is only half as efficient due to spreading a fixed amount of resources across the sources (cf. Houpt and Townsend, 2011). While we do not have a strong argument for a fixed-capacity coactive baseline over an unlimited-capacity parallel model, our focus is not to determine whether individual participants exhibit super, unlimited, or limited workload capacity in the RMP task. Rather, our focus is on the extent to which redundancy effects in the RMP task vary as a function of individual differences in performance on other well-established working memory span tasks. This focus minimizes the issue of specifying a baseline model because redundancy effects are operationalized experimentally, as given by the magnitude of differences between performance

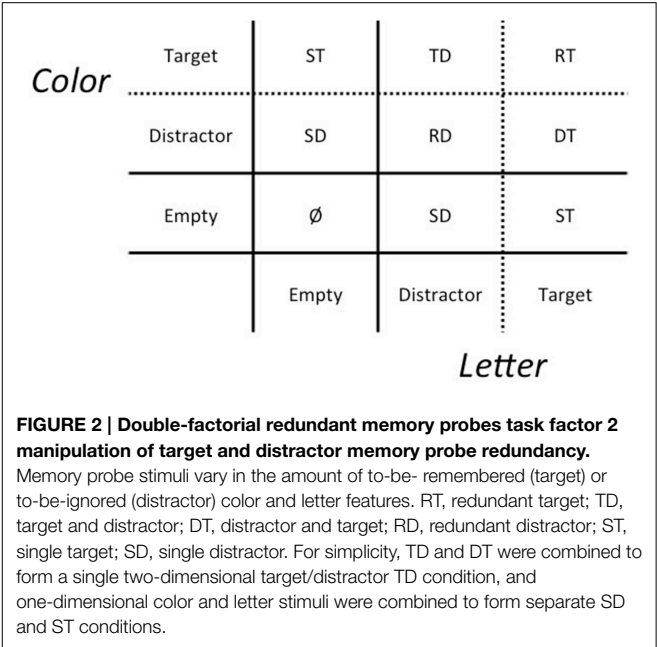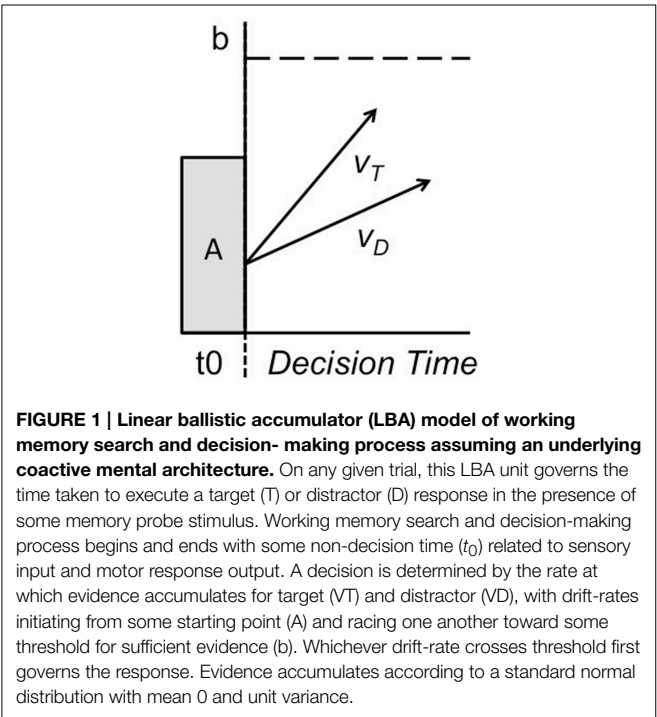indicators obtained under redundancy and singleton conditions.

As in **Figure 1**, the current LBA model had 5 parameters (t_0, A, b, v, and $s = 1$) that were assumed to govern the process of scanning short-term memory and deciding whether a given memory probe contained target (match) or distractor (non-match) information. Although alternative sequential sampling models are capable of characterizing RMP task performance (e.g., Ratcliff, 1978), these models tend to lead to similar conclusions (Donkin et al., 2011). The current LBA model used full RT distributions for correct and incorrect choices to estimate the rate at which evidence for target and distractor responses accumulate during the memory search process. A decision is made whenever the first accumulation process reaches an internal threshold criterion for sufficient evidence. In **Figure 1**, the b parameter represents the threshold of sufficient evidence for a response. High b-values reflect a preference for more information before making a decision. The A parameter represents the amount of evidence in each accumulator at the beginning of the trial. Higher values of A reflect a preference for responding fast. The t_0 parameter represents elements of the RT distribution that are not associated with the decision-making process, such as perceptual encoding or motor execution latencies. Higher values of t_0 reflect slower perceptual encoding and response execution. The v parameter represents the average rate of evidence accumulation for either the target (v_T) or distractor (v_D). High values of v reflect steeper or faster rates of evidence accumulation. The s parameter represents the standard deviation of the v parameter estimate, and is set to 1. Here, an accuracy adjusted drift rate, denoted (V), operationalized the process of accumulating accurate evidence for target and distractor decisions. The V measure was calculated by subtracting v obtained on incorrect trials from v on correct trials ($V = v_{correct} - v_{incorrect}$).

In terms of LBA parameters, our baseline prediction was formalized as $V_{RedundantProbe} = 0.5(V_{SingleProbe1} + V_{SingleProbe2})$. Specifically, redundancy effects were evaluated as the inequality resulting from contrasting V obtained under redundancy conditions vs. the V obtained under singleton conditions, e.g., $V_{RedundantTarget}$ versus $0.5(V_{ColorTarget} + V_{LetterTarget})$. Note that using a single information accumulator to represent information accumulation for the redundant probe trials, and assuming that drift rate is a linear combination of the drift rate of the single probe processes, implies a coactive (i.e., information pooling) process. The "fixed-capacity" comes from the fact that we scale the sum by 0.5, or one over the number of information sources, when we take the average of the single probe drift rates.

The LBA model output t_0, A, and b parameter values, along with 10 separate drift-rates, reflecting correct ($v_{correct}$) and incorrect ($v_{incorrect}$) evidence accumulation rates over each of the memory probe conditions (RT, ST, TD, RD, ST). Five accuracy adjusted drift-rates (V) were then derived by subtracting $v_{incorrect}$ from $v_{correct}$ for each condition separately, yielding the V_{RT}, V_{ST}, V_{TD}, V_{RD}, and V_{SD} values.

The present research investigated two main aims. The first was to examine the effects of redundancy on performance in a novel task designed to study the interaction between divided-attention and short-term memory processes in working memory, which we call the redundant memory probes (RMP) task. Illustrated in **Figure 2**, and described in greater detailed later, the RMP task systematically varied the amount of to-be-remembered (target) and to-be-ignored (distractor) information present during short-term memory search. Consistent with previous research, choice accuracy, mean response time (mRT), and LBA model drift-rate measures were used to quantify redundancy effects in the RMP task. Based on previous research, it was hypothesized

FIGURE 1 | Linear ballistic accumulator (LBA) model of working memory search and decision- making process assuming an underlying coactive mental architecture. On any given trial, this LBA unit governs the time taken to execute a target (T) or distractor (D) response in the presence of some memory probe stimulus. Working memory search and decision-making process begins and ends with some non-decision time (t_0) related to sensory input and motor response output. A decision is determined by the rate at which evidence accumulates for target (VT) and distractor (VD), with drift-rates initiating from some starting point (A) and racing one another toward some threshold for sufficient evidence (b). Whichever drift-rate crosses threshold first governs the response. Evidence accumulates according to a standard normal distribution with mean 0 and unit variance.

FIGURE 2 | Double-factorial redundant memory probes task factor 2 manipulation of target and distractor memory probe redundancy. Memory probe stimuli vary in the amount of to-be- remembered (target) or to-be-ignored (distractor) color and letter features. RT, redundant target; TD, target and distractor; DT, distractor and target; RD, redundant distractor; ST, single target; SD, single distractor. For simplicity, TD and DT were combined to form a single two-dimensional target/distractor TD condition, and one-dimensional color and letter stimuli were combined to form separate SD and ST conditions.

that a redundant-target (RT) condition would yield higher accuracy, faster mean reaction time (mRT), and larger LBA model drift-rates when contrasted against single-target (ST) conditions ($V_{RT} > V_{ST}$). A redundant-distractor condition also was hypothesized to yield lower accuracy, slower mRT, and smaller drift-rates when contrasted against the single-distractor (SD) condition ($V_{RD} < V_{SD}$). Mixed-target and distractor (TD and DT) conditions also were included to investigate the effects of overlapping target-distractor information on choice accuracy, mRT, and drift-rates, although we did not have any a priori predictions about the ordering of those drift rates relative to the other trial types (V_{TD}, V_{DT}?V_{ST}).

The second aim was to examine whether individuals classified as having high or low working memory capacity (WMC), as determined by performance on traditional dual span tasks, differed in the magnitude of redundancy gain and loss effects on the RMP task. This extreme groups approach was used to determine whether individuals who are known to differ on well-established measures of WMC also differ with regard to their sensitivity to redundancy gain and loss effects and overall efficiency in working memory visual search. Based on previous working memory individual differences research, it was hypothesized that individuals with low WMC would show lower accuracy, slower mRT, smaller drift-rates, and be more susceptible to distractor information while processing target information than those with high WMC. We also expected to find an interaction between experimentally driven redundancy effects and WMC individual differences. Specifically, we hypothesized that the magnitude of redundancy effects would depend on WMC individual differences, such that individuals with low WMC would show less redundancy gain and loss effects.

2. Materials and Methods

2.1. Participants
2.1.1. Sample Characteristics
The sample consisted of 170 young adults (96 men, 74 women; $\chi^2 = 2.85, p > 0.05$) ranging in age from 18 to 30 (mean = 20.89 ± 2.31). The sample was 77% White, 8% African American, 6% Asian, Indian, or Middle Eastern, 6% Hispanic or Latino, and 3% multiple ethnicities. Men were older than women [$t_{(168)} = 1.96, p < 0.05$]. However, gender was not associated with differences on any other study variable.

2.1.2. Study Recruitment
Participants were recruited from a subject pool of participants who completed a larger study on the personality, cognitive, and decision making correlates of substance use and antisocial behavior problems in young adults. Participants in the larger study were recruited using advertisements posted around the campus and surrounding community of a large Midwestern university. Advertisements were also placed in local and student newspapers. Advertisements were designed to attract individuals with varying degrees of lifetime problems with substance use and impulse control. This approach has been effective in attracting responses from individuals who vary in performance on cognitive tasks assessing intelligence, associative learning,

short-term memory, working memory, and approach-avoidance decision making (Finn et al., 2002, 2009; Endres et al., 2011, 2014).

Advertisement respondents were telephone screened for inclusion criteria of being between 18 and 30 years of age, able read/speak English, at least 6th grade education, and without a history of psychosis or head trauma. On the day of testing subjects were further screened to ensure participants did not use alcohol or drugs in the past 12 hours, were not experiencing symptoms of withdrawal or fatigue, and had a breath alcohol content of 0.0%.

Participants in the current sub-study were recruited based on a stratified random sample of main study participants ($N = 507$). Participants who completed the entire main study protocol were categorized as having low, moderate, or high histories of substance use and antisocial behavior based on an unsupervised cluster analysis of participant self-reported history with alcohol, drugs, childhood conduct problems, and adult antisocial behavior. A total of 180 participants (60 from each of the three groups) were solicited for participation in the present study with a final response rate of 94.44%. Based on previous research noting a negative association between executive cognitive functioning (e.g., intelligence, associative learning, and working memory) and individual's history of substance use and antisocial behavior (Finn et al., 2009), participants in the current stratified sample also were expected to vary greatly with respect to working memory and executive decision-making ability.

2.1.3. Dual Span Tasks
Working memory capacity (WMC) was assessed using two different complex-span tests, the Operation-Word Span test (OW; Conway and Engle, 1994) and a modified version of the Auditory Consonant Trigram test (AC; Brown, 1958; Finn et al., 2009; Endres et al., 2011). These tasks operationalize WMC as the total number of primary memory items that can be correctly recalled after performing a second unrelated cognitive task. The OW test was experimenter based and assessed the total number of words that were correctly recalled after performing a mathematical operation. For example, participants were asked to determine whether a mathematical operation was correct and presented with a word to-be-remembered ($2 \times 5 = 12$? DOG). After a series of operation-word trials, participants were asked to recall the words in there correct order of presentation in the series. The AC test also was experimenter based and assessed the total number of consonant letters, from a string of letters (e.g., r, d, t, and l), that could be remembered after counting backwards by 3's from a random three-digit number (e.g., 379) for a pre-determined length of time (e.g., 18 or 36 s). Several studies indicated that the OW and AC tests are valid indicators of the limited capacity nature of working memory, wherein accuracy decreases as a function of increases in primary memory items and secondary cognitive loads (Engle et al., 1999; Endres et al., 2011). Consistent with previous research, a composite WMC factor score was created by estimating the covariance among the total number of items correctly recalled on the OW and AC tasks using maximum likelihood extraction (Engle et al., 1999; Finn

et al., 2009; Endres et al., 2011). This WMC factor score variable was eventually dichotomized to reflect individual differences in high and low WMC in repeated measure analyses. Individuals were classified as having low or high WMC based on a median split (median = 0.03) of maximum likelihood estimated WMC factor scores (Cronbach's Alpha = 0.67, mean = 0, SD = 0.88, skew = −0.34, kurtosis = −0.36).

2.1.4. Redundant Memory Probe Tasks

The redundant memory probes (RMP) task was designed to study the interaction between divided- attention and short-term memory processes in working memory. The task used basic study-test (Sternberg, 1966) and varied response mapping (Schnieder and Shiffrin, 1977) procedures embedded within a double-factorial design Townsend and Wenger (2004) to examine the effects of redundant target and distractor information on the processes of searching short-term memory for color and letter information.

The study-test procedure (**Figure 3**) involved the initial rehearsal of memory lists varying in length and composition of color and letter items (Factor 1), followed by the serially matching of 16 memory-test probes with and without redundant target and distractor features (Factor 2). During the study phase, participants rehearsed memory lists containing either 1 or 3 color items and 1 or three letter items for a period of time lasting 1 s per memory list item. Memory lists were 2, 4, or 6 items in length, and there were 4 list types (1-color/1-letter,
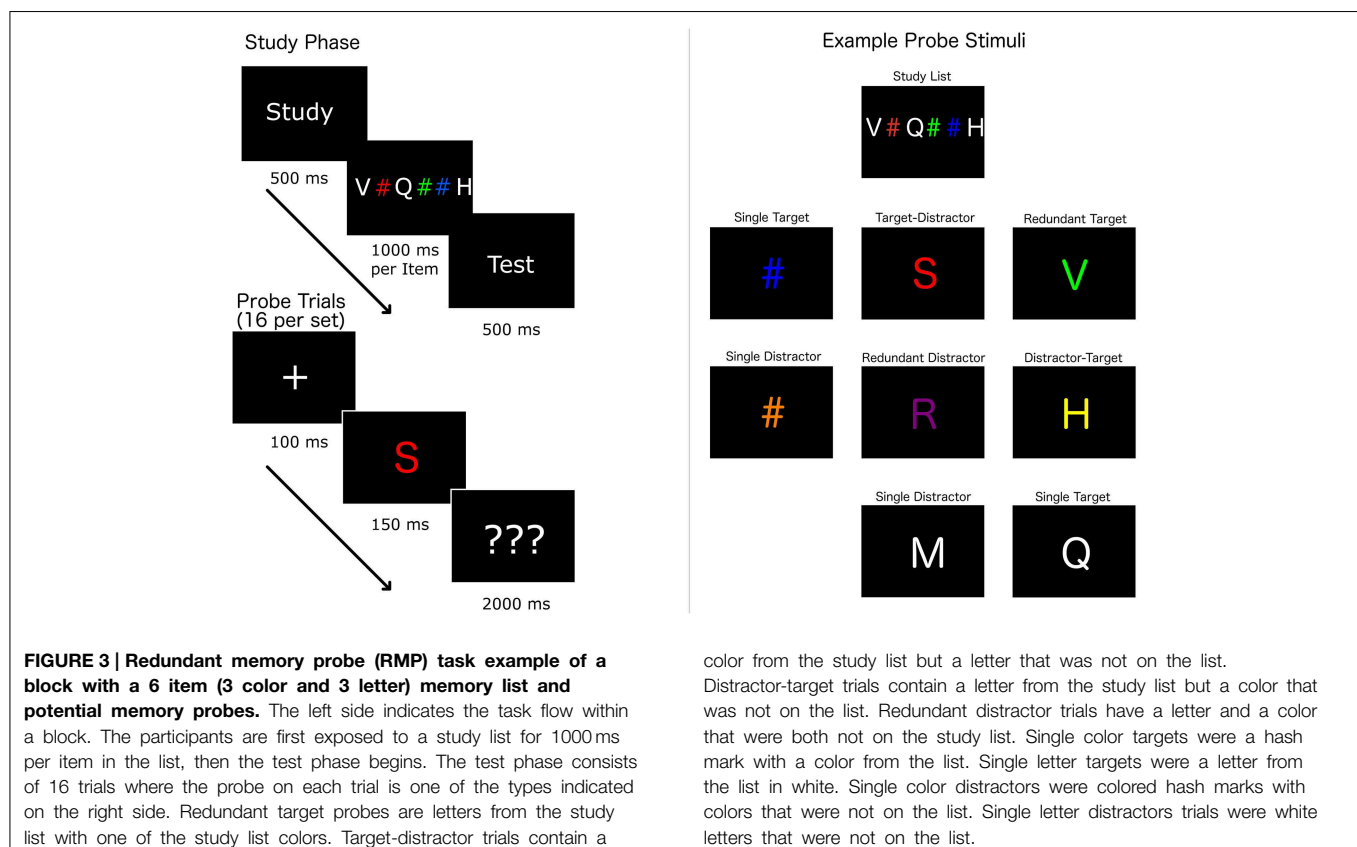
1-color/3-letter, 3-color/1-letter, and 3- color/3-letter) each with 6 different memory sets, totaling 24 lists in the task.

During the test phase, participants were briefly shown memory-test probes. Each probe was a single character. Probes that were colored (non-white) letters are referred to as dual probes. Probes that were either a white letter or a colored hash symbol are referred to as single probes. Probes could have 0, 1, or 2 target or distractor features. There were 8 probe types (**Figure 2**): redundant dual targets (RT) or distractors (RD), mixed color and letter dual targets and distractors (TD and DT), single color or letter targets (ST), and single color or letter distractors (SD).

Note that the participants were asked to say yes if either the color or letter of the probe was in the memory set. Hence, the dual probes to which the participants should have responded no (distractors) were defined by the conjunction of the color being outside of the memory set and the letter being outside of the memory set. The probes for which both color and letter were in the memory set had redundant target information. Memory test probes representing targets in a given study-test procedure could be distractors in other study-test sets (varied response mapping procedure), which was assumed to generate proactive interference.

2.1.5. Dependent Measures

Consistent with previous research, choice accuracy, mRT, and LBA model drift-rate estimates, which incorporates

FIGURE 3 | Redundant memory probe (RMP) task example of a block with a 6 item (3 color and 3 letter) memory list and potential memory probes. The left side indicates the task flow within a block. The participants are first exposed to a study list for 1000 ms per item in the list, then the test phase begins. The test phase consists of 16 trials where the probe on each trial is one of the types indicated on the right side. Redundant target probes are letters from the study list with one of the study list colors. Target-distractor trials contain a

color from the study list but a letter that was not on the list. Distractor-target trials contain a letter from the study list but a color that was not on the list. Redundant distractor trials have a letter and a color that were both not on the study list. Single color targets were a hash mark with a color from the list. Single letter targets were a letter from the list in white. Single color distractors were colored hash marks with colors that were not on the list. Single letter distractors trials were white letters that were not on the list.

both accuracy and RT information, were used to investigate redundancy effects on test-phase performance by contrasting RT and RD with ST and SD, respectively. Performance estimates were aggregated across Factor 1, study set size, because memory probe redundancies were manipulated during the test phase (Factor 2). As in **Figure 2**, performance estimates also were aggregated across the mixed TD and DT, as well as single target (ST) and single distractor (SD) test probe types, because the task was designed so that: (a) color and letter elements had equal a priori stimulus presentation probabilities across the 24 study lists and 8 test probe types, and (b) target-distractor discriminability was held constant for the different color and letter elements of study lists and test probes.

2.2. Data Analyses

Separate 2×2 repeated measures ANOVAs were used to examine the within-subjects effects of redundant information on RMP task performance measures as a function of between-subjects differences in WMC on dual span tasks. Based on previous research, the within-subjects factor in repeated measures analyses reflected planned comparisons for redundancy gain (RT vs. ST conditions), loss (RD vs. SD), and mixed (TD vs. ST) effects. Planned comparisons were conducted separately for gain, loss and mixed effects. Based on subject recruitment, the

dichotomized (median split) WMC factor score variable was used as the between-subjects factor in all repeated measures analyses. Analyses were conducted separately for choice accuracy (percent correct), mRT (on correct trials), and accuracy adjusted LBA drift-rate performance measures. Within-subjects and between-subjects effect sizes were examined with partial eta-square estimates.

3. Results

3.1. Descriptive Statistics

The low ($n = 85$) and high ($n = 85$) WMC groups did not differ in gender composition ($\chi^2 = 2.16, p > 0.05$) or average age [$t_{(168)} = 1.06, p > 0.05$]. However, groups did differ in average IQ [$t_{(167)} = -3.66, p < 0.001$] and years of education [$t_{(168)} = -3.66, p < 0.001$].

3.2. Individual LBA Model Fits

Model fit was examined by using subject's LBA model parameters to simulate accuracy and RT data, and then comparing these simulations to subject's actual accuracy and RT data. For example, **Figure 4** shows one subject's LBA model simulated defective cumulative density functions (CDF) plotted against that subject's actual defective CDFs. In **Figure 2**, LBA model simulated CDFs for correct and incorrect responses in RT, TD,
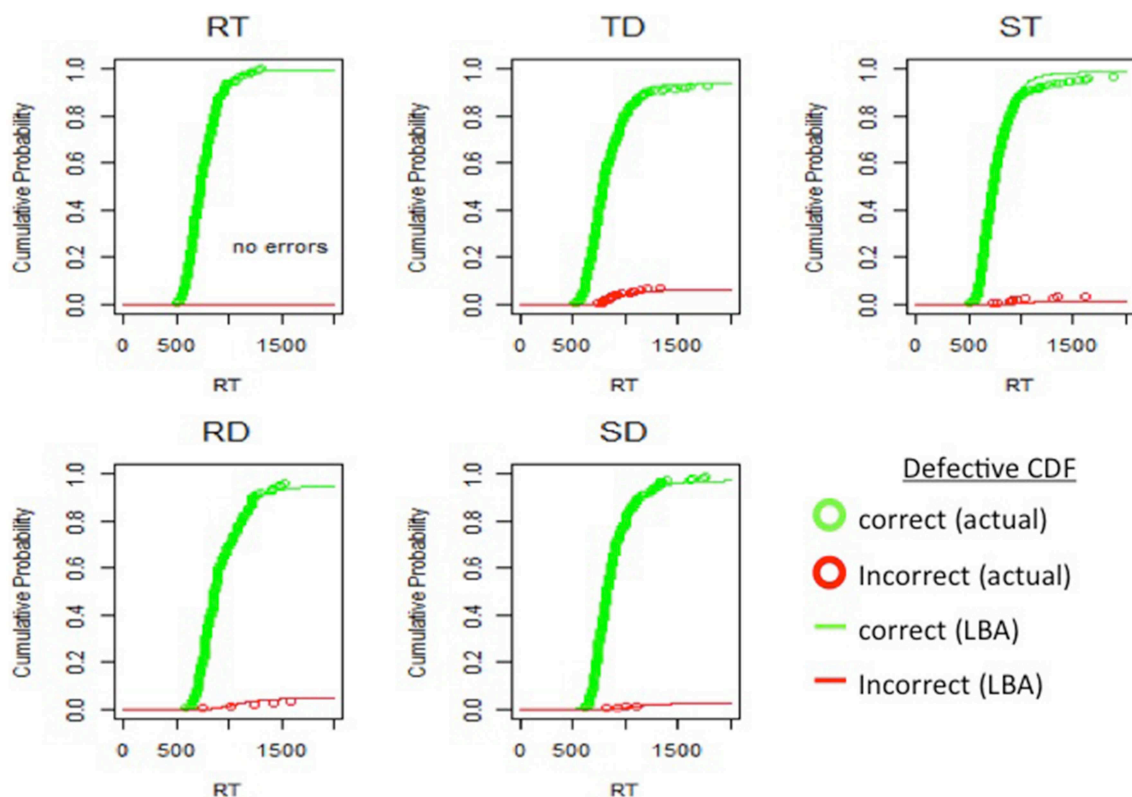
FIGURE 4 | Example subject's defective cumulative density functions illustrating the probability of observing correct (green font) and incorrect (red font) responses on or before some response time (RT). Subject's actual (open circles) data LBA simulated (lines) data plotted against each other for RT, redundant target; TD, mixed target/distractor; RD, redundant distractor; ST, single target; SD, single distractor conditions.

ST, RD, and SD test-probe conditions showed consistent overlap with actual CDFs collected in these respective conditions. The mean parameter value and standard deviation across participants is shown in **Table 1**.

3.3. Effects of WMC on LBA Model Non-Decision Time, Starting Point, and Threshold

No WMC group differences were found for LBA model parameters t_0 [$t_{(168)} = 0.67, p > 0.05$], A [$t_{(168)} = -0.16, p > 0.05$], or b [$t_{(168)} = -1.36, p > 0.05$]. For the High EMW capacity group, mean non-decision time, starting point, and threshold were $73.01 \pm 65.4, 7.30 \pm 1.28$ and 8.66 ± 0.22 respectively. For the low EMW capacity group, mean non-decision time, starting point, and threshold were $67.12 \pm 48.38, 7.33 \pm 1.28$, and 8.65 ± 0.23 respectively. These results suggest WMC individual differences are not involved in RMP task decision-making processes related to early perceptual coding and later response execution latencies, nor setting preferences for response types or sufficient evidence for responding.

3.4. Effects of Redundant Target Information and WMC on RMP Task Performance
3.4.1. Accuracy

Figure 5A, hit rates were facilitated by redundant-target information. These effects did not depend on WMC differences, even though those with high WMC were generally better at recognizing targets than those with low WMC. Within subjects tests showed target percent correct (PC) was higher for redundant color and letter targets, relative to single color targets

or single letter targets [RT > ST, $F_{(168)} = 7.14, p < 0.01$, partial $\eta^2 = 0.04$]. Between subjects tests showed those classified as high WMC had higher overall target PC than those classified as low WMC [$F_{(168)} = 6.67, p < 0.01, \eta_2 = 0.04$]. No interaction between redundant targets and WMC differences was found for target PC [$F_{(168)} = 0.38, p > 0.05, \eta_2 < 0.01$].
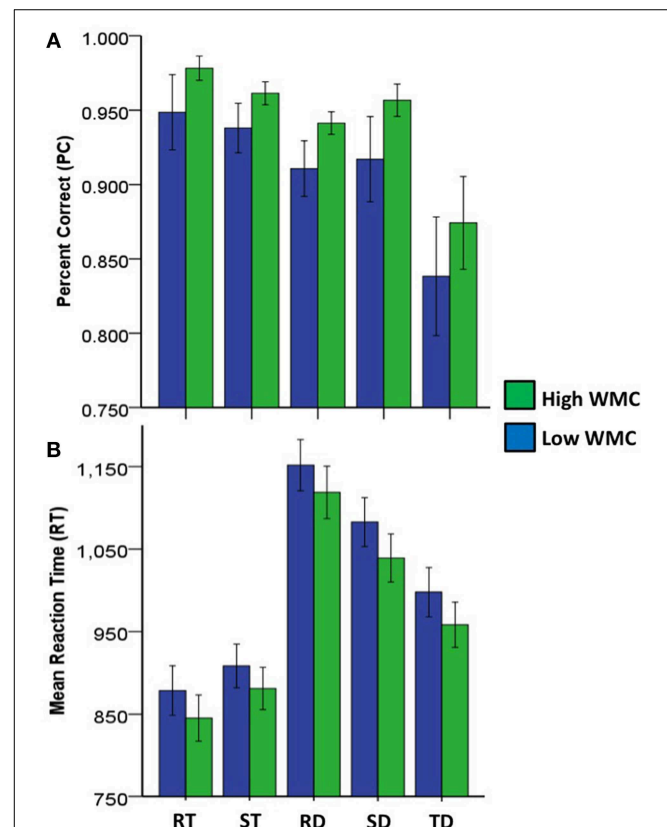
3.4.2. Correct Trials mRT

Figure 5B, shows mRT on for hits were facilitated by redundant target information, and these effects did not depend on WMC differences. Although those with high WMC tended to be faster at recognizing targets than those with low WMC, these differences did not reach statistical significance.

Within subjects tests showed mRT was shorter for redundant color and letter targets, relative to single color targets or single letter targets [RT < ST, $F_{(168)} = 116.65, p < 0.001$, partial $\eta^2 = 0.41$]. Between subjects tests showed those classified as high WMC did not differ in mRT from those classified as low WMC in overall mRT for targets [$F_{(168)} = 2.46, p > 0.05$, partial $\eta^2 = 0.01$]. No interaction between redundant targets and WMC differences was found for mRT [$F_{(168)} = 0.99, p > 0.05$, partial $\eta^2 = 0.01$].

TABLE 1 | Accuracy adjusted drift-rates by redundancy condition and percentile grouping.

Condition	Low 20% (n = 34)		High 80% (n = 34)	
	Mean	(SD)	Mean	(SD)
Non-decision time (t_0)	64.34	50.24	73.56	65.61
Starting point (A)	7.35	1.33	7.22	1.25
Threshold criterion (b)	8.67	0.25	8.58	0.22
Redundant target (V_{RT})	3.00	0.94	3.68	1.32
Single target (V_{ST})	2.63	0.74	3.09	0.63
Target and distractor (V_{TD})	1.96	0.93	2.33	0.83
Redundant distractor (V_{RD})	2.19	0.71	2.37	0.62
Single distractor (V_{SD})	2.25	0.73	2.80	0.82
	Low 50% (n = 85)		**High 50% (n = 85)**	
Non-decision time (t_0)	67.12	48.38	73.01	65.40
Starting point (A)	7.33	1.28	7.30	1.22
Threshold criterion (b)	8.65	0.23	8.60	0.22
Redundant target (V_{RT})	3.30	1.37	3.66	1.34
Single target (V_{ST})	2.86	0.78	3.11	0.64
Target and distractor (V_{TD})	1.95	1.08	2.28	0.89
Redundant distractor (V_{RD})	2.24	0.80	2.43	0.69
Single distractor (V_{SD})	2.39	0.87	2.69	0.71

FIGURE 5 | Bar graphs with 95% confidence intervals for mean accuracy (A) and response time (B) by redundancy condition and working memory capacity (WMC) groupings. RT, redundant target; TD, target and distractor; DT, distractor and target; RD, redundant distractor; ST, single target; SD, single distractor.

3.4.3. LBA Drift-Rates

Figure 6 shows accuracy adjusted drift-rates (V) were facilitated by redundant-target information; and, these effects did not depend on WMC differences, even though those with high WMC were generally more efficient in target recognition than those with low WMC. Within subjects tests showed V was larger for redundant color and letter targets, relative to single color targets or single letter targets [$V_{RT} > V_{ST}$, $F_{(168)} = 25.03$, $p < 0.001$, partial $\eta^2 = 0.13$]. Between subjects tests showed those classified as high WMC had larger overall V for targets than those classified as low WMC [$F_{(168)} = 5.41$, $p < 0.05$, partial $\eta^2 = 0.03$]. No interaction between redundant targets and WMC differences was found for V [$F_{(168)} = 0.36$, $p > 0.05$, partial $\eta^2 < 0.019$].

3.5. Effects of Redundant Distractor Information and WMC on RMP Task Performance
3.5.1. Accuracy

Figure 5A, shows redundant-distractor information had an inhibitory effect on correct rejection rates, but these effects did not reach statistical significance. However, those with high WMC were generally better at recognizing distractors than those with low WMC. Within subjects tests showed PC for redundant color and letter distractors was not significantly different from PC for single color distractors or single letter distractors [RT = ST, $F_{(168)} = 3.27$, $p > 0.05$, partial $\eta^2 = 0.02$]. Between subjects tests showed those classified as high WMC had higher distractor PC than those classified as low WMC [$F_{(168)} = 9.25$, $p < 0.01$, partial $\eta^2 = 0.05$]. No interaction between conjunctive distractors and WMC differences was found for PC [$F_{(168)} = 0.57$, $p > 0.05$, partial $\eta^2 < 0.01$].
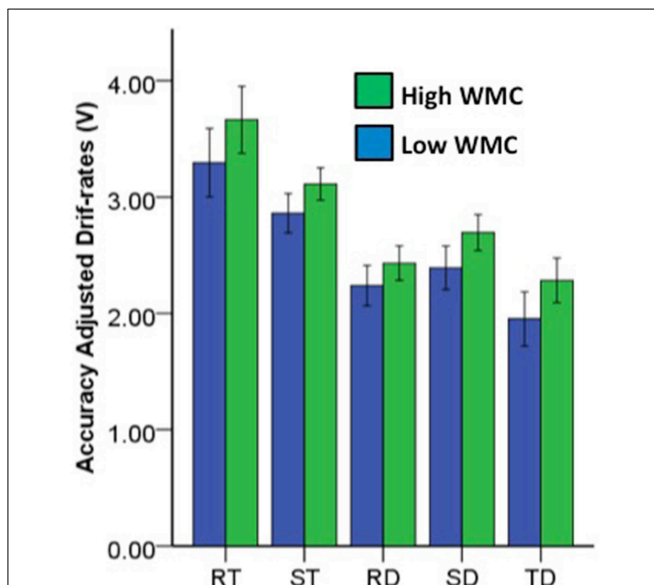
3.5.2. Correct Trials mRT

Figure 5B, shows mRT on correct trials was inhibited for redundant distractors, and these effects did not depend on WMC differences. Those with high WMC were generally faster at recognizing distractors than those with low WMC, but these effects did not reach statistical significance. Within subjects tests showed mRT was longer for redundant color and letter distractors, relative to single color distractors or single letter distractors [RD > SD, $F_{(168)} = 273.75$, $p < 0.001$, partial $\eta^2 = 0.62$]. Between subjects tests showed those classified as high WMC did not differ from those classified as low WMC in distractor mRT [$F_{(168)} = 3.26$, $p > 0.05$, $\eta^2 = 0.02$]. No interaction between conjunctive distractors and WMC differences was found for mRT [$F_{(168)} = 3.26$, $p > 0.05$, partial $\eta^2 < 0.01$].

3.5.3. LBA Drift-Rates

Figure 6 shows accuracy adjusted drift-rates (V) reduced for redundant-distractor information. These effects did not depend on WMC differences, even though those with high WMC were generally more efficient at recognizing distractors than those with low WMC. Within subjects tests showed V was smaller for redundant color and letter distractors, relative to single color distractors or single letter distractors [$V_{RD} < V_{SD}$, $F_{(168)} = 9.86$, $p < 0.01$, partial $\eta^2 = 0.06$]. Between subjects tests showed those classified as high WMC had larger overall V for distractors than those classified as low WMC [$F_{(168)} = 6.40$, $p < 0.05$, partial $\eta^2 = 0.04$]. No interaction between conjunctive distractors and WMC differences was found for V [$F_{(168)} = 0.69$, $p > 0.05$, partial $\eta_2 < 0.01$].

3.6. Effects of Mixed Target/Distractor Information and WMC on RMP Task Performance
3.6.1. Accuracy

Figure 5A, shows mixed target-distractor information had an inhibitory effect on hit rates, and these effects did not depend on WMC differences. Those with high WMC were better at recognizing targets while ignoring distractors than those with low WMC, but these effects did not reach statistical significance. Within subjects tests showed PC was lower for mixed color and letter targets and distractors, relative to single color targets or single letter targets [TD < ST, $F_{(168)} = 76.32$, $p < 0.001$, partial $\eta^2 = 0.31$]. Between subjects tests showed those classified as high WMC did not significantly differ from those classified as low WMC in PC for mixed color and letter targets and distractors [$F_{(168)} = 3.47$, $p > 0.05$, $\eta^2 = 0.02$]. No interaction between mixed color and letter targets and distractors and WMC differences was found for PC [$F_{(168)} = 0.34$, $p > 0.05$, partial $\eta^2 < 0.01$].

3.6.2. Correct Trials mRT

Figure 5B, shows mRT on correct trials was inhibited by mixed target-distractor information, and these effects did not depend on WMC differences. Those with high WMC were generally faster at recognizing targets while ignoring distractors than those with low WMC, but these effects did not reach statistical significance. Within subjects tests showed mRT was longer for mixed color

FIGURE 6 | Bar graphs with 95% confidence intervals for mean LBA model accuracy adjusted drift-rates by redundancy condition and working memory capacity (WMC) groupings. RT, redundant target; TD, target and distractor; DT, distractor and target; RD, redundant distractor; ST, single target; SD, single distractor.

and letter targets and distractors, relative to single color targets or single letter targets [TD > ST, $F_{(168)} = 513.49, p < 0.001$, partial $\eta^2 = 0.75$]. Between subjects tests showed those classified as high WMC did not differ from those classified as low WMC in mRT for mixed color and letter targets and distractors [$F_{(168)} = 3.05, p > 0.05, \eta^2 = 0.02$]. No interaction between mixed color and letter targets and distractors and WMC differences was found for mRT [$F_{(168)} = 2.74, p > 0.05, \eta^2 = 0.02$].

3.6.3. LBA drift-rates

Figure 6 shows accuracy adjusted drift-rates (V) were inhibited by mixed target-distractor information. These effects did not depend on WMC differences, even though those with high WMC were generally more efficient at recognizing targets while ignoring distractors than those with low WMC. Within subjects tests showed V was smaller for mixed color and letter targets and distractors, relative to single color targets or single letter targets [$V_{TD} < V_{ST}, F_{(168)} = 175.79, p < 0.001$, partial $\eta^2 = 0.51$]. Between subjects tests showed those classified as high WMC had larger V for mixed color and letter targets and distractors than those classified as low WMC [$F_{(168)} = 6.38, p < 0.05$, partial $\eta^2 = 0.04$]. No interaction between mixed color and letter targets and distractors and WMC differences was found for V [$F_{(168)} = 0.37, p > 0.05$, partial $\eta^2 < 0.01$].

3.7. Additional Analyses

To examine the stability of our findings, we conducted supplemental analyses using a more extreme percentile grouping criterion for dual span task WMC factor scores than a median split. As shown in Table 1, for adjusted drift rates, the direction and pattern of repeated measures effects did not differ by characterizing extreme (Low and High) WMC groups using a 20% and 80% (top) or using a 50% and 50% (bottom) percentile grouping. Regardless of 20/80 and 50/50 percentile grouping, results showed high WMC had larger drift-rates (V) than low WMC (i.e., main effect of group), but redundancy gain (RT vs. ST) and loss (RD vs. SD) did not depend on WMC individual differences (i.e., no group by redundancy condition interaction). Critically, both analyses show high EWM had larger drift-rates (V) than low EWM (i.e., main effect of group), but redundancy gain (RT vs. ST) and loss (RD vs. SD) effects did not depend on EWM capacity individual differences (i.e., no interaction between group and redundancy effects).

4. Discussion

The main findings of the present study were twofold. First, working memory visual-search processes were found to be both facilitated and inhibited under a novel redundant memory probes (RMP) task using accuracy, RT, and LBA measures of "how much" (i.e., capacity) and "how fast" (i.e., efficiency) information is processed. Second, although individuals classified as having high or low WMC with traditional dual span tasks differed in accuracy, RT, and rates of evidence accumulation on the RMP task, groups did not differ in the magnitude of facilitation (redundancy gain) and inhibition (redundancy loss) effects observed under the RMP task. When taken together, these results

suggest redundant information reliably facilitates and inhibits the efficiency or speed of working memory visual search, and these effects are independent of more general limits and individual differences in the capacity or space of working memory.

4.1. Redundancy Effects on Working Memory Visual Search

Consistent with previous research, results showed that memory probes with redundant-target features significantly improved or facilitated the accuracy and mean RT of working memory visual search relative to memory probes with only one target feature (i.e., redundancy gain). In contrast, results showed that memory probes with redundant-distractor features significantly reduced or inhibited the accuracy and mean RT of working memory visual search relative to memory probes with only one distractor feature (i.e., redundancy loss). Similarly, inhibition effects also were found for memory probes with mixed target and distractor features relative to memory probes with one distractor feature. These results also were confirmed with an LBA model of decision-making accuracy and RT that implicitly assumed a coactive mental architecture with fixed-capacity drove the rate or efficiency in which internal evidence accumulates (drift-rates) during working memory visual search. For this model, drift-rates were (i) larger (facilitated) for redundant target probes than for single target probes, (ii) smaller (inhibited) for redundant distractor probes than for single distractor probes, and (iii) smaller (inhibited) for mixed target and distractor probes than for single target probes.

In the context of Eidels et al. (2010)'s findings, the current evidence of redundancy gains in LBA model drift-rates suggest that the RMP task facilitated participant's workload efficiency to that of "super-capacity," such that increases in the amount of to-be-processed target information lead to an increase in the rate at which evidence accumulated during working memory visual-search process. This interpretation of the current findings is inconsistent with the dominant conceptualization of working memory processes being limited capacity in nature (Baddeley, 2000). Crucially, the expectation for limited capacity would be that of inhibition or a decrease in workload efficiency, such that redundant target conditions lead to reduced accuracy, RT, and drift-rates relative to single target conditions. Therefore, the limited-capacity assumption did not hold in the present study, because evidence of "super capacity" processing was found via significant redundancy gain effects. However, the limited-capacity assumption did hold under distractor probe conditions, such that accuracy, RT, and drift-rates where impeded when contrasting (i) redundant-distractor vs. single-distractor conditions, and (ii) mixed target/distractor conditions vs. single-target conditions (see **Figure 6**).

One explanation for the present findings could be that the locus of working memory limited capacity is specific to short-term memory processes, and not necessarily divided-attention processes. That is, perhaps domain-specific short-term memory space is limited in capacity and can hold only a certain amount of contents, while controlled divided-attention speed is not limited in efficiency or workload capacity and can be facilitated or inhibited by the stimulus-context. Toward this end, a key

limitation of the present research was that we did not take into account variability in performance as a function of variability in memory-set size (i.e., Factor 1). Specifically, RMP task memory lists were either 2, 4, or 6 items long, and thus, it could be that facilitation and inhibition effects on workload capacity during working memory visual search depend on memory list or set size. Future work with the RMP task should attempt to disentangle the interactive effects of memory set size (short-term memory) and memory probe redundancy (divided-attention search).

Another possible explanation for the present finding of "super capacity" processing under redundant-target conditions is that these effects were simply an artifact of implicitly selecting a fixed-capacity coactive process as a baseline for our LBA model. Perhaps fitting an LBA model that assumed a more conservative UCIP baseline would not yield evidence of facilitation. Therefore, the present findings are limited by questions concerning LBA model specification, and the exact configuration of mental processes driving performance in the RMP task. Future work with the RMP task might attempt to identify the best fitting baseline model at the individual subjects level, and/or use the standard UCIP model to determine the extent to which model derived differences in workload capacity (i.e., super, unlimited, or limited capacity classifications) correspond with differences in WMC on dual span tasks.

4.2. Working Memory Capacity Effects on Working Memory Visual Search

Consistent with previous research, results showed that individuals classified as high WMC on traditional dual span tasks had generally more accurate and faster RMP task performance than those classified as low WMC. These results also were confirmed with the LBA model of performance that indicated higher WMC was associated with higher drift-rates. Evidence of a link between WMC and RMP task drift-rates is consistent with previous research demonstrating that WMC individual differences are predicted by drift-rates obtained under other simple reaction time tasks (Schmemiedek et al., 2007). Our findings also could be interpreted to suggest that capacity and efficiency measurements of working memory processing could stem from the same underlying source of individual differences, such that greater working memory "capacity" or processing "space" is associated with greater working memory "efficiency" or processing "speed."

However, our results also suggest an important caveat in that redundancy gain and loss effects were not dependent on WMC. Specifically, both high and low WMC individuals showed comparable redundancy gains (facilitation) and losses (inhibition) effects in the RMP task. In fact, low and high WMC groups showed comparable evidence of "super-capacity" processing for redundant targets and "limited capacity" processing for mixed and redundant-distractors. This could be interpreted to mean that the efficiency with which individuals integrate information in working memory (i.e., workload capacity) may not depend on individual differences in working memory capacity or space limitations. However, it is important to point out that our sample recruitment and extreme groups approach may limit the generalizability of the present findings.

Mainly, the use of a dichotomized WMC variable and categorical analysis (i.e., repeated measures) method limited the statistically power of the current results. Perhaps other dimensional or factor analytic methods might reveal an interaction between WMC individual differences and redundancy effects. However, it is suspected that any potential interaction effects revealed by dimensional or factor analytic approaches would be weak at best, given that the current analyses did not reveal statistical trends in favor of rejecting the null hypothesis of an interaction between WMC differences and redundancy effects.

Finally, limitations in analytic approach notwithstanding, the results of the current study have broader implications for clinical research, because working memory impairments are known to characterize individuals with a history of substance use and antisocial behavior (Finn et al., 2009; Endres et al., 2011, 2014). Current results using the extreme group approach revealed that individuals with low WMC showed poorer RMP task performance than those with high WMC. Indeed, these effects could be largely due to clinical problems, given that individuals with low WMC also tend to have a greater history of chronic, severe, and co-occurring substance abuse and antisocial behavior than those with high WMC. In this regard, another study limitation was that participants were recruited based on individual differences in clinical history, but such individual differences were not included as covariates in repeated measures analyses. Perhaps redundancy gain and loss effects are more or less apparent in those with a history of substance use and antisocial behavior. This has important clinical implications because, to the extent that the RMP task could be used to disentangle the interaction between working memory subsystems, it would be interesting to know whether the source of working memory impairments stems from deficits in divided attention, short-term memory, or both. To our knowledge, research has yet to identify the exact psychological processes and mechanisms driving working memory impairments in substance use and antisocial behavior. It is also unclear whether individuals with such conditions are more or less sensitive to redundancy information in working memory tasks. Such knowledge and specificity could provide valuable information to emerging treatment models for substance use and antisocial behavior problems that utilize working memory training or remediation as a means to improve self-regulation and impulse control. Future research with the RMP task should examine the effects of individual differences in externalizing disorders on performance, and attempt to uncover the latent psychological mechanisms driving the known working memory impairments associated with this condition.

4.3. Linear Ballistic Accumulator Model of the Redundant Memory Probes Task

Lastly, results from the current study added to the growing body of research applying quantitative modeling approaches to the study of individual differences (Neufeld et al., 2002; Yechiam et al., 2005; Johnson et al., 2010; Endres et al., 2011, 2014). Here, evidence showed that measures of performance accuracy and RT we not always sensitive to differences in RMP task condition and dual span task related WMC. Specifically, for the

3 possible RMP task effects: RT vs. ST, RD vs. SD, and TD vs. ST, the accuracy (percent correct) measure detected 2 of 3, the RT (mean) measure detected 2 of 3, and the LBA drift-rates (accuracy adjusted) measure detected 3 of 3. For the 3 group effects that were possible for each RMP task effect, the accuracy (percent correct) measure detected 2 of 3, the RT (mean) measure detected 0 of 3, and the LBA drift-rates (accuracy adjusted) measure detected 3 of 3. There were no significant interaction effects between task and group for any of the 3 contrasts. These comparisons could be interpreted to mean that LBA model drift-rates were more psychometrically reliable than accuracy and RT measures, showing the greatest sensitivity to task and group main effects, while being equally selective at ruling out task by group interactions. However, it is important to note that a key limitation with the current LBA model was its specification. Specifically,

we implicitly assumed that a fixed-capacity, coactive mental architecture drove visual search processes for all subjects, rather than taking steps to identify exactly which mental architecture was driving visual-search processes in the RMP task. Future quantitative modeling work should investigate this issue of model specification and identify whether RMP visual search is best represented by a coactive, parallel or serial mental architecture.

Acknowledgments

References

Baddeley, A. D. (1986). *Working Memory*. New York, NY: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A. D., and Hitch, G. (1974). "Working memory," in *The Psychology of Learning and Motivation*, ed G. Bower (New York, NY: Academic Press), 47–89.

Baddeley, A. D., and Logie, R. (1999). "Working memory: the multi-component model," in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, eds A. Mijake and P. Shah (Cambridge, MA: MIT University Press), 28–61.

Brown, J. (1958). Some tests of the decay of immediate memory. *Q. J. Exp. Psychol.* 10, 12–21. doi: 10.1080/17470215808416249

Brown, S. D., and Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cogn. Psychol.* 57, 153–178. doi: 10.1016/j.cogpsych.2007.12.002

Burns, D., Houpt, J., Townsend, J., and Endres, M. (2013). Functional principal components analysis of workload capacity functions. *Behav. Res. Methods* 45, 1048–1057. doi: 10.3758/s13428-013-0333-2

Conway, A. R. A., and Engle, R. W. (1994). Working memory and retrieval: a resource-dependent inhibition model. *J. Exp. Psychol.* 123, 354–373. doi: 10.1037/0096-3445.123.4.354

Conway, A. R. A., Kane, M. J., Bunting, B. F., Hambrick, Z. D., Wilhelm, O., and Engle, R. W. (2005). Working memory span tasks: a methodological review and users guide. *Psychon. Bull. Rev.* 12, 769–786. doi: 10.3758/BF03196772

Donkin, C., Averell, L., Brown, S., and Heathcote, A. (2009). Getting more from accuracy and response time data: methods for fitting the linear ballistic accumulator. *Behav. Res. Methods* 4, 1095–1110. doi: 10.3758/BRM.41.4.1095

Donkin, C., Brown, S., Heathcote, A., and Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: different models for response time, same conclusions about psychological mechanisms? *Psychon. Bull. Rev.* 18, 61–69. doi: 10.3758/s13423-010-0022-4

Eidels, A., Donkin, C., Brown, S., and Heathcote, A. (2010). Converging measures of workload capacity. *Pychon. Bull. Rev.* 17, 763–771. doi: 10.3758/PBR.17.6.763

Endres, M., Donkin, C., and Finn, P. (2014). An information processing/associative learning account of behavioral disinhibition in externalizing psychopathology. *Exp. Clin. Psychopathol.* 22, 122–132. doi: 10.1037/a0035166

Endres, M. J., Rickert, M. E., Bogg, T., Lucas, J., and Finn, P. R. (2011). Externalizing psychopathology and behavioral disinhibition: working memory mediates signal discriminability and reinforcement moderates response bias in approach-avoidance learning. *J. Abnorm. Psychol.* 120, 336–351. doi: 10.1037/a0022501

Engle, R., Tuholski, S., Laughlin, J., and Conway, A. R. (1999). Working memory, short- term memory, and general fluid intelligence: a latent variable approach. *J. Exp. Psychol.* 128, 309–331. doi: 10.1037/0096-3445.128.3.309

Finn, P., Miller, M., Rickert, M. E., Lucas, J., Bogg, T., Bobova, L., et al. (2009). Reduced cognitive ability in alcohol dependence: examining the role of covarying externalizing psychopathology. *J. Abnorm. Psychol.* 188, 100–116. doi: 10.1037/a0014656

Finn, P. R., Mazas, C. A., Justus, A. N., and Steinmetz, J. E. (2002). Early-onset alcoholism with conduct disorder: go / no go learning deficits, working memory capacity, and personality. *Alcohol. Clin. Exp. Res.* 19, 148–157. doi: 10.1111/j.1530-0277.2002.tb02524.x

Houpt, J., Blaha, L., McIntire, J., Havig, P., and Townsend, J. (2014). Systems factorial technology with r. *Behav. Res. Methods* 46, 307–330. doi: 10.3758/s13428-013-0377-3

Houpt, J., and Townsend, J. (2011). An extension of sic predictions to the wiener coactive model. *J. Math. Psychol.* 55, 267–270. doi: 10.1016/j.jmp.2011.02.002

Houpt, J., and Townsend, J. (2012). Statistical measures for workload capacity analysis. *J. Math. Psychol.* 56, 341–355. doi: 10.1016/j.jmp.2012.05.004

Johnson, S. A., Blaha, L., Houpt, J., and Townsend, J. T. (2010). Information processing in autism spectrum disorders. *J. Math. Psychol.* 54, 53–72. doi: 10.1016/j.jmp.2009.06.006

Neufeld, R. W. J., Vollick, D., Carter, J. R., Boksman, K., and Jette, J. (2002). Application of stochastic modeling to the assessment of group and individual differences in cognitive functioning. *Psychol. Assess.* 14, 279–298. doi: 10.1037/1040-3590.14.3.279

Raab, D. (1962). Statistical facilitation of simple reaction times. *Trans. N.Y. Acad. Sci.* 24, 574–590. doi: 10.1111/j.2164-0947.1962.tb01433.x

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85:59. doi: 10.1037/0033-295X.85.2.59

Schmemiedek, F., Oberauer, K., Wilhelm, O., Süß, H., and Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *J. Exp. Psychol. Gen.* 136, 414–429. doi: 10.1037/0096-3445.136.3.414

Schnieder, W., and Shiffrin, R. (1977). Controlled and automatic human information processing: detection, search, and attention. *Psychol. Rev.* 84, 1–66. doi: 10.1037/0033-295X.84.1.1

Sternberg, S. (1966). High-speed scanning in human memory. *Science* 153, 652–654. doi: 10.1126/science.153.3736.652

Townsend, J., and Eidels, A. (2011). Workload capacity spaces: a unified methodology for response time measures of efficiency as workload is varied. *Psychon. Bull. Rev.* 18, 659–681. doi: 10.3758/s13423-011-0106-9

Townsend, J. T., and Nozawa, G. (1995). Spatiotemporal properties of elementary perception: an investigation of parallel, serial, and coactive processing. *J. Math. Psychol.* 39, 321–359. doi: 10.1006/jmps.1995.1033

Townsend, J. T., and Wenger, M. J. (2004). A theory of interactive parallel processing: new capacity measures and predictions for a

response time inequality series. *Psychol. Rev.* 111, 1003–1035. doi: 10.1037/0033-295X.111.4.1003

Yechiam, E., Busemeyer, J. R., Stout, J. C., and Bechara, A. (2005). Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychol. Sci.* 16, 973–978. doi: 10.1111/j.1467-9280.2005.01646.x

Conflict of Interest Statement: The Guest Associate Editor Cheng-Ta Yang declares that, despite having collaborated with author Joseph W Houpt, the review process was handled objectively and no conflict of interest exists. 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.

Dyslexia and configural perception of character sequences

Joseph W. Houpt [1,2]*, Bethany L. Sussman [2], James T. Townsend [2] and
Sharlene D. Newman [2]

[1] Department of Psychology, Wright State University, Dayton, OH, USA, [2] Department of Psychological and Brain Sciences,
Indiana University, Bloomington, IN, USA

Developmental dyslexia is a complex and heterogeneous disorder characterized by
unexpected difficulty in learning to read. Although it is considered to be biologically based,
the degree of variation has made the nature and locus of dyslexia difficult to ascertain.
Hypotheses regarding the cause have ranged from low-level perceptual deficits to higher
order cognitive deficits, such as phonological processing and visual-spatial attention.
We applied the capacity coefficient, a measure obtained from a mathematical cognitive
model of response times to measure how efficiently participants processed different
classes of stimuli. The capacity coefficient was used to test the extent to which
individuals with dyslexia can be distinguished from normal reading individuals based on
their ability to take advantage of word, pronounceable non-word, consonant sequence
or unfamiliar context when categorizing character strings. Within subject variability of
the capacity coefficient across character string types was fairly regular across normal
reading adults and consistent with a previous study of word perception with the capacity
coefficient—words and pseudowords were processed at super-capacity and unfamiliar
characters strings at limited-capacity. Two distinct patterns were observed in individuals
with dyslexia. One group had a profile similar to the normal reading adults while the
other group showed very little variation in capacity across string-type. It is possible
that these individuals used a similar strategy for all four string-types and were able to
generalize this strategy when processing unfamiliar characters. This difference across
dyslexia groups may be used to identify sub-types of the disorder and suggest significant
differences in word level processing among these subtypes. Therefore, this approach
may be useful in further delineating among types of dyslexia, which in turn may lead to
better understanding of the etiologies of dyslexia.

Keywords: capacity, dyslexia, configural processing, word superiority effect, individual differences

1. Introduction

Developmental dyslexia is a neurobiologically based, lifelong learning disability that specifically
affects the ability to read skillfully and is estimated to be present in 5–17.5% of children (Shay-
witz, 1998). Reading deficits in dyslexia are considered unexpected and independent of factors
such as intelligence and opportunity (see however Stanovich, 1996). There is no consensus on
the etiology or core deficit in dyslexia and several theories have been proposed. It is generally
associated with deficits in spelling, phonological/orthographical processing, rapid auditory pro-
cessing, and short-term verbal memory (Ramus, 2003; Shaywitz and Shaywitz, 2005). Dyslexia has

also been linked to other more domain general impairments such as automaticity (Nicolson and Fawcett, 2011), magnocellular functioning (Stein, 2001), and temporal auditory processing (Tallal, 1980). While phonological awareness has remained the most consistent explanatory marker (Ramus, 2003) of dyslexia, the cause of phonological impairment remains controversial. Dyslexia is often diagnosed in childhood and many dyslexic readers may build reading proficiency in adolescence and adulthood, however, reading often remains slow and effortful and there remains a phonological processing deficit (Wilson and Lesaux, 2001; Shaywitz and Shaywitz, 2005).

1.1. The Word Superiority Effect and Dyslexia

From the early days of experimental psychology, researchers have noted that normal reading adults are better at perceiving letters in the context of a word than alone or in random sequences (e.g., Cattell, 1886). Even when the informativeness of a word context is eliminated through careful experimental control (Reicher, 1969; Wheeler, 1970) normal reading adults perform better with a word context. The pervasive advantage is frequently referred to as the word superiority effect. The word superiority effect is a classical example of a configural superiority effect (cf. Pomerantz et al., 1977), but there is still some uncertainty as to the nature of the context advantage. Possible explanations have ranged from holistic processing of the word form (e.g., Healy, 1994) to independent processing of letters with some correction of letter level errors based on word level properties (e.g., Massaro, 1973; Pelli et al., 2003). Given that there is argument about the presence of a superiority effect, we focus on the degree of superiority rather than the locus of the superiority effect in this paper.

Given the robustness of the word superiority effect, one might inquire as to whether the effect is intact among individuals with developmental dyslexia. With dyslexia, reading is a generally slower and more effortful process. Potential loci of the reading deficit range from sub-word level, such as letter-phoneme correspondence (e.g., Blau et al., 2009; Blomert, 2011), to sentence level syntactic deficits. Tests of word superiority isolate one attribute of reading performance, and the extent to which individuals with dyslexia have a reduced or absent word superiority effect may be informative as to the nature of their deficits. Likewise, variation in the word superiority effect when comparing those with dyslexia and controls may also inform our understanding of the nature of the word superiority effect in the normal reading population.

Although research on dyslexia and the word superiority effect is limited, Grainger et al. (2003) have compared children with dyslexia and reading-age matched controls on the Reicher-Wheeler task (the standard paradigm for measuring the word superiority effect). Despite clear differences between the groups in ability to pronounce pseudowords, both groups were significantly better at identifying letters in the context of a word than in a non-word. The magnitude of the difference between words and non-words was nearly the same in both groups, and, if anything, slightly larger in the dyslexia group. This same basic effect was replicated by Ziegler et al. (2008), although they found statistically significant superiority effects in only response times, not accuracy.

Since the original demonstrations of the word superiority effect, researchers have also shown a pseudoword superiority effect: letters are more easily identified in pronounceable non-words (henceforth referred to as pseudowords to distinguish from unpronounceable non-words) than letters alone (e.g., McClelland and Johnston, 1977) or letters in non-word contexts (e.g., Baron and Thurston, 1973; Spoehr and Smith, 1975). Given that difficulty pronouncing pseudowords is one of the identifying characteristics of developmental dyslexia (for review, see Rack et al., 1992), one might predict that there would be a more dramatic difference between those with dyslexia and controls in the magnitude of a pseudoword superiority effect. Nonetheless, Grainger et al. (2003) also found no difference between groups on the pseudoword superiority effect: The effect was present in both the children with dyslexia and the reading-age matched controls and the magnitude was roughly the same in both groups. Hence, any explanation of the differing ability to pronounce pseudowords cannot depend solely on processes involved in the pseudoword superiority effect. In particular, Grainger et al. claim that this finding rules out the common explanation of dyslexia as a deficit in letter (or letter clusters) to phoneme translation.

A third finding in the Grainger et al. work was that, with both dyslexic and control groups of children, there was no difference in the magnitude of the word superiority effect and of the pseudoword superiority effect. That is, the increase in performance for letters in words over letters in isolation was roughly the same size as the increase in performance for letters in pseudowords over letters in isolation. In contrast, the normal-reading adults in their study had a larger advantage for word context compared to pseudoword context, a difference that has been found in many other studies (Manelis, 1974; McClelland and Johnston, 1977; Estes and Brunn, 1987; Jacobs and Grainger, 1994).

Houpt et al. (2014) recently demonstrated a new approach to measuring the word superiority effect based on response times to whole letter strings rather than accuracy of single letter identification. Their approach is based on a comparison of an individuals response latency to a full string, such as a word or pseudoword, to his predicted response time if he had identified each letter independently and in parallel. This method has multiple potential advantages for studying word superiority among those with dyslexia. First, it is an individualized measure so we can study both differences across groups as well as heterogeneity within those with dyslexia. Second, even though compensated dyslexic adults may increase word recognition and accuracy, reading is often still less automatic, fluid, and fast (Lefly and Pennington, 1991; Shaywitz et al., 1999), so the fact that the Houpt et al. approach is based on response times may make it more likely to pick up on differences between the groups. Finally, it is a model based approach, so the results can inform models of word perception by both normal-reading adults and those with dyslexia.

The main statistic used by Houpt et al. (2014) was the capacity coefficient (Townsend and Nozawa, 1995; Townsend and Wenger, 2004; Houpt and Townsend, 2012), which uses the cumulative reverse hazard function of the response times to predict hypothetical independent, parallel performance and compare it to participants actual performance. For more details see

(Houpt et al., 2014). For each participant the cumulative reverse hazard function is estimated from single character conditions by the sum over all response times less than a given time of 1/number of response times less than or equal to t, i.e.,

$$K(t) = 1/n \sum 1/Y(t).$$

The independent parallel model prediction for a participant is given by summing the cumulative reverse hazard functions over each of the characters (Townsend and Wenger, 2004; Houpt et al., 2013). The participants actual performance with words (or pseudowords, etc.) is then compared to the predicted independent, parallel performance to get a measure of the degree of the advantage or disadvantage of the context.

$$C(t) = K_{Letter1} + K_{Letter2} + K_{Letter3} + K_{Letter4} - K_{word}$$

When the capacity coefficient is positive, indicating participants performed better with context, it is referred to as super-capacity. If the capacity coefficient is negative, which occurs if participants perform worse, it is referred to as limited capacity. Finally, if their performance is approximately equal to the predicted independent parallel model, it is referred to as unlimited-capacity.

The participants reported in Houpt et al. (2014), who had no reported reading difficulties, were nearly all super-capacity with words and pseudowords, while they tended to be limited-capacity with unpronounceable non-words and were nearly all limited capacity with upside-down, unpronounceable, non-words and unfamiliar characters (Katakana). They found that words and pseudowords were higher capacity than the other string-types. However, unlike the larger advantage for words over pseudowords normally reported (including for adults in Grainger et al., 2003), they only found higher capacity for words compared to pseudowords when the stimuli were not masked.

There are multiple potential outcomes to applying the capacity approach to analyzing dyslexia. If the time based measures follow the accuracy based results of Grainger et al., then we would expect to see super-capacity for words and pseudowords and unlimited or limited capacity for non-words for both dyslexic and control participants. With normal reading adults, we would also expect to see higher capacity with words than with pseudowords, although this prediction is less certain given that Houpt et al. only found the difference in capacity in one of their two experiments. If the deficits present in dyslexia are specific to word perception speed, but not accuracy, then we would expect word and pseudoword capacity to be unlimited or limited, more on par with non-word capacity. We would also predict that the participants with dyslexia would have generally lower capacity with words and pseudowords than the control group.

2. Method

To measure the cumulative hazard function for responses to strings, we had a block of trials dedicated to each string type in which the same target and distractors were used. Targets were all four character strings: "care" for the word blocks, "lerb" for the pseudoword blocks, "rlkf" for the non-word blocks and

"サイクオ" for the unfamiliar character blocks. For each target, a set of four distractors was chosen that was within the same category, e.g., all of the distractors for the word-target block were also words. Each distractor was created by changing a single character in the target string, with one distractor for a change in each character position, e.g., for the target "rlkf," the distractors were "vlkf," "rtkf," "rlhf," and "rljk." This is essentially the same task as Houpt et al. (2014).

To measure the cumulative hazard function for characters in isolation, we had blocks of trials in which participants needed to discriminate between each of the two possible characters in each position. For example, because "vlkf" was a distractor for the target "rlkf," we had a block of trials during which the participants were required to distinguish between "v" and "r" in isolation. The full set of stimuli we used are shown in **Table 1**.

2.1. Participants

Participants were 19 students (Mean age = 21; 15 female) recruited from the Indiana University community. 11 participants had a formal dyslexia diagnosis and one dyslexia participant was left handed. Two of the participants with dyslexia (both Male) were dropped from the analyses because they did not complete 2 days of each of the experimental sessions. All control participants had no history of neurological conditions. All participants provided written informed consent, as approved by the Institutional Review Board of Indiana University, Bloomington. The participants completed a battery of tests to measure cognitive performance. They completed the Wechsler Abbreviated Scale of Intelligence (WASI; Weschler, 1999), Word Attack (pseudoword naming) from the Woodcock-Johnson III tests of Achievement (Woodcock et al., 2001), the Edinburgh Handedness Questionnaire (Oldfield, 1971), Dyslexia Checklist (Vinegrad, 1994), and the Adult Reading History Questionnaire (Lefly and Pennington, 2000). As shown in **Table 2**, the groups did not differ on intelligence measurements, but did differ on measures of phonological processing and verbal working memory. Also, although all but one participant reported being right handed, the groups differed in degree of handedness with the dyslexics having a weaker absolute handedness measure.

Groups did not differ in age or intelligence measures. On average, verbal IQ was higher than non-verbal IQ ($M = 7.26$, $SD = 9.63$, $p < 0.005$), but this did not differ by group.

2.2. Stimuli

Table 1 gives the complete list of stimuli used for both the single character and exhaustive trials for each type, which are a subset of the stimuli in Houpt et al. (2014). There were four categories of stimuli used: words, pronounceable non-words (pseudowords), unpronounceable non-words and strings of Katakana characters. All strings used were four characters long. Word frequency counts (based on Kucera and Francis, 1967) are available in the appendix of Houpt et al. (2014). Pseudowords were taken from the ARC Non-word Database (Rastle et al., 2002). The neighborhood size and summed frequency of the neighbors for each of the pseudowords are also included in the appendix of Houpt et al. (2014). Strings and characters were presented in black Courier font on a gray background. Characters were approximately 0.33°

TABLE 1 | Full set of character sequences used for stimuli.

	Target		Distractors			Single character						
Word	care	bare	cure	cave	card	c	b	a	u	r	v	e	d
Pseudoword	lerb	nerb	larb	lemb	lerf	l	n	e	a	r	m	b	f
Non-word	rlkf	vlkf	rtkf	rlhf	rljk	r	v	l	t	k	h	f	k
Katakana	サイクオ	ヘイクオ	サナクオ	サイフオ	サイクノ	サ	ヘ	イ	ナ	ク	フ	オ	ノ

TABLE 2 | Descriptive measures of participant groups.

	Control (n = 8)		Dyslexia (n = 9)		t	BF
	M	SD	M	SD		
Age	21.5	(2)	21.3	(1.3)	0.20	0.43
WASI verbal	113	(12.6)	118	(10.9)	−0.87	0.55
WASI non-verbal	105.4	(5.5)	110.7	(10.3)	−1.34	0.73
WASI full	110.5	(9.6)	115.6	(11.3)	−1.00	0.58
Verbal - Non-verbal	7.6	(9.2)	7.33	(9.34)	0.06	0.42
Handedness	81.8	(17.6)	57.9	(19.9)	2.63	3.23
Dyslexia checklist	4	(3.3)	14.9	(3.3)	−6.79	2120
Reading history	29.9	(11.2)	67.3	(10.3)	−7.14	3650
Reading span	3.1	(0.7)	2.1	(0.22)	3.55	17.4
Word attack (GE)	15.26	(3.7)	7.36	(1.95)	4.92	159

BF refers to the Bayes Factor comparing a model in which there is a difference between groups to a model in which there is no difference between groups. BF larger than 1 indicates evidence in favor of a difference with > 3.2 considered substantial evidence, > 10 strong and > 100 decisive. BF below 1 indicates evidence in favor of no difference between groups (< 0.31 substantial, < 0.1 strong, < 0.01 decisive).

horizontally and between $0.30°$ and $0.45°$ vertically. Strings were about $1.5°$ horizontally.

2.3. Procedure

All experimental conditions were run using Presentation® software version 14.9 (www.neurobs.com). Stimuli were presented on a 17″ Dell CRT monitor running in 1280 × 1024 mode. Participants used a two-button mouse for their responses. Participants were paid $8 per session, and received a $20 bonus upon completion of all 10 sessions. Each session lasted between 45 and 60 min. The first session was dedicated to general cognitive and reading ability assessment. The second through ninth sessions were each dedicated to one of the four stimulus types (e.g., word, pseudoword, . . .), so there were two sessions of each type. The order of string-types was randomized across participants. At the beginning of each session, we read the participant the general instructions for the task while those instructions were presented on the screen. The instructions encouraged participants to respond as quickly as possible while maintaining a high level of accuracy. Each session was divided into five blocks, one block of string stimuli and a block for each of the corresponding single character stimuli. The final session was a dedicated EEG session, although those data are not further discussed here.

Each block began with a screen depicting the button corresponding to each of the categories. Participants first completed 30 practice trials of the stimulus type in that block. Next, participants completed 170 trials. Half of the trials were with the target stimulus and the other half were divided evenly among the distractor set. Each trial began with a 500 ms presentation of the block instruction screen which included a diagram of a computer mouse that depicted which button to press for the target and distractors, respectively. One button of the mouse was associate with the target string (e.g., "care") and the other button was associated with the distractor(s) (e.g., "bare," "cure," "cave," and "card"). In the single character trials, there was only one stimulus associated with each button (e.g., left button: "c"; right button: "b"). The instruction screen was followed by a 500 presentation of a fixation cross. The stimulus was then presented for 100 ms. Participants had a maximum of 1600 ms to respond. Participants did not receive feedback about the correctness of their response. The session order was counterbalanced among the participants so that participants completed the different types on different days and in different orders.

2.4. Analysis

All data were analyzed using R statistical software (R Development Core Team, 2011). We computed Bayesian ANOVA of the correct target response times using the BayesFactor package (Rouder et al., 2012). The Bayes factor (BF) approach to ANOVA uses model comparison to give evidence for or against including independent variables as predictors for the dependent variables. The BF indicates the ratio of posterior probability of observed data given the model for a pair of models. A rough scale for interpretation of the BF is as follows: <0.01 decisive evidence against; <0.1 strong evidence against; 0.31 substantial evidence against; 0.32–1; minimal evidence against; 1–3.2 is minimal evidence for; >3.2 substantial evidence for; >10 strong evidence for; >100 decisive evidence for (Jeffreys, 1961). Capacity analyses were completed using the sft package (Houpt et al., 2013).

3. Results

3.1. Mean Response Time and Accuracy

For each analysis, we computed the Bayes Factor for a full model, which included string-type (word, pseudoword, random, or Katakana), target/distractor, day (1 or 2), and group (control or dyslexia), relative to a subject intercept only model. We then compared that Bayes factor to successively simpler models which were derived by first removing interactions terms then main effects while maintaining a component for any lower order effects that were included in an interaction term.
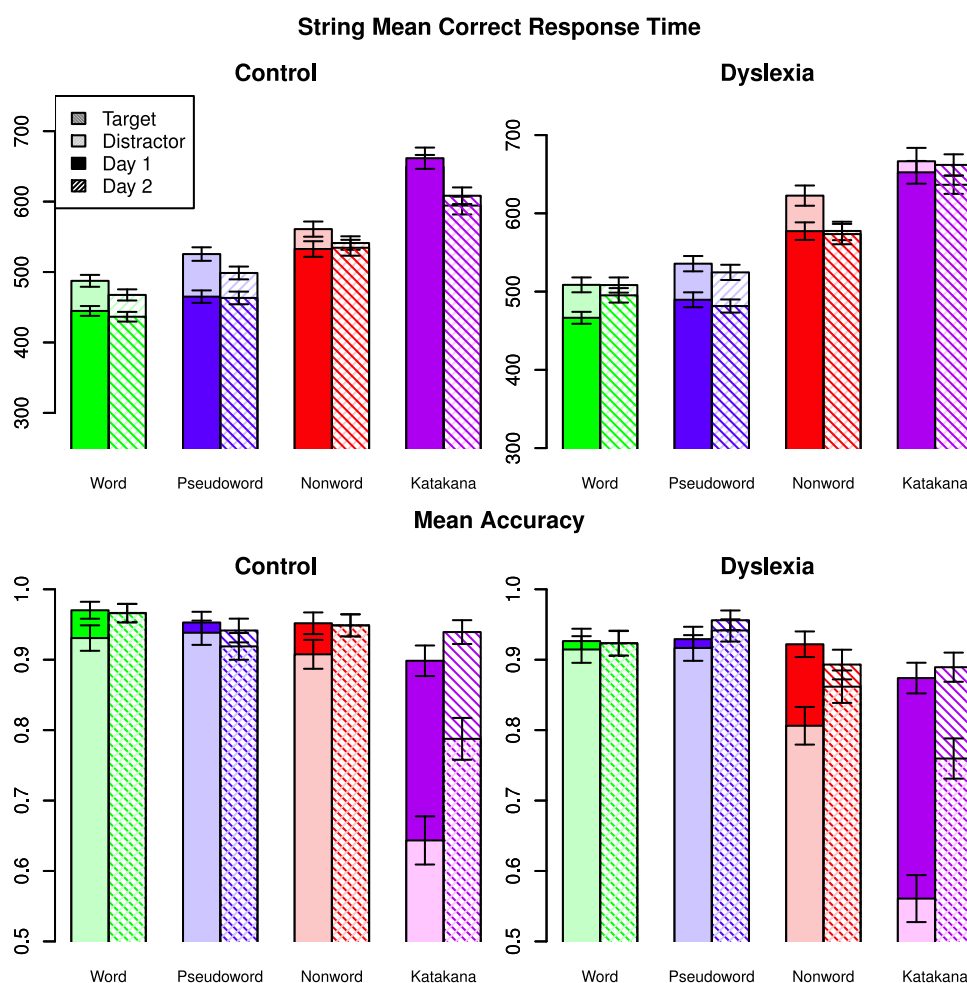
Accuracy and mean correct response times with the string blocks for each string-type are shown in **Figure 1** with error bars representing the 95% credible intervals from the full model. The highest Bayes factor model for correct response times included a three-way interaction among string-type, day and group along with two-way interactions between string-type and target/distractor and day and target/distractor. This model had a Bayes factor of 19.9 (strong evidence) over the next best model, which included a group by target/distractor interaction and was otherwise the same. There was decisive evidence for the best model over all other models (*BF* > 125).

Analysis of the posterior of the full model indicated that the three-way interaction was driven by the control group speeding up on Katakana on Day 2 compared to Day 1, while the dyslexia group was relatively faster on non-words on Day 2 compared to Day 1. The string-type by target/distractor interaction was driven by a cross-over targets being slower for words and pseudowords and faster for Katankana. The string by day interaction,

marginalized across group, showed a cross-over between faster performance for words on Day 1 relative to Day 2 and slower performance for Katakana on Day 1 relative to Day 2. A marginal interaction between string-type and group was mostly driven by faster performance by the controls on the non-word stimuli.

Marginalized over the other factors, words were faster than pseudowords (Posterior Mean = 20.7, 95% HDI = [15.6, 25.7]), non-words (Posterior Mean = 89.3, 95% HDI = [84.1, 94.6]), and Katakana (Posterior Mean = 164, 95% HDI = [159, 170]). Additionally pseudowords were faster than non-words (Posterior Mean = 68.7, 95% HDI = [64.4, 73.4]) and Katakana (Posterior Mean = 144, 95% HDI = [138, 149]) and non-words were faster than Katakana (Posterior Mean = 74.9, 95% HDI = [69.6, 80.3]). Targets were slower than distractors (Posterior Mean = −20.1, HDI = [−23.6, −16.6]). Response time on Day 1 were slower than on Day 2 (Posterior Mean = −14.7, HDI = [−18.6, −10.9]). There was not clear evidence for one group being faster than the other overall (Posterior Mean of Control minus Dyslexia = −25.4, HDI = [−106, 54.5]).

FIGURE 1 | Mean correct response times and mean accuracies for all string types across days, targets/distractors and group. Error bars indicate 95% confidence intervals.

The highest Bayes Factor model for accuracy included a three-way interaction among string-type, day and target/distractor and a two-way interaction between string-type and group. There was strong evidence for this model over a model which also included a group by target interaction ($BF = 12.0$) and over a model that included a group by day interaction ($BF = 29.8$). There was decisive evidence for the best model over all other models ($BF > 159$).

The three-way interaction in accuracy comes from the large increase in distractor performance across days on Katakana and a slight increase in performance for distractors relative to target on words and non-words compared to a unchanged relative performance on the pseudowords across days. Overall there was a larger increase in performance for Katakana than the other string-types, with the smallest changes in the word and pseudoword blocks. Between groups, there was a larger difference in accuracy in the non-word blocks and the smallest difference for pseudowords. Between targets and distractors, the largest difference was for Katakana and the smallest differences were for the word and pseudoword string-types. Generally, distractor performance improved more between the days than target performance.

Marginalized over the other factors, accuracy with words was nearly the same as accuracy on pseudowords (Posterior Mean = 0.00323, 95% HDI = [−0.00667, 0.00129]), slightly better than non-words (Posterior Mean = 0.0351, 95% HDI = [0.0254, 0.0448]), and much better than Katakana (Posterior Mean = 0.146, 95% HDI = [0.136, 0.156]). Additionally pseudowords were slightly more accurate than non-words (Posterior Mean = 0.0318, 95% HDI = [0.0219, 0.0416]) and much more accurate than Katakana (Posterior Mean = 0.143, 95% HDI = [0.133, 0.153]) and non-words were more accurate than Katakana (Posterior Mean = 0.111, 95% HDI = [0.101, 0.121]). Targets were more accurate than distractors (Posterior Mean = 0.0724, HDI = [0.0654, 0.0794]). Accuracy on Day 2 was higher than on Day 1 (Posterior Mean = 0.0326, HDI = [0.0256, 0.0395]). There was not clear evidence for one group being more accurate than the other overall (Posterior Mean of Control minus Dyslexia = 0.0353, HDI = [−0.0476, 0.117]).

Mean correct response time and accuracy with the single character blocks for each type are shown in **Figure 2**.

As in the string data, the best model included a three-way interaction among character-type, day and group. There
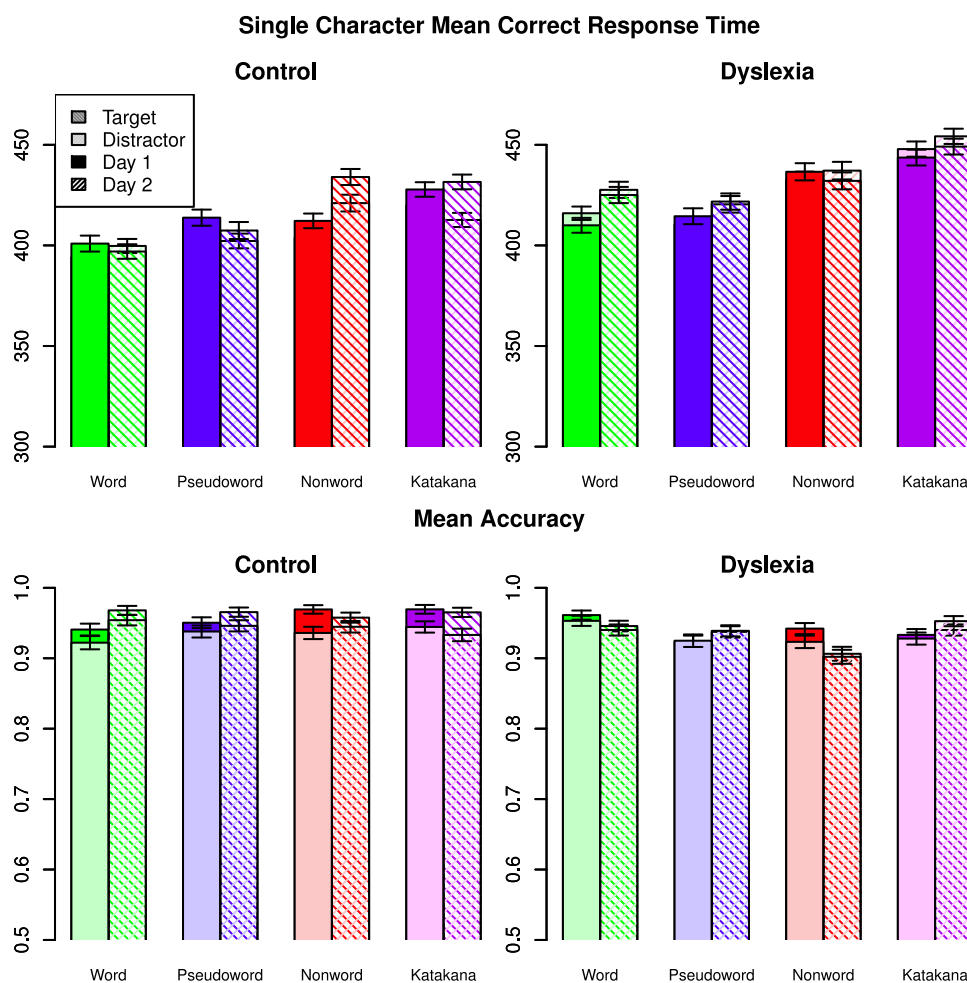
FIGURE 2 | Mean correct response times and mean accuracies on single characters for all types across days, targets/distractors and group. Error bars indicate 95% confidence intervals.

were also two-way interactions between character-type and day, character-type and group, day and group, and group and target/distractor. There was very strong evidence for this model over the next best, which also included a character-type by target/distractor interaction ($BF = 65.7$), and the third best model which included a day by target/distractor interaction ($BF = 81.2$). There was decisive evidence for the best model over all others ($BF > 2600$).

The three-way interaction was driven by the slow-down for control participants on the non-word task between days, while there was no such change for the dyslexia group. The character-type by day interaction was also mainly due to the slow down on the non-words between days. The control group was relatively faster on words and Katakana, while there was a smaller group difference on the non-word characters and nearly no group differences on the characters from pseudowords. The control group slowed down less from Day 1 to Day 2 than the dyslexia, although the magnitude of this difference was small. Control participants were had a relatively larger speed up for distractors over targets on Day 1 than dyslexia participants compared to the second day.

There was a small response time advantage for word characters compared to pseudoword characters when the other factors were marginalized (Posterior Mean = -2.89, HDI = $[-4.73, -1.00]$) and large advantages for word characters over non-word characters (Posterior Mean = -17.5, HDI = $[-19.3, -15.6]$) and Katakana characters (Posterior Mean = -26.1, HDI = $[-27.9, -24.2]$). Pseudoword characters were faster than non-word characters (Posterior Mean = -14.6, HDI = $[-16.5, -12.7]$) and Katakana characters (Posterior Mean = -23.2, HDI = $[-25.0, -21.3]$). Non-word characters were faster than Katakana characters (Posterior Mean = -8.55, HDI = $[-10.4, -6.67]$). The marginal group response times were again indistinguishable (Posterior Mean = -15.6, HDI = $[-74.0, 40.7]$).

For the single character accuracy data, the best fit model again included the three-way interaction among character-type, day and group. There was decisive evidence for this model over all alternative models ($BF \geq 138$). There was a small advantage for word characters over pseudoword characters (Posterior Mean = 0.00756, HDI = $[0.00372, 0.0115]$) and non-word characters (Posterior Mean = 0.0129, HDI = $[0.00906, 0.0168]$) but not a clear difference between word and Katakana characters (Posterior Mean = 0.00239, HDI = $[-0.00146, 0.00624]$). Participants were slightly more accurate characters with pseudoword characters than non-word characters (Posterior Mean = 0.00531, HDI = $[0.00138, 0.00918]$) but less accurate with pseudoword characters compared to Katakana characters (Posterior Mean = -0.00517, HDI = $[-0.00905, -0.00126]$). Participants were also slightly less accurate with non-word characters than with Katakana characters (Posterior Mean = -0.0105, HDI = $[-0.0143, -0.00658]$). There were no clear marginal differences between days (Posterior mean of Day 2 minus Day 1 = 0.00232, HDI = $[-0.000389, 0.00505]$ or groups (Posterior mean of Control minus Dyslexia = 0.0148, HDI = $[-0.0494, 0.0789]$).

Because response time distributions tend to be skewed, and these data are no exception, we also ran an analysis on the log-transformed response time data and found no difference in which model had the highest Bayes factor and only a small difference in the magnitude of that Bayes factor compared to the next best model for the string data ($BF = 17.8$) and resulted in stronger evidence for the character data ($BF = 217$).
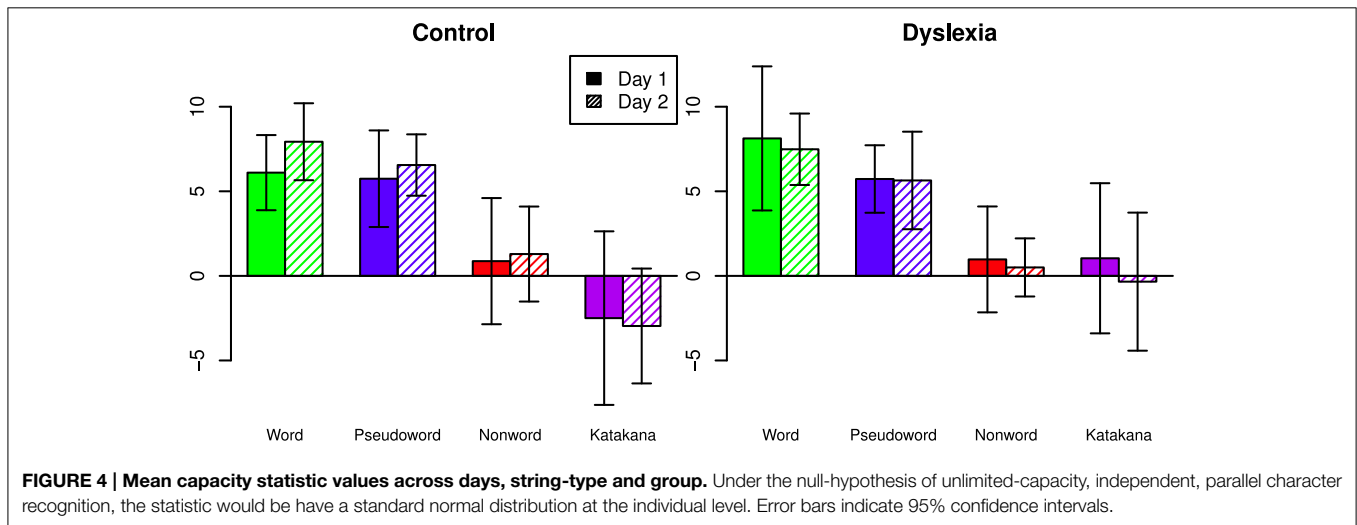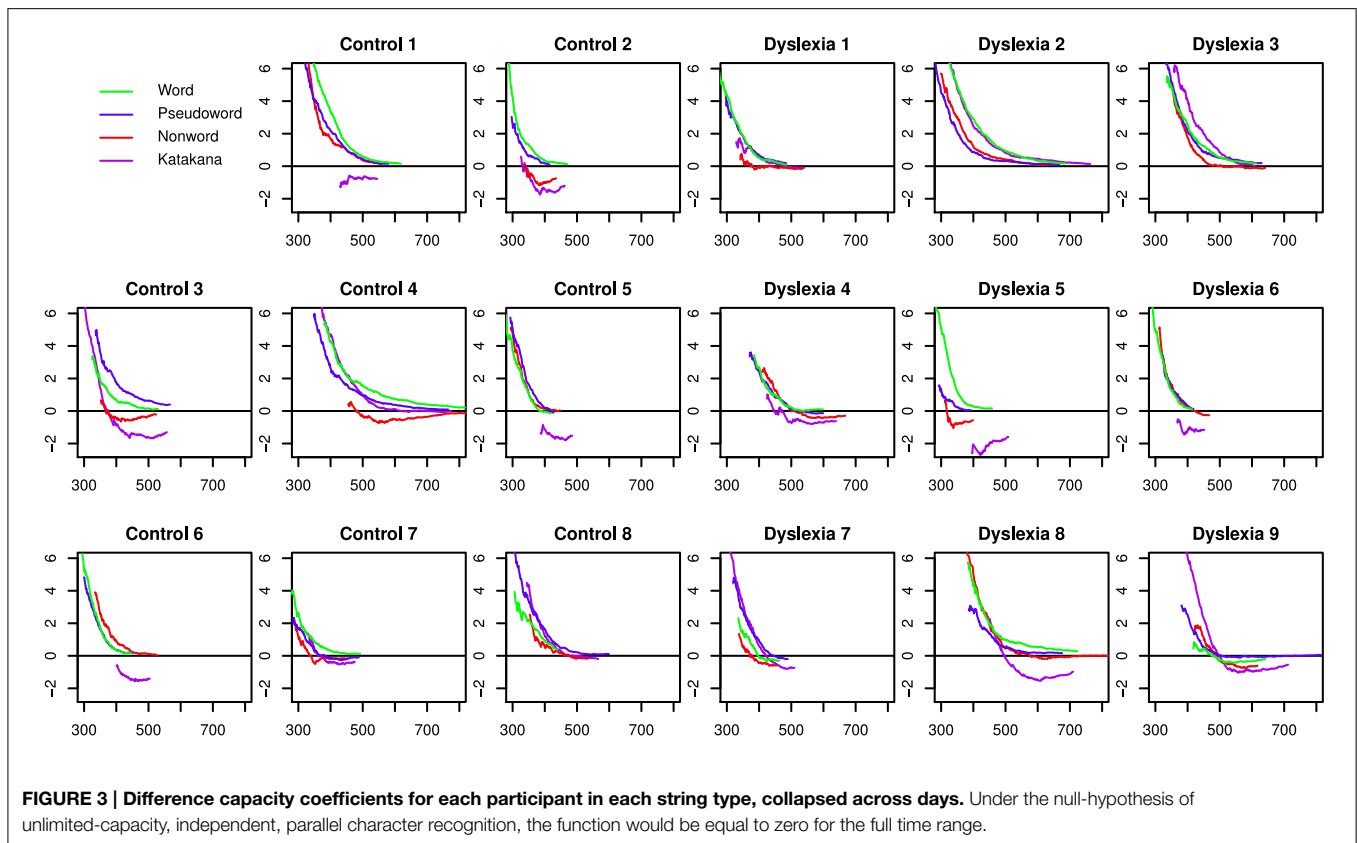
3.2. Capacity Analyses

Capacity coefficients are shown for each individual (collapsed across days) in **Figure 3**. Using the capacity statistic from Houpt and Townsend (2012), participants tended to be super-capacity in the Word (Control: Day 1 = 7/8, Day 2 = 8/8; Dyslexia: Day 1 = 7/9, Day 2 = 7/8 significantly better than baseline) and Pseudoword string-types (Control: Day 1 = 7/8, Day 2 = 8/8; Dyslexia: Day 1 = 9/9, Day 2 = 7/8 significantly better than baseline). **Figure 4** summarizes the overall capacity statistic for each group on each day. There was more variable performance with Katakana (Control: Day 1 = 2/8 above and 5/8 below, Day 2 = 1/8 above and 5/8 below; Dyslexia: Day 1 = 3/9 above and 3/9 below, Day 2 = 3/9 above and 4/9 below) and the non-words (Control: Day 1 = 3/8 above and 2/8 below, Day 2 = 2/8 above and 2/8 below; Dyslexia: Day 1 = 4/9 above and 4/9 below, Day 2 = 1/8 above and 2/8 below).

The best model based on a Bayesian ANOVA measuring day, group and string-type predicting the individual capacity z-scores included a group by string-type interaction as well as main effects of group and string-type. The evidence was nearly equivocal when compared to a model with only a main effect of string-type ($BF = 1.73$) but had at least substantial evidence over all other models ($BF \geq 4.00$). **Table 3** shows the Bayes Factor for the best model relative to all models over which there was not very strong or decisive evidence.

Capacity z-scores were close between words and pseudowords (Posterior Mean = 1.44, HPD = $[-0.367, 3.19]$) and higher for words than non-words (Posterior Mean = 6.29, HPD = $[4.49, 8.10]$) and Katakana (Posterior Mean = 8.35, HPD = $[6.56, 10.1]$). Pseudoword capacity z-scores were higher than both non$-$words (Posterior Mean = 4.86, HPD = $[3.06, 6.64]$) and Katakana (Posterior Mean = 6.91, HPD = $[5.08, 8.75]$). Non-words had higher capacity z-scores than Katakana (Posterior Mean = 2.06, HPD = $[0.292, 3.82]$). There was nearly no marginal difference between groups (Posterior Mean = -0.526, HPD = $[-3.16, 2.04]$).

The capacity z-score gives a summary of the capacity function across time. To check for differences in the shape of capacity coefficient functions, we tested the factor scores obtained from functional principal components analysis (fPCA) of the capacity coefficients (Burns et al., 2013). fPCA is a dimensionality reduction technique that is essentially the same as the more familiar principal components analysis for vectors. The main difference in fPCA is that the data are described in terms of a linear combination of *functions* rather than vectors.

Because the best model of capacity effects did not include day and better estimates of capacity functions lead to more accurate principal component representation, these analysis were performed with data collapsed across day. The fPCA indicated that

FIGURE 3 | Difference capacity coefficients for each participant in each string type, collapsed across days. Under the null-hypothesis of unlimited-capacity, independent, parallel character recognition, the function would be equal to zero for the full time range.

FIGURE 4 | Mean capacity statistic values across days, string-type and group. Under the null-hypothesis of unlimited-capacity, independent, parallel character recognition, the statistic would be have a standard normal distribution at the individual level. Error bars indicate 95% confidence intervals.

the variation across capacity functions was well-represented by three factors related to early, middle and late response time regions (see **Figure 5**).

According to the Bayes factor analyses reported in **Table 4**, there was clear evidence of variation in the capacity functions due to string type in the middle and late time regions. Evidence was present, but less clear, against an effect of group. The analysis was nearly equivocal with respect to meaningful variation in the early time region beyond the variation due to individual subject.

A visual inspection of the individual participant capacity plots in **Figure 3** suggest different patterns of results across string-types for different participants. First, some participants showed much higher capacity for words and pseudowords than for Katakana, with lower capacity for non-words, but not as low as Katakana (e.g., Controls 1, 2, and 3 and Dyslexia 5). This is basically the pattern of results reported in Houpt et al. (2014). Another set of participants had mostly similar capacity functions across string-type (e.g., Controls 7 and 8 and Dyslexia 9).

To investigate these patterns of differences and the extent to which they may be predictive of the basic behavioral measures, we used *k*-means clustering on the fPCA scores. Inspection of a scree plot indicated three clusters would be appropriate for these data. The capacity functions represented by the three cluster means are shown in **Figure 6**. The pattern in Cluster 2 is most similar to the results in Houpt et al. (2014) whereas Cluster 3 represents the participants who had less variation in capacity across string-type. Similar to Cluster 2, Cluster 1 had higher capacity for words and pseudowords and limited capacity for Katakana, but Cluster 2 also had fairly limited capacity for non-words. Control participants were all in either Cluster 2 (4/8) or Cluster 3 (3/8) except Participant C4, who was in Cluster 1. Four of the nine Dyslexia participants were in Cluster 1, three in Cluster 3 and

two in Cluster 2. Note that neither dyslexia status nor the reading and cognitive performance measures contributed to discovering the clusters.
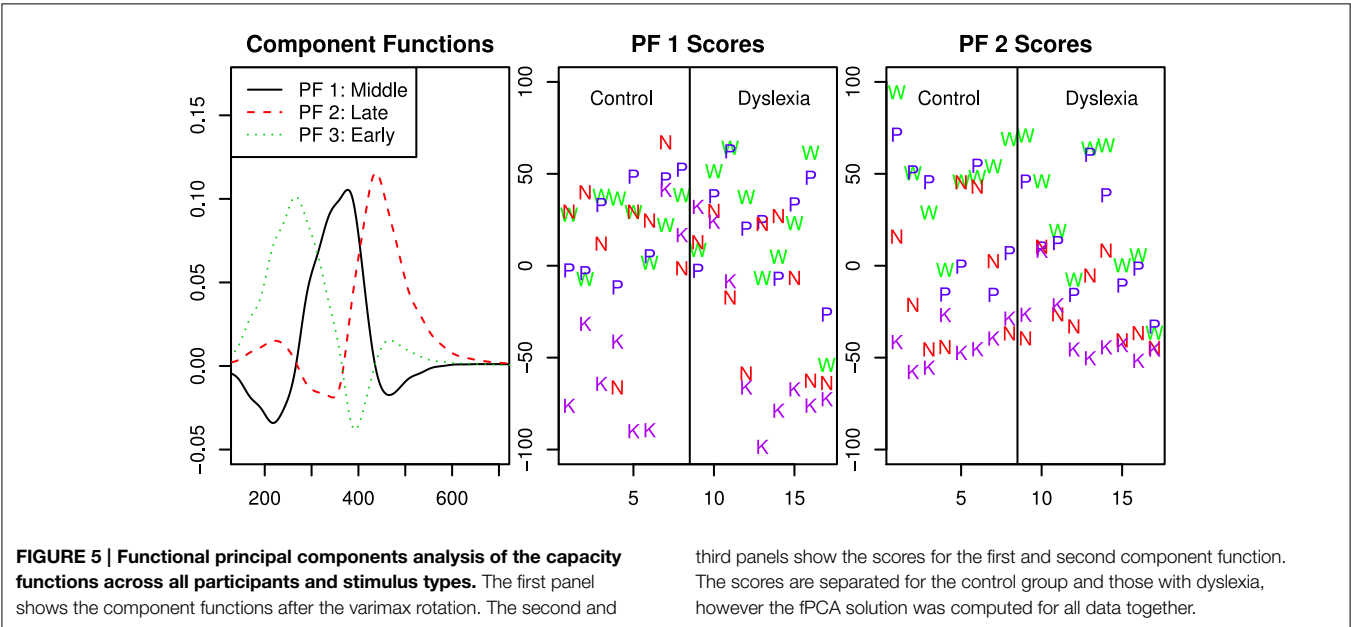
Probing deeper into the connection between the capacity task and the reading and cognitive task, we also examined the variation in those measures across clusters. **Figure 7** shows the distribution (after standardizing across participants) of the basic behavioral measures across each cluster. Generally speaking, Cluster 1 was distinguished in these measures by being have lower handedness scores and lower scores on the Grade Equivalent Word Attack; Cluster 2 had lower Dyslexia checklist scores, higher reading span scores and lower reading history scores; and Cluster 3 had slightly lower verbal IQ scores. Despite the pattern of differences across the measures, Bayesian ANOVAs did not indicate strong evidence either for or against differences among the clusters on any single measure ($0.4 \leq \text{BF} \leq 2.5$ due to the small number of participants in the study.

4. Discussion

In the current study we aimed to explore word perception differences in dyslexia using a novel approach, capacity measures designed to investigate response time latencies. We compared participants with dyslexia and age-matched controls on

TABLE 3 | Bayes Factors for the highest model relative to the next best models for predicting capacity z-scores.

Model	BF
String-type + Subject	1.00
String-type + Group + Subject	0.417
String-type + Day + Subject	0.194
String-type × Group + String-type + Group + Subject	0.181

FIGURE 5 | Functional principal components analysis of the capacity functions across all participants and stimulus types. The first panel shows the component functions after the varimax rotation. The second and third panels show the scores for the first and second component function. The scores are separated for the control group and those with dyslexia, however the fPCA solution was computed for all data together.

TABLE 4 | Bayes Factors relative to the highest model for predicting fPCA capacity scores.

Model	Middle (D1)	Late (D2)	Early (D3)
String-type + Subject	1.00	1.00	1.00
String-type + Group + Subject	0.437	0.633	0.345
String-type × Group + String-type + Group + Subject	0.131	0.270	0.083
Group + Subject	2.96×10^{-6}	8.25×10^{-11}	0.194
Subject	7.83×10^{-6}	1.83×10^{-10}	0.580
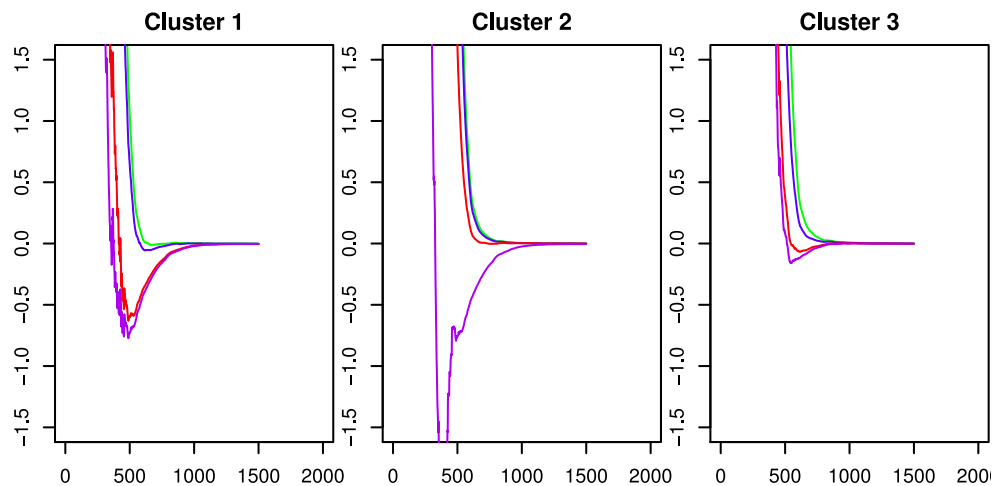
FIGURE 6 | Capacity functions representing the center of each of the three k-means clusters. These are derived by using the mean vector of the cluster on the fPCA scores to as factor weights to determine the functions. The colors indicate the string types using the same scheme as the preceding figures. Word: Green; Pseudoword: Blue; Nonword: Red; Katakana: Purple.
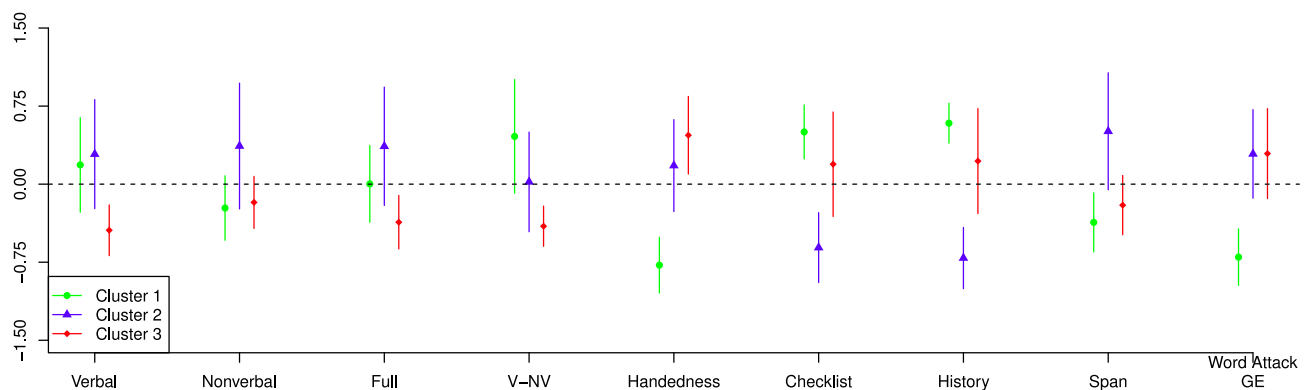
FIGURE 7 | Representation of the variation in the diagnostic tests across the clusters which were derived from the capacity analysis. The points (triangles, squares, and circles) represent the mean value and the lines represent the standard error of the mean.

a discrimination task with four types of stimuli: words, pseudowords, non-words, and Katakana.

The lack of a marginal level difference in either response time, accuracy or capacity based on dyslexia diagnosis replicates and extends the basic finding of Grainger et al. (2003) and the replication in Ziegler et al. (2008): Word superiority effects are present at a group level for those with a dyslexia diagnosis and at a similar magnitude to age-matched control groups. This finding is extended in this paper to a new group, college aged students, and a new paradigm, the design from Houpt et al. (2014).

However, in our current study, the response latency showed a three-way interaction between group, string-type, and day, suggesting that there are some subtle differences between controls and dyslexics. Additionally, the mean capacity results were similar to those found in a previous study by Houpt et al. (2014) using this technique—words and pseudowords had similarly higher

capacity than non-words and non-words had higher capacity than Katakana. Interestingly, when the capacity results were inspected, individual differences emerged such that three different capacity profiles emerged. One group was similar to the non-dyslexics reported in the Houpt study while the other two groups had capacity profiles that diverged in important ways.

The *k*-means clustering analysis indicated three distinct capacity profiles. In an attempt to characterize these three profiles we also explored the cognitive/behavioral scores of the individuals that composed them. The profile that most resembled (Houpt et al., 2014), Cluster 2, had scores more similar to those expected of normal reading adults (i.e., lower dyslexia checklist and reading history scores and higher reading span scores). Indeed, the two dyslexic participants whose capacity profiles were included in Cluster 2 had the lowest dyslexia checklist and reading history scores among those with dyslexia.

Like Cluster 2, the capacity profile for Cluster 1 showed high capacity for words and pseudowords and lower capacity for Katakana, but also showed lower capacity for non-words that was similar to Katakana. The individuals that made up Cluster 1 on average had low Word Attack scores and reading span scores, and high reading history and dyslexia checklist scores, all of which are indicative of dyslexia. The one control participant who was included in this cluster had the highest dyslexia checklist and reading history scores. Interestingly, with the exception of one dyslexic in Cluster 3, the dyslexics in Cluster 1 showed the lowest Word Attack Grade Equivalent scores (all below 7th grade) and the members of this group appear to show an efficiency divide between pronounceable and non-pronounceable string-types. Low performance on Word Attack, particularly in college students, may suggest that the grapheme-to-phoneme processes for this group are particularly affected. This may prompt a "whole word" strategy when reading. They do appear to be efficient in visually recognizing whole regular words and whole pseudowords. The efficiency for pseudowords may be due to repetition causing them to be processed more like words. Although the participants in Cluster 1 had low Word Attack scores, the pseudowords in this study were four-letter, single-syllable pseudowords that are relatively easy to pronounce. Therefore, they may have treated pseudowords like words once they were learned (e.g., on day 2). However, this may not be possible for non-pronounceable consonant strings or foreign characters because they were unable to be learned as words (e.g., non-words are orthotactically invalid and Katakana is not linguistically meaningful). A study by Siegel et al. (1995) suggested that dyslexics with low phonological awareness rely more on orthographic processing. Specifically, they noted a group of dyslexics with poor performance on Word Attack, but high orthographic awareness compared to controls with higher Word Attack scores.

The final profile identified by k-means clustering, Cluster 3, revealed little differences among the four stimulus types. The individuals who showed this profile included both dyslexic and control participants. In terms of test scores, only Word Attack and verbal IQ differentiate Cluster 3 and Cluster 1. On average, individuals in Cluster 3 had higher Word Attack scores and lower verbal IQ. This suggests that these individuals may not have a weaknesses related to grapheme-to-phoneme conversion, but may have deficits in other language-related processes that account for the lower verbal IQ. The finding that the capacity scores were similar across stimulus types suggests that individuals in Cluster 3 used a generalized strategy. Because all participants were naive to Katakana, a generalized strategy could not have depended on linguistic processing but may instead have depended on visual feature processing. This strategy is apparently very efficient and able to handle complex unfamiliar visual stimuli. It is possible that this is a global, holistic process. Some evidence to support such a strategy comes from a study examining high school students that found dyslexics were faster, but not more accurate, at detecting impossible objects (von Károlyi, 2001). They found that these students relied on global processing of the objects (e.g., recognizing features simultaneously and discerning if they contradict each other). While Katakana does not have any inherent contradictory features in this study, if we

situate the target Katakana string as the goal this contextualizes the distractor strings as somewhat contradictory. It is possible that the participants who were efficient at Katakana (as well as the other string-types) were processing the strings as whole objects. It is also possible that many of the dyslexic members of Cluster 3 were especially good at Katakana because language processing could not "get in the way." They may then have been able to generalize a visual, non-linguistic strategy into the other categories.

While it may be that individuals in Cluster 3 used a non-linguistic strategy, an alternative explanation is that a linguistic strategy was used for non-word and Katakana stimuli. In an MEG study of visual word recognition in dyslexia, Salmelin et al. (1996) found that non-dyslexics displayed a typical sharp negativity around 180 ms in temporo-occipital regions to words, but dyslexics only activated this region after 200 ms with a slowly increasing signal that peaked closer to 450 ms. Some of the participants in the current study also participated in an pilot EEG session of the task after completing the study. Generally, participants who showed a profile similar to Cluster 3 failed to show a sharp left N170 in response to the stimuli, but instead showed a more gradual negativity in less lateralized posterior electrodes that peaked between 220-350 ms; this pattern was fairly consistent across string-type (Sussman et al., 2011). In contrast, a control with a non-clumping capacity pattern, similar to Cluster 2, showed a more typical pattern of an N180 in left temporo-occipital electrodes for words, pseudowords, and non-words; but for Katakana did not show this N180 response. The correspondence between our EEG data and (Salmelin et al., 1996) potentially suggests that the participants who show similarity in capacity across all four string-types are generalizing a strategy from words to Katakana and not vice versa. This also suggests, however, that the presumed compensatory strategy they are using requires visual language processes. Interestingly, the Cluster 3 pattern is not unique to the dyslexia participants and was, in fact, used by some controls. That most (all three dyslexics and one control) of the subjects in Cluster 3 showed super-capacity for Katakana suggests that the strategy was more generalized across string-types, but not always efficient. It is possible that particularly the dyslexics in this group are more practiced at using a generalized strategy. Further research is necessary to determine the strategy being used by individuals in Cluster 3.

Together with the results from Grainger et al. (2003) and Ziegler et al. (2008), these results indicate that there is no general deficit in orthographic recognition, either at the single character or configural level, with dyslexia. Some of the participants with dyslexia were differentiated from most of the control participants in this task, but the main difference was in their performance on non-words. Given the low Word Attack scores, it is unlikely that the participants with dyslexia are using phonological information for better performance in the word and pseudoword condition, so they are potentially relying on information from the orthographic configurations. The subgroup that performed worse on non-words may have relied more on statistical regularities in letter combinations (cf. Pelli et al., 2003) than the participants who were not much worse with non-words. Although previous research has shown that the effect of orthographic regularity

across languages (English and German) is similar across participants with and without dyslexia (Landerl et al., 1997), in future research it would be worthwhile to investigate whether there is a difference in the effect of orthographic regularity associated with the different capacity profiles reported herein.

One potential limitation of the current study, and of the approach in Houpt et al. (2014), is that only a single string is used for each string-type. In the standard Reicher–Wheeler paradigm, a different word is used on each trial. Because the repeated presentation of the string, there is ample opportunity for the participants to use encoding strategies that are efficient for those particular strings, but are not necessarily representative of the participants' ability across the whole class of string-types that is represented by that string. Despite this possibility, Houpt et al. (2014) found a clear differentiation among the string types. Although it is more parsimonious to assume, that the same perceptual process differences underly the word and pseudoword superiority effects observed in both (Houpt et al., 2014) and the Reicher-Wheeler design, it leaves open the possibility that the individual differences in this study were due to differences in perceptual learning rather than differences in more general, stable, perceptual encoding strategies.

Another limitation of the current work is that the participants were undergraduate and graduate students at a major university. These participants may not be representative of the wider range of adults with dyslexia. Furthermore, these participants have had many years of reading practice to develop strategies for ameliorating the effects of dyslexia. In future work, it would be informative to use this paradigm with younger children who have not had access to as many years of remediation training as the

participants in this study. This would facilitate further connection between the effects reported here and the previous studies of dyslexia and the word superiority effect (Grainger et al., 2003; Ziegler et al., 2008). It would be particularly interesting to test if the same clusters of capacity performance emerge with younger participants or perhaps if there is some effect of remediation training on the capacity patterns. More generally speaking, this is a relatively small sample of participants for individual differences research and we hope to expand these results to a much larger sample.

To conclude, the results presented here emphasize the importance of exploring individual differences. The dyslexic group, like the control group, is not homogeneous; they do not all process word and word-like strings in the same way. Here, when examining capacity profiles, three different subgroups were observed and there were both control and dyslexic participants in each of these groups. While it is difficult to detect these patterns by only examining the accuracy data from tasks designed to explore the word superiority effect (e.g., Grainger et al., 2003), by using response latency data to predict independent, parallel processing, group differences emerged. These types of analyses may prove to be informative and provide information regarding how individuals are processing word stimuli, which can then be used to develop remediation tools that are tailored to an individual dyslexic.

Funding

References

Baron, J., and Thurston, I. (1973). An analysis of the word-superiority effect. *Cogn. Psychol.* 4, 207–228.

Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., and Blomert, L. (2009). Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Curr. Biol.* 19, 503–508. doi: 10.1016/j.cub.2009.01.065

Blomert, L. (2011). The neural signature of orthographic–phonological binding in successful and failing reading development. *Neuroimage* 57, 695–703. doi: 10.1016/j.neuroimage.2010.11.003

Burns, D. M., Houpt, J. W., Townsend, J. T., and Endres, M. J. (2013). Functional principal components analysis of workload capacity functions. *Behav. Res. Methods* 45, 1048–1057. doi: 10.3758/s13428-013-0333-2

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind* 11, 63–65.

Estes, W. K., and Brunn, J. L. (1987). Discriminability and bias in the word-superiority effect. *Percept. Psychophys.* 42, 411–422.

Grainger, J., Bouttevin, S., Truc, C., Bastien, M., and Ziegler, J. (2003). Word superiority, pseudoword superiority, and learning to read: a comparison of dyslexic and normal readers. *Brain Lang.* 87, 432–440. doi: 10.1016/S0093-934X(03)00145-7

Healy, A. F. (1994). Letter detection: a window into unitization and other cognitive processes in reading text. *Psychon. Bull. Rev.* 3, 333–344.

Houpt, J., Townsend, J., and Donkin, C. (2014). A new perspective on visual word processing efficiency. *Acta Psychol.* 145, 118–127. doi: 10.1016/j.actpsy.2013.10.013

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., and Townsend, J. T. (2013). Systems factorial technology with R. *Behav. Res. Methods* 46, 307–330. doi: 10.3758/s13428-013-0377-3

Houpt, J. W., and Townsend, J. T. (2012). Statistical measures for workload capacity analysis. *J. Math. Psychol.* 56, 341–355. doi: 10.1016/j.jmp.2012.05.004

Jacobs, A. M., and Grainger, J. (1994). Models of visual word recognition: sampling the state of the art. *J. Exp. Psychol.* 20, 1311–1334.

Jeffreys, H. (1961). *The Theory of Probability, 3rd Edn.* Oxford.

Kucera, H., and Francis, W. (1967). *Computational Analysis of Present-day American Engish*. Providence, RI: Brown University Press.

Landerl, K., Wimmer, H., and Frith, U. (1997). The impact of orthographic consistency on dyslexia: a German–English comparison. *Cognition* 63, 315–334.

Lefly, D. L., and Pennington, B. F. (1991). Spelling errors and reading fluency in compensated adult dyslexics. *Ann. Dyslexia* 41, 141–162.

Lefly, D. L., and Pennington, B. F. (2000). Reliability and validity of the adult reading history questionnaire. *J. Learn. Disabil.* 33, 286–296. doi: 10.1177/002221940003300306

Manelis, L. (1974). The effect of meaningfulness in tachistoscopic word perception. *Percept. Psychophys.* 16, 182–192.

Massaro, D. W. (1973). Perception of letters, words, and nonwords. *J. Exp. Psychol.* 100, 349–353.

McClelland, J. L., and Johnston, J. C. (1977). The role of familiar units in perception of words and nonwords. *Percept. Psychophys.* 22, 249–261.

Nicolson, R. I., and Fawcett, A. J. (2011). Dyslexia, dysgraphia, procedural learning and the cerebellum. *Cortex* 47, 117–127. doi: 10.1016/j.cortex.2009.08.016

Oldfield, R. C. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia* 9, 97–113.

Pelli, D. G., Farell, B., and Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature* 423, 752–756. doi: 10.1038/nature01516

Pomerantz, J. R., Sager, L. C., and Stoever, R. J. (1977). Perception of wholes and of their component parts: some configural superiority effects. *J. Exp. Psychol.* 3, 422.

R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Rack, J. P., Snowling, M. J., and Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: a review. *Read. Res. Q.* 27, 29–53.

Ramus, F. (2003). Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Curr. Opin. Neurobiol.* 13, 212–218. doi: 10.1016/S0959-4388(03)00035-7

Rastle, K., Harrington, J., and Coltheart, M. (2002). 358,534 nonwords: the ARC nonword database. *Q. J. Exp. Psychol.* 55A, 1339–1362. doi: 10.1080/02724980244000099

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *J. Exp. Psychol.* 81, 274–280.

Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Bayes factors for anova designs. *J. Math. Psychol.* 56, 356–374. doi: 10.1016/j.jmp.2012.08.001

Salmelin, R., Kiesil, A. P., Uutela, K., Service, E., and Salonen, O. (1996). Impaired visual word processing in dyslexia revealed with magnetoencephalography. *Ann. Neurol.* 40, 157–162.

Shaywitz, S. E. (1998). Dyslexia. *N. Engl. J. Med.* 338, 307–312.

Shaywitz, S. E., Fletcher, J. M., Holahan, J. M., Shneider, A. E., Marchione, K. E., Stuebing, K. K., et al. (1999). Persistence of dyslexia: the connecticut longitudinal study at adolescence. *Pediatrics* 104, 1351–1359.

Shaywitz, S. E., and Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biol. Psychiatry* 57, 1301–1309. doi: 10.1016/j.biopsych.2005.01.043

Siegel, L. S., Share, D., and Geva, E. (1995). Evidence for superior orthographic skills in dyslexics. *Psychol. Sci.* 6, 250–254.

Spoehr, K. T., and Smith, E. E. (1975). The role of orthographic and phonotactic rules in perceiving letter patterns. *J. Exp. Psychol.* 104, 21–34.

Stanovich, K. E. (1996). Toward a more inclusive definition of dyslexia. *Dylexia* 2, 154–166.

Stein, J. (2001). The magnocellular theory of developmental dyslexia. *Dyslexia* 7, 12–36. doi: 10.1002/dys.186

Sussman, B. L., Houpt, J. W., Townsend, J. T., and Newman, S. D. (2011). "EEG correlates of visual word processing efficiency in dyslexia," in *Poster Presented at: Society for Neuroscience Annual Meeting* (Washington, DC).

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain Lang.* 9, 182–198.

Townsend, J. T., and Nozawa, G. (1995). Spatio-temporal properties of elementary perception: an investigation of parallel, serial and coactive theories. *J. Math. Psychol.* 39, 321–360.

Townsend, J. T., and Wenger, M. J. (2004). A theory of interactive parallel processing: new capacity measures and predictions for a response time inequality series. *Psychol. Rev.* 111, 1003–1035. doi: 10.1037/0033-295X.111.4.1003

Vinegrad, M. (1994). A revised adult dyslexia checklist. *Educare* 48, 21–23.

von Károlyi, C. (2001). Visual-spatial strength in dyslexia: rapid discrimination of impossible figures. *J. Learn. Disabil.* 34, 380–391. doi: 10.1177/002221940103400413

Weschler, D. (1999). *Weschler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Harcourt Assesment.

Wheeler, D. D. (1970). Processes in word recognition. *Cogn. Psychol.* 1, 59–85.

Wilson, A. M., and Lesaux, N. K. (2001). Persistence of phonological processing deficits in college students with dyslexia who have age-appropriate reading skills. *J. Learn. Disabil.* 34, 394–400. doi: 10.1177/002221940103400501

Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* Itasca, IL: Riverside Publishing.

Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F., Perry, C., et al. (2008). Developmental dyslexia and the dual route model of reading: simulating individual differences and subtypes. *Cognition* 107, 151–178. doi: 10.1016/j.cognition.2007.09.004

Individual differences in attention influence perceptual decision making

Michael D. Nunez[1]*, Ramesh Srinivasan[1,2] and Joachim Vandekerckhove[1,3]

[1] Department of Cognitive Sciences, University of California, Irvine, Irvine, CA, USA
[2] Department of Biomedical Engineering, University of California, Irvine, Irvine, CA, USA
[3] Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA, USA

Sequential sampling decision-making models have been successful in accounting for reaction time (RT) and accuracy data in two-alternative forced choice tasks. These models have been used to describe the behavior of populations of participants, and explanatory structures have been proposed to account for between individual variability in model parameters. In this study we show that individual differences in behavior from a novel perceptual decision making task can be attributed to (1) differences in evidence accumulation rates, (2) differences in variability of evidence accumulation within trials, and (3) differences in non-decision times across individuals. Using electroencephalography (EEG), we demonstrate that these differences in cognitive variables, in turn, can be explained by attentional differences as measured by phase-locking of steady-state visual evoked potential (SSVEP) responses to the signal and noise components of the visual stimulus. Parameters of a cognitive model (a diffusion model) were obtained from accuracy and RT distributions and related to phase-locking indices (PLIs) of SSVEPs with a single step in a hierarchical Bayesian framework. Participants who were able to suppress the SSVEP response to visual noise in high frequency bands were able to accumulate correct evidence faster and had shorter non-decision times (preprocessing or motor response times), leading to more accurate responses and faster response times. We show that the combination of cognitive modeling and neural data in a hierarchical Bayesian framework relates physiological processes to the cognitive processes of participants, and that a model with a new (out-of-sample) participant's neural data can predict that participant's behavior more accurately than models without physiological data.

Keywords: electroencephalography (EEG), steady-state visual evoked potential (SSVEP), Phase-locking, hierarchical Bayesian modeling, diffusion models, individual differences, perceptual decision making

1. INTRODUCTION

The joint analysis of physiological and behavioral data has been a topic of recent interest. In a string of publications, a number of research groups (Forstmann et al., 2010; Turner et al., 2013; Cassey et al., 2014) have presented work in which neurophysiological data are linked to parameters of cognitive or behavioral process models (see also Palmeri et al., in preparation). The goal of these modeling exercises is not only to evaluate the predictive power of brain activity for behavior, but also to elucidate the nature of this prediction. The use of cognitive models with neural data and cognitive parameters permits more psychologically interpretable labeling of the neurophysiological measurements, providing links between brain activity, cognition, and behavior.

In the present paper, we apply a cognitive model constrained by EEG data to fit accuracy and response times of multiple individuals from a perceptual decision making task. The goal of the model fit is twofold: (1) to demonstrate the superior generalizability of such a model as compared to model variants without neural input components and (2) to evaluate the hypothesis that individual differences in enhancement or suppression of visual

attention, as measured by EEG, contribute to individual differences in cognition and thus to individual differences in accuracy and/or reaction time in the task.

In order to show out-of-sample generalizability, we first fit the model to a training set of participants and obtain the requisite (population-level) linking parameters, and then make predictions about the behavior of a new participant to which the model was not trained. In the sections that follow, we will describe (1) the cognitive process model that we have chosen, (2) the task to which it is applied and the EEG data that we collected, (3) a series of three models of increasing complexity, of which the model with external attentional EEG covariates is the most complex, (4) the results of the generalization exercise and (5) evaluation of the hypothesis.

1.1. STEADY-STATE VISUAL EVOKED POTENTIALS AS A MEASURE OF ATTENTION

In this study, we will demonstrate how attentional mechanisms can explain individual differences in perceptual decision making as estimated by a cognitive model. In a typical visual attention

experiment, the signal stimulus is attended and preferentially processed while competing stimuli (i.e., visual noise) are not further processed. A number of studies have demonstrated that a measure of the deployment of attention can be obtained by using flickering stimuli and electroencephalographic (EEG) recordings of the (frequency tagged) steady-state visual evoked potentials (SSVEPs) (Morgan et al., 1996; Müller et al., 1998; Ding et al., 2006; Bridwell and Srinivasan, 2012; Garcia et al., 2013). SSVEPs are narrow band responses at the visual flicker frequencies and flicker harmonics of a stimulus (Regan, 1977). When a stimulus is attended, the SSVEP is enhanced, and when a stimulus is not attended or suppressed, the SSVEP is diminished. This approach has been used to investigate individual differences in attention strategy in detection and discrimination tasks. Bridwell et al. (2013) found that only a subset of participants could deploy the optimal attention strategy and modify their strategy by the task demands. An SSVEP approach has also been used to show that individuals are trained by their own experiences. Individuals with attentional training due to a history of fast-action video gaming have been found to preferentially suppress noise rather than enhance the signal, and those individuals performed better at vigilance tasks (Krishnan et al., 2013).

1.2. DIFFUSION MODELS FOR TWO-CHOICE RESPONSE TIMES

Diffusion models are a class of sequential-sampling models for reaction time (RT) and response data that can capture the joint distribution of RT and accuracy in speeded choice tasks. This family of models has been useful in explaining between- and within-participant variability in two-alternative forced choice decision making experiments (Vandekerckhove et al., 2008, 2011). Diffusion models also add to the analyses of participants' behavior by assuming underlying cognitive processes which have some empirical validation (Voss et al., 2004). In particular, they assume that at each trial, participants obtain relative evidence from a stimulus over time until sufficient evidence is accumulated to exceed the threshold for one of the two choices (Stone, 1960; Link and Heath, 1975; Ratcliff, 1978). This process of relative evidence accumulation is modeled as a Wiener diffusion process (or *Brownian motion*) and can be thought of as a continuous random walk process—that is, a random walk process where in each infinitesimal time step, the evidence increases by a random amount according to a normal distribution with some mean and some instantaneous variance (Ratcliff, 1978). A visual representation of the model is provided in **Figure 1**.

Fitting RT and choice behavior using the diffusion model is a useful behavioral analysis tool since the model's parameters have interpretable psychological correlates. The drift rate δ_j represents the mean rate of evidence accumulation of participant j during their decision process. The drift rate is thought to reflect the quality of evidence the participant obtains during an experimental trial (Ratcliff et al., 2001). The diffusion coefficient ς_j is the parameter that represents the amount of variability in the evidence accumulation process within one trial (i.e., the instantaneous variance). The bias parameter β_j is the proportion of bias a participant j has in favor of choice A over choice B (it should be noted that we fix the bias parameter to $\frac{1}{2}$ in this paper since we
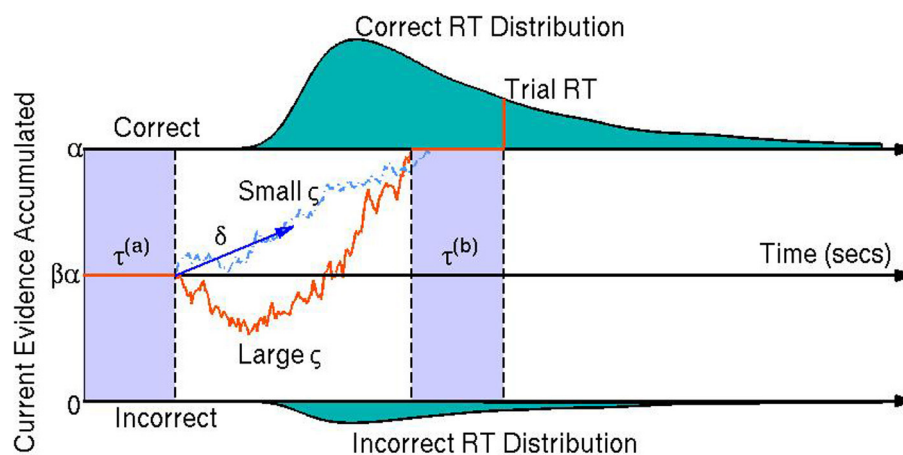
FIGURE 1 | A visual representation of the diffusion model. The orange line represents the participant's stochastic evidence accumulation process during one trial. When a participant accumulates enough evidence over time for a correct or incorrect response (graphically represented by the top and bottom boundaries at 0 and α, respectively) a decision is made. The drift rate δ is the mean rate of evidence accumulation (evidence units per second) during the participant's decision time on one trial. The bias parameter β represents a bias of the participant toward one choice or the other (set to $\frac{1}{2}$ when the model parameters are expressed in terms of correct over incorrect evidence instead of choice A over choice B evidence). The non-decision time τ is the portion of the participant's reaction time (RT) during the trial not associated with decision making, equal to the sum of encoding/preprocessing time $\tau^{(a)}$ and motor response

time $\tau^{(b)}$ which are not estimable. The boundary separation parameter α represents the amount of relative evidence needed to make a decision. Another parameter is the variability in the evidence accumulation process, the diffusion coefficient ς. In this trial the diffusion coefficient is large in comparison to a smaller diffusion coefficient as shown by the light blue dashed line. The teal shaded areas represent the correct (top) and incorrect (bottom) reaction time distributions. In this example systematic component of the decision making process is positive $\delta > 0$ indicating a mean trend toward correct responses. However, incorrect responses can still be reached due to the random component of the decision making process (the diffusion coefficient ς). Larger ς indicate that a participant would be increasing likely to make faster decisions, but have closer to chance performance (i.e., an accuracy of β).

model behavior as correct vs. incorrect trials instead of choice A over choice B trials). The non-decision time τ_j is the amount of time during the response process that is not associated with the decision making process, such as preprocessing of the stimulus and/or motor response time. Finally, the boundary separation parameter α_j represents the amount of relative evidence needed to make a decision and is typically manipulated by task instructions emphasizing either speed or accuracy (Ratcliff et al., 2001; Voss et al., 2004). It is important to note that the model is not identifiable unless we constrain at least one of the parameters that pertain to the evidence dimension (i.e., diffusion coefficient ς, drift rate δ, or boundary separation α).

1.3. THE CASE FOR HIERARCHICAL BAYESIAN MODELS

Recent advances in mathematical psychology have introduced hierarchical Bayesian versions of cognitive models (Rouder et al., 2005; Vandekerckhove et al., 2011). The advantages of these hybrid modeling–measurement strategies include more principled (Bayesian) statistical inference, increased statistical power (Vandekerckhove et al., 2010), and interpretability of results in terms of psychological concepts rather than statistical summary (Vandekerckhove, 2014). The use of cognitive models as measurement tools has become known as *cognitive psychometrics* (e.g., Batchelder, 2010).

The hierarchical Bayesian process modeling framework is ideally suited for the joint analysis of multiple modes of data—(Turner et al., 2013) describe three such joint modeling strategies and (Vandekerckhove, 2014) describes a fourth. One strategy afforded by hierarchical Bayesian models involves constraining the estimation of cognitive process models by introducing the brain data as (fixed) covariate information. This strategy carries the disadvantage that it does not by default allow for measurement variance on the neurophysiological side, but has the advantage of being relatively straightforward to implement in a computationally efficient fashion. By conditioning the estimation of the cognitive parameters on brain data (or other external covariates), it is expected that unexplained variability between participants can be reduced, and consequently that such a model should perform better in generalization tests.

Interindividual variability (i.e., variability in the participant-level cognitive parameters; changes over subscript j) in diffusion models has been previously analyzed by fitting a diffusion model to each participant individually then comparing parameters across model fits. The individual differences were then gauged by statistical analyses on the models' resulting maximum likelihood parameter estimates (Ratcliff et al., 2001; Wagenmakers et al., 2008). Some limitations to this technique are that large sample sizes are needed for diffusion model parameter estimation, that shared condition-level differences across individuals cannot be easily evaluated (Wagenmakers, 2009; Vandekerckhove et al., 2011), and that statistical uncertainty is not propagated across stages of the analysis. Hierarchical Bayesian methods along with Monte Carlo sampling techniques allow for the estimation of complex models. These methods have been used to explain individual differences in the diffusion model and other cognitive models without the need for large sample sizes (Lee, 2008; Lee and Newell, 2011; Vandekerckhove et al., 2011). Additionally, the

hierarchical framework allows for between-participant variability to be explained when each participant's diffusion model parameters are functionally related to known exogenous data (e.g., physiological data).

1.4. CONSTRAINING MODEL PARAMETERS WITH EEG DATA

We assume that brain activity compels cognition, which in turn drives participant behavior. Assuming attention constrains one or more of the cognitive processes in perceptual decision making, then as a consequence of attentional mechanisms we expect SSVEPs to help explain between-participant variability in the parameters of the diffusion model and thus between-participant variability in RT and accuracy. In one study, an occipital SSVEP amplitude was shown to track visual sensory evidence over the time course of a trial, suggesting that SSVEPs can reflect the evidence accumulation process itself (O'Connell et al., 2012). The experimental stimulus used in this study involves a flickering signal overlaid on time-varying visual noise, designed to evoke separate SSVEP responses to the signal and the visual noise, which we expect will explain individual differences in the model parameters and behavior.

We hypothesize increased within-trial evidence accumulation rates, reflected by increased drift rates, for those subjects who suppressed attention to the visual noise. We further hypothesize that another benefit of attention for RT and accuracy is a result of reduced within-trial variability in the accumulation of evidence. Thus, we predict an across-individuals relationship between enhanced attention to the signal and decreased diffusion coefficients.

As mentioned above, one of the parameters of the diffusion model must be fixed rather than estimated (either diffusion coefficient ς, drift rate δ, or boundary separation α). For the present study a variable boundary separation across conditions is not a valid interpretation of the data since the changes between conditions occur unannounced, leaving the participant with no opportunity to adapt strategies (e.g., switch between a speed or accuracy strategy) in response to stimulus changes. In our parameterization, we leave the diffusion coefficient ς free to vary, set α to one evidence unit, and assume no bias ($\beta = \frac{1}{2}$) toward correct responses. The joint density f of RT t and accuracy w of this simplified diffusion model is given in Equation 1. The density is derived from the limiting approximation given by Ratcliff (1978) where $z = \frac{1}{2}\alpha$ and $\alpha = 1$.

$$
\begin{cases}
f(t, w = 0 \mid \varsigma^2, \tau, \delta) = \pi \varsigma^2 e^{-\frac{1}{2}\left[\frac{\delta}{\varsigma} + \delta^2(t-\tau)\right]} \\
\qquad \sum_{k=1}^{+\infty} \left[k \sin\left(\frac{1}{2}\pi k\right) e^{-\frac{1}{2}k^2\pi^2\varsigma^2(t-\tau)} \right] \\
f(t, w = 1 \mid \varsigma^2, \tau, \delta) = f(t, w = 0 \mid \varsigma^2, \tau, -\delta)
\end{cases}
\tag{1}
$$

In what follows, we will use the effect of attention, as measured by SSVEPs, to constrain diffusion model parameter estimates (in our case δ_j, ς_j, and τ_j). In particular, we assume that, on each trial, a participant's attention is reflected in phase locking (i.e., SSVEPs) to the attended visual signal and decreased phase locking to the unattended visual noise.

We will demonstrate that the hierarchical Bayesian SSVEP-driven diffusion model has predictive ability as well as descriptive ability—more specifically, that our ability to predict each participant's accuracy and RT behavior is improved by including the SSVEP measures of attention processes.

2. MATERIALS AND METHODS

2.1. PARTICIPANTS

The following study was approved by the University of California, Irvine Institutional Review Board and was performed in accordance with APA standards. Informed consent was obtained from each of the seventeen participants (8 females and 9 males) who took part in the study. The mean age of 16 of the participants was 25 with an age range of 21–30. Another participant was over 45 years of age. Sixteen participants self-identified as being right handed while another identified as being left or ambidextrous. All participants had at least 20/30 vision or corrected vision as measured by a visual acuity chart available on the internet (Olitsky et al., 2013). No participants reported any history of neurological disorder. Each participant completed the experiment in one session within 2.5 h.
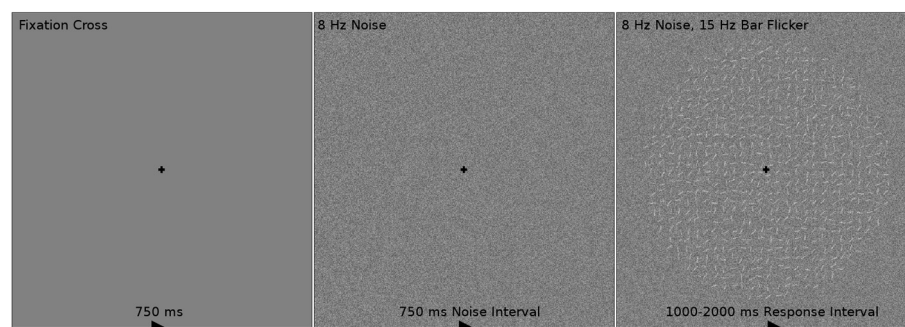
2.2. EXPERIMENTAL STIMULUS

The participants were given a two-alternative forced-choice perceptual decision making task in which they were asked to differentiate the mean rotation of bars within a circular field of bars that deviated randomly from mean rotation. One half of the trials had a mean bar rotation of 45° while the other half had a mean rotation of 135°. The bar field was flickered against a time-varying noise pattern.

The participants viewed each trial of the experimental stimulus on a monitor in a dark room. The time course of one trial is shown in **Figure 2**. Participants were positioned such that the entire circular field of small oriented bars had a visual angle of 9.5°. Within each trial the participant first saw a black cross for 750 ms in the middle of the screen on which they were instructed to maintain fixation throughout the trial. The participant then observed visual contrast noise changing at 8 Hz for 750 ms; this time period of the trial will be referred to later in this paper as the *noise interval*. The participant then observed a circular field of small oriented bars flickering at 15 Hz overlaid on the square field of visual noise pattern changing at 8 Hz and responded during this time frame, henceforth referred to as the *response interval*. The visual noise and bar field are modulated at constant rates (8 and 15 Hz, respectively) to evoke frequency-tagged signal and noise responses in the cortex which we measured as steady-state visual evoked potentials (SSVEPs). The SSVEP responses at the signal frequencies (15 Hz and its harmonics) and at the contrast noise frequencies (8 Hz and its harmonics) were used to measure the effect of attention to the signal stimulus and noise stimulus. The display time of the response interval was sampled between 1000 and 2000 ms from a uniform distribution. After this display period the black fixation cross was shown in isolation for 250 ms to alert the participant the trial was over and to collect any delayed responses.

Three levels of variance of bar rotation and three levels of contrast noise were used to modulate the task difficulty. In the first level of bar rotation variance, each bar was drawn from a uniform $\mathcal{U}(-30°, 30°)$ distribution centered on the mean angle. In the two other levels, the rotations of each bar were drawn from $\mathcal{U}(-35°, 35°)$ and $\mathcal{U}(-40°, 40°)$, respectively. The three levels of contrast noise were 30% contrast noise, 45% contrast noise and 60% contrast noise. The 30% contrast noise condition was obtained by the addition of a random draw from a $\mathcal{U}(-15\%, 15\%)$ distribution to the luminance of each pixel in a square field. Baseline luminance was 50%. The other contrast noise conditions were obtained similarly. Each participant was shown 90 trials from each bar rotation-noise condition combination.

The bar rotation (BR) variance manipulation was hypothesized to modulate each participant's diffusion coefficient since the participant would have more variable information in harder trials. Considering each bar's rotation as a unit of information contributing to a "left" or "right" response, information would be more variable in trials that sampled the BRs from wider uniform distributions. It was thought that contrast noise would degrade

FIGURE 2 | The time course of one trial of the experimental stimulus. The participant first fixated on a black cross for 750 ms indicating the beginning of a trial. The participant then observed visual contrast noise changing at 8 Hz for 750 ms while maintaining fixation. A circular field of small oriented bars flickering at 15 Hz overlaid on the changing visual noise was then shown to the participant for 1000–2000 ms. The task was to indicate during this response interval whether the bars were on average oriented toward the "top-right" (45° from the horizontal line; as in this example) or "top-left" (135°) corners. It was assumed that the participant's decision making process began at the start of the response interval. After the response interval, the fixation cross was shown in isolation for 250 ms to alert the participant that the trial was over and to collect remaining responses.

the amount of information each bar gave to the decision process thus leading to smaller drift rates in trials with higher noise contrast.

2.3. BEHAVIOR AND EEG COLLECTION

Participants first completed a training session of 36 trials each. Participants were asked to complete a second training set if their percentage accuracy was subjectively judged by the experimenter to not converge to a stable value. Each participant completed 6 blocks of 90 trials each for a total of 540 trials with breaks between each block of variable time. Each trial lasted randomly (uniformly) from 2.75–3.75 s. Participants were asked to respond during the 1–2 s response interval as accurately as possible, with no-answer trials considered as incorrect. To maintain participant performance, auditory feedback was given after the response interval to the alert the participant if they were correct or incorrect. Performance feedback was also provided between blocks by displaying on the screen the percentage of trials answered correctly in that block. The behavioral data consists of each participant's accuracy and RT during each trial.

High-density electroencephalography (EEG) was collected using Electrical Geodesics, Inc.'s 128-channel Geodesic Sensor Net and Advanced Neuro Technology's amplifier with electrodes sitting on the participant's scalp throughout the duration of the experiment. Electrical activity from the scalp was recorded at a sampling rate of 1024 samples per second with an online average reference using Advanced Neuro Technology's digitization software. The EEG data was then imported into MATLAB for offline analysis.

Linear trends were removed from the EEG data. As we were only interested in 1–50 Hz EEG, the following filters were applied to each channel: (1) A high pass Butterworth filter with a 1 Hz pass band with 1 dB ripple and 0.25 Hz stop band with 10 dB attenuation, (2) a stopband Butterworth filter with 59 and 61 Hz pass bands with 1 dB ripple and 59.9–60.1 Hz stop band with 10 dB attenuation (to remove power-line noise), and (3) a low pass Butterworth filter with a 50 Hz pass band with 1 dB ripple and 60 Hz stop band with 10 dB attenuation. Artifactual data thought to be generated by phenomena outside of the cortex were removed from the EEG data using a paradigm involving Independent Component Analysis (ICA): First, any trials or channels were rejected that had time-courses unusual for cortical activity and/or had properties that ICA is deemed to not extract well, such as trials with high frequency activity indicative of muscle activity, trials or channels with high 60 Hz amplitude indicative of power-line noise suggesting poor electrode-to-skin connection, or trials with sudden high amplitude peaks that cannot be generated by cortical activity (Delorme et al., 2007). Second, ICA was used to remove linear mixtures of channel time-courses that did not subjectively correspond to EEG data in spatial map on the scalp, in power spectrum, and/or in event-related potential (ERP). Typical artifactual components include: those components with spatial maps of highly weighted electrodes near the eyes suggestive of eye movements, those components with high amplitudes at high frequencies and low amplitudes at low frequencies suggestive of muscle activity, and spatial maps of highly weighted

singular electrodes suggestive of poor electrode-scalp connectivity. A final cleaning step was performed by rejecting any trials that had high amplitudes not typical of cortical electrical activity.

For each participant, steady-state visual evoked potentials (SSVEPs) to the visual noise and signal (the circular bar field) were found at each electrode. In this experiment a steady-state response was defined by the consistency in phase at the frequencies of the stimulus (8 and 15 Hz) and the harmonic frequencies of the stimulus (16, 24, 32, 40, 48, 30, and 45 Hz). The uniformity of phase across trials was measured by the Phase Locking Index (PLI) across trials. The PLI is a statistical characterization of phase synchronization resulting from an experimental stimulus and has been shown to be successful in characterizing cortical signals (Rosenblum et al., 1996; Sazonov et al., 2009). The PLI ignores signal amplitude and ranges from 0 (all trials out-of-phase) to 1 (all trials in-phase; Tallon-Baudry et al., 1996). The equation used for PLI is provided in Equation 2. PLI is the average of $\approx$ 540 trials of amplitude normalized Fourier coefficients of the time interval. For each electrode e and participant j, PLI is defined as a function of frequency f.

$$\mathrm{PLI}_{ej}(f) = \left| \frac{1}{540} \sum_{i=1}^{540} \frac{F_{iej}(f)}{|F_{iej}(f)|} \right| \tag{2}$$

The steady-state responses to the visual noise were analyzed based on both the 750 ms noise interval and the first 1000 ms of the response interval while the steady-state responses to the signal were analyzed based only on the first 1000 ms of the response interval. Because steady-state responses located in parietal electrodes have been successfully related to attentional mechanisms in past studies (Ding et al., 2006; Bridwell and Srinivasan, 2012), electrical activity at parietal electrodes was hypothesized to be most descriptive of cognitive processes in the visual decision making task. The subject mean PLI at all frequencies averaged over parietal channels is shown in **Figure 3**. Topographic maps of the distribution of the PLI are shown at the fundamental and first two harmonics for signal and noise frequencies. It is clear that the SSVEP is broadly distributed over frontal, parietal, and occipital networks, as has been found in other studies (Ding et al., 2006; Bridwell and Srinivasan, 2012; Krishnan et al., 2013). The mean PLIs over prefrontal, frontal, central, parietal, and occipital electrode groups for each of the evoked frequencies were used as predictors in the model.

We expect the evoked cortical networks to change dependent upon the flicker frequencies of the stimulus (Ding et al., 2006; Bridwell and Srinivasan, 2012), as shown by the stimulus response in **Figure 3** where the spatial distributions of the fundamental and harmonic responses are quite different. However, we do not expect the behavior of these harmonics to be uncorrelated. To avoid multicollinearity, we performed two principal components analyses (PCAs; on the noise and signal frequencies separately) to obtain a smaller number of PLI measures from uncorrelated cortical networks. The first PCA reduced 60 PLI variables (5 cortical locations by 6 noise harmonics in both the noise and response intervals) to 16 principal components.
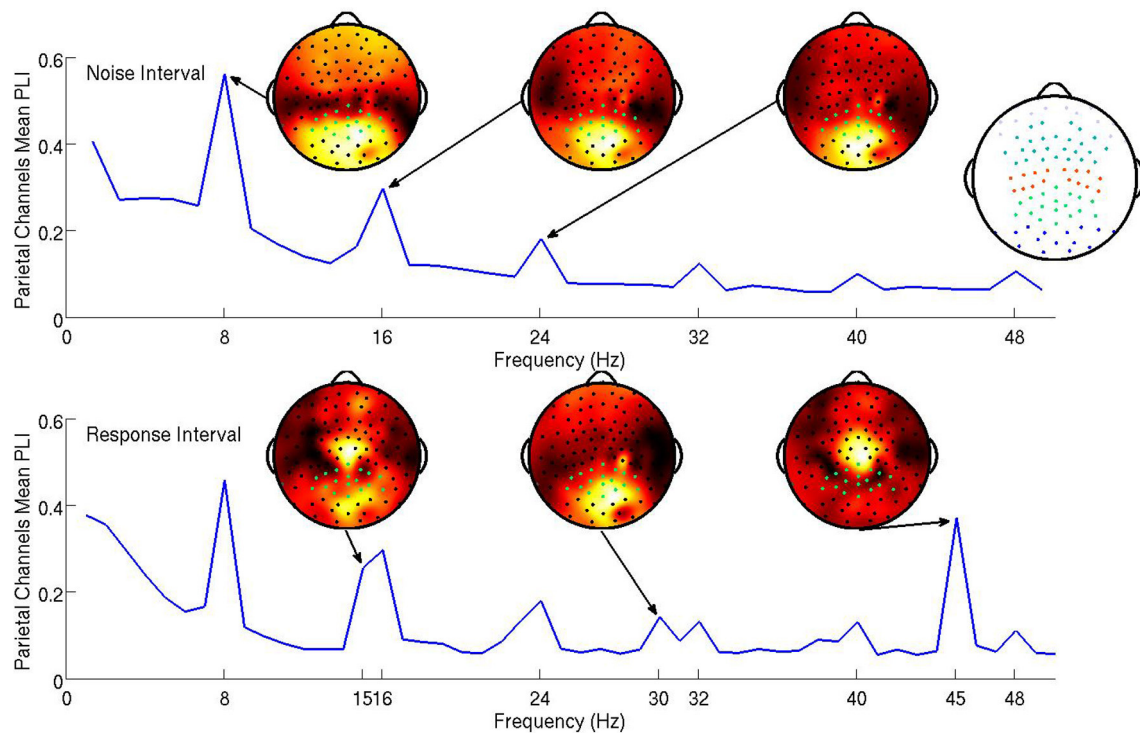
FIGURE 3 | The subject mean parietal channel PLI at all frequencies in the noise interval (top panel) and the response interval (bottom). The resolution of the the PLI spectra in the top plot is approximately 1.3 Hz due to the PLI being a function of Fourier transforms of 750 ms epochs. The resolution of the PLI spectra in the bottom plot is 1 Hz as the Fourier transforms are of 1000 ms epochs. The 15 and 16 steady-state responses during the response interval are separable when using 1000 ms epochs. Also shown are subject mean PLI topographies (at 8, 16, and 24 Hz during the noise interval and 15, 30, and 45 Hz during the response interval, each on a standardized scale) indicating where the maximum subject mean PLI is located on the scalp in relation to the parietal electrodes (highlighted green). It is clear from these topographies that using only parietal electrodes will not capture all of the steady-state response information. An index of electrode locations is also provided in the top right. Prefrontal, frontal, central, parietal, and occipital electrode groups are colored light blue, teal, orange, green, and blue, respectively.

The second PCA transformed 15 PLI variables (5 cortical locations by 3 signal harmonics) to 15 principal components. Our criteria for which principal components to include in the hierarchical Bayesian models were (1) based upon the improvement of in-sample predictive power as we increased the number of principal components, resulting in candidate principal components and (2) then based upon the out-of-sample predictive power of the candidate principal components.

2.4. HIERARCHICAL BAYESIAN MODELS

All trials from every participant were used for model fitting except those trials in which there was deemed to be EEG artifact and those trials during which the participant made no response or responded more than once. Since our models do not account for non-decision making trials, exceedingly fast trials (faster than 250 ms) were excluded as well.

The marginal likelihood for the model—that is, the predicted distribution of the data conditional on all parameters—is the first passage time distribution of a Wiener process with constant drift. We call this probability density function the *Wiener distribution*. For each trial i, subject j, and condition k, the observed accuracy w_{ijk} and RT t_{ijk} were combined in a two-element vector $\mathbf{y}_{ijk}$. These values were then assumed to be drawn from a joint distribution:

$$\mathbf{y}_{ijk} \sim \mathcal{W}(\delta_{ijk}, \varsigma_{ijk}, \tau_{ijk}). \quad (3)$$

We applied a sequence of three models—each adding a new feature—to the data.

2.5. MODEL 1: NO INDIVIDUAL DIFFERENCES

We assumed in **Model 1** that all three diffusion model parameters were constant across participants (i.e., that all participants were *identical*), and depended only on the experimental condition k. The diffusion model was fit to the RT and accuracy data of all 17 participants under the assumption that all participants had the same drift rate δ_k, diffusion coefficient ς_k, and non-decision time τ_k that were variable across condition k but not variable across participant j. Here k denotes both the particular BR condition and the particular contrast noise condition-level, $k = 1, \ldots, 9$. A graphical representation of **Model 1** is provided in **Figure 4A**.

The assumptions of the model, together with the prior distributions for the parameters, appear below. The priors for the drift rate δ_k and non-decision time τ_k were truncated normal distributions due to the knowledge of the natural constraints of the diffusion model and prior knowledge of acceptable values for similar tasks. Note that the second parameter of the normal distributions below represent the variance.
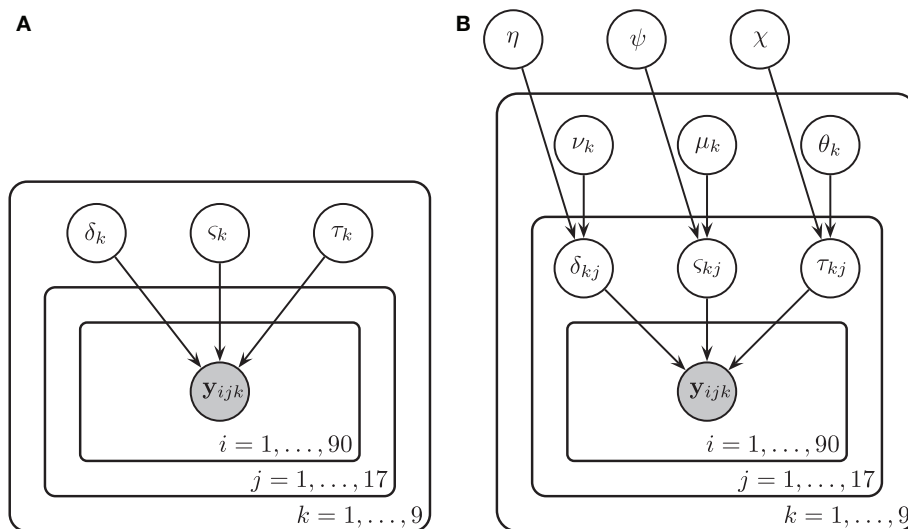
FIGURE 4 | A graphical representation of Model 1 (A) and Model 2 (B). In **Model 1**, drift rates δ_k, diffusion coefficients ς_k, and non-decision times τ_k were assumed to vary over conditions k but remain invariant across participants j and trials i. There were three bar rotation conditions and three contrast noise conditions. Here k denotes each bar rotation and contrast noise pair. In **Model 2**, drift rates δ_{jk}, diffusion coefficients ς_{jk}, and non-decision times τ_{jk} were assumed to vary over both conditions and participants. Each of these parameters are in turn assumed to be drawn from normal distributions with means that varied over conditions k and with variances that did not vary across conditions.

$$\delta_{jk} = \delta_k, \quad \delta_k \sim \mathcal{N}(0.0, 5) \in (-9, 9)$$
$$\varsigma_{jk} = \varsigma_k, \quad \varsigma_k \sim \mathcal{N}(0.5, 4)$$
$$\tau_{jk} = \tau_k, \quad \tau_k \sim \mathcal{N}(0.3, 4) \in (0, 1)$$

2.6. MODEL 2: INDIVIDUAL DIFFERENCES

In **Model 2** we assumed that participants differ but are draws from a single superordinate population (i.e., participants are *exchangeable*). Consequently, the drift rate δ_{jk}, diffusion coefficient ς_{jk}, and non-decision time τ_{jk} varied by both subject j and condition k. Subject-level parameters were assumed to be drawn from normal distributions with means that were variable over condition only. Variances were assumed to be invariant across conditions to maintain model simplicity (i.e., the model assumes *homoscedasticity* in the parameters). The prior distributions of the parameters are listed below.

$$\left(\delta_{jk} \mid \nu_k, \eta\right) \sim \mathcal{N}(\nu_k, \eta) \in (-9, 9), \quad \nu_k \sim \mathcal{N}(0.0, 5), \quad \eta \sim \Gamma(6, 0.10)$$
$$\left(\varsigma_{jk} \mid \mu_k, \psi\right) \sim \mathcal{N}(\mu_k, \psi), \quad\quad\quad \mu_k \sim \mathcal{N}(0.5, 4), \quad \psi \sim \Gamma(4, 0.05)$$
$$\left(\tau_{jk} \mid \theta_k, \chi\right) \sim \mathcal{N}(\theta_k, \chi) \in (0, 1), \quad \theta_k \sim \mathcal{N}(0.3, 4), \quad \chi \sim \Gamma(5, 0.01)$$

A graphical representation of **Model 2** is provided in **Figure 4B**.

2.7. MODEL 3: INDIVIDUAL DIFFERENCES WITH NEURAL CORRELATES

With **Model 3**, we will attempt to explain any individual differences in cognitive parameters by introducing the neural data as explanatory variables. The model is similar to **Model 2**, but additionally includes a regression structure to explain variability in subject-level model parameters with steady-state PLI values.

In order to avoid multicollinearity, PLIs were first subjected to a principal component analysis (PCA), and the resultant independent components were used as predictors. The PCA was performed on the noise and signal frequencies separately. The first PCA reduced 60 PLI variables to 16 principal components and the second PCA transformed 15 PLI variables into 15 components. The criterion used to determine which principal components to include was the out-of-sample predictive power of each model. Predictive power was measured as R^2_{pred}, a measure of the percentage of total between-subject variance explained, in this case of the correct-RT medians of each condition. The equation used for R^2_{pred} is provided in the Supplemental Materials.
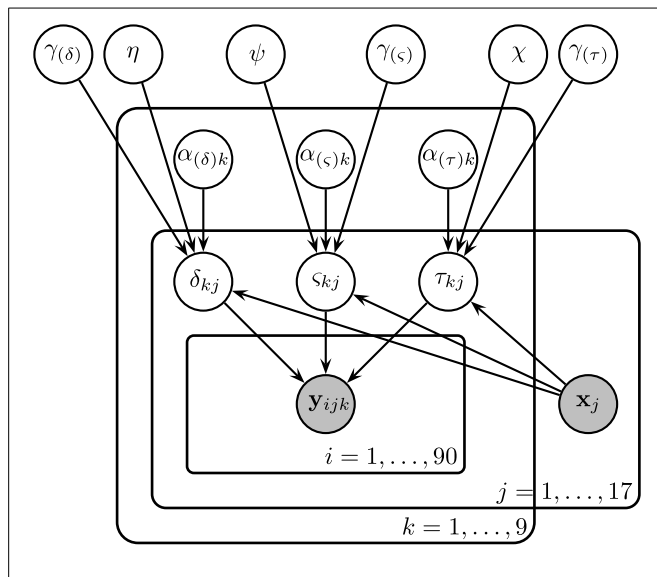
Subject-level drift rates δ_{jk}, diffusion coefficients ς_{jk}, and non-decision times τ_{jk} were assumed to be drawn from normal distributions with means of the form $\alpha_k + \boldsymbol{x}_j^\mathsf{T} \boldsymbol{\gamma}$ where α_k is condition k's effect on the subject-level cognitive parameter, $\boldsymbol{x}_j$ is a vector of principal components, and $\boldsymbol{\gamma}$ is a vector of regression coefficients (i.e., the effect of each principal component on the cognitive parameter). The graphical representation of the model is provided in **Figure 5**. The priors of the variance parameters are the same as in **Model 2**. Weakly informative prior distributions of $\mathcal{N}(0.0, 10)$ were given to the weight variables that make up the vectors $\boldsymbol{\gamma}_{(\delta)}$, $\boldsymbol{\gamma}_{(\varsigma)}$, and $\boldsymbol{\gamma}_{(\tau)}$. The other hyperpriors and priors were:

$$\left(\delta_{jk} \mid \alpha_{(\delta)k}, \boldsymbol{\gamma}_{(\delta)}, \eta\right) \sim \mathcal{N}\left(\alpha_{(\delta)k} + \boldsymbol{x}_j^\mathsf{T} \boldsymbol{\gamma}_{(\delta)}, \eta\right) \in (-9, 9), \quad \alpha_{(\delta)k} \sim \mathcal{N}(0.0, 5)$$
$$\left(\varsigma_{jk} \mid \alpha_{(\varsigma)k}, \boldsymbol{\gamma}_{(\varsigma)}, \psi\right) \sim \mathcal{N}\left(\alpha_{(\varsigma)k} + \boldsymbol{x}_j^\mathsf{T} \boldsymbol{\gamma}_{(\varsigma)}, \psi\right), \quad\quad\quad\quad \alpha_{(\varsigma)k} \sim \mathcal{N}(0.5, 4)$$
$$\left(\tau_{jk} \mid \alpha_{(\tau)k}, \boldsymbol{\gamma}_{(\tau)}, \chi\right) \sim \mathcal{N}\left(\alpha_{(\tau)k} + \boldsymbol{x}_j^\mathsf{T} \boldsymbol{\gamma}_{(\tau)}, \chi\right) \in (0, 1), \quad\quad \alpha_{(\tau)k} \sim \mathcal{N}(0.3, 4)$$

2.8. POSTERIOR SAMPLING

We used the JAGS software (Plummer, 2003) to analyze the data by drawing samples from the joint posterior distribution of the parameters of the hierarchical models. To compute the likelihood function associated with the assumed decision making process (the Wiener distribution), we used the *jags-wiener*

FIGURE 5 | Graphical representation of Model 3. Drift rates δ_{jk}, diffusion coefficients ς_{jk}, and non-decision times τ_{jk} were assumed to vary over both conditions and participants. Each of these parameters are assumed to be drawn from normal distributions with means of the form $\alpha_k + x_j^\mathsf{T} \gamma$, where x_j is the vector of SSVEP responses of subject j, and with variances that did not vary across conditions. As an example, $\alpha_{(\tau)k}$ is the condition effect on the non-decision time and $\gamma_{(\tau)}$ reflects the change in non-decision time (seconds) due to a one SSVEP unit difference across two participants.

module (Wabersich and Vandekerckhove, 2013). This allowed us to explain accuracy and response time distributions within conditions and across subjects. For each model, samples from the posterior distributions of the parameters were found by running JAGS with six Markov Chain Monte Carlo (MCMC) chains of length 21000, with 1000 burn-in (discarded) samples and a thinning parameter of 10 (keeping only every 10th sample) resulting in six joint posterior distribution estimates of 2000 samples each. We used the $\hat{R}$ statistic to compare within-chain variance to between-chain variance in order to assess convergence of the MCMC algorithm (Gelman and Rubin, 1992).

2.9. POSTERIOR PREDICTIVE DISTRIBUTIONS

To quantify model fit, in-sample posterior predictive distributions of accuracy-RTs from 5000 simulated experiments were estimated by sampling from the posterior distributions of subject-level parameters for each of the three models. That is, $s = 1, \ldots, 5000$ samples were randomly drawn from the subject-level posterior distributions of the model parameters producing 5000×1 column vectors for each drift rate $\delta_{jk}^{(s)}$, diffusion coefficient $\varsigma_{jk}^{(s)}$, and non-decision time $\tau_{jk}^{(s)}$. The samples $\left(\delta_{jk}^{(s)}, \varsigma_{jk}^{(s)}, \tau_{jk}^{(s)}\right)$ were used to generate accuracy-RT samples from the Wiener distribution [with the rejection sampling algorithm described in Tuerlinckx et al. (2001)].

In order to find candidate PLI predictors for **Model 3** and also to gauge the ability of each model type to predict new subjects' behavioral data, *in-sample* and *out-of-sample* posterior predictive distributions were generated using the PLI coefficients and posterior distributions of the *condition*-level parameters to find

predictive distributions of the *subject*-level parameters. This procedure does not use samples from the subject-level posterior distributions directly, but estimates the subject-level parameters from the posteriors of the condition-level parameters and EEG covariates before finding a posterior predictive distribution of accuracy-RTs. Samples from the posterior predictive distribution of subject j's mean drift rate on a trial in condition k are drawn from a normal distribution with mean $\boldsymbol{\alpha}_{(\delta)k}^{(s)} + \mathbf{G}_{(\delta)}^{(s)}\boldsymbol{x}_j$ where $\boldsymbol{x}_j$ is the vector of subject j's principal component PLI values, $\boldsymbol{\alpha}_{(\delta)k}^{(s)}$ are samples from the posterior distribution of condition k's effect on drift rate, and $\mathbf{G}_{(\delta)}^{(s)}$ is a matrix consisting of samples from the posterior distributions of the PLI coefficients for drift rate. For in-sample prediction, we fit different possible forms of **Model 3**, with different numbers of principal components, 17 times each to generate in-sample posterior distributions to find candidate principal components. Then for out-of-sample prediction, we fit different possible forms of **Model 3**, with the resulting candidate principal components, 17 times with each participant removed from the data set. In the previously mentioned example, both the condition effect on drift rate and PLI coefficients are estimated from the model with all subjects except j for out-of-sample prediction.

3. RESULTS

For all models and all parameters, convergence of the Monte Carlo chains was satisfactory: $\hat{R} \leq 1.01$ for all parameters ($\hat{R} \geq 1.10$ is conventionally taken as evidence for non-convergence; Gelman and Rubin, 1992).

3.1. MODEL 1: NO INDIVIDUAL DIFFERENCES

Marginal posterior distributions of the parameters of **Model 1** are plotted in the Supplemental Materials' Figure 8. The variability of evidence units gained per second ς_k increased as BR variance grew. Evidence units gained per second, drift rate δ_k, was found to decrease both with larger contrast noise and larger BR. The parameter estimates seem to show a complex interaction effect of BR and contrast noise on non-decision time τ_k. However, the results from **Model 2** will indicate that **Model 1** is sufficiently misspecified that this interaction cannot be interpreted in a meaningful way.

3.2. MODEL 2: INDIVIDUAL DIFFERENCES

The marginal posterior distributions of the condition-level parameters are shown in Figure 8 of the Supplementary Materials. At the condition level, the effects of the experimental manipulations on drift rate and the diffusion coefficient remain similar to the results of **Model 1**: Mean drift rates ν_k were found to decrease as BR variance grew, smaller mean drift rates were observed in the high visual noise condition, and mean diffusion coefficients μ_k increased as BR variance grew. Main effects on the condition-level non-decision time not clearly observable in **Model 1** were found in **Model 2**. Mean non-decision time θ_k was slow when the BR variance was high, and participants were estimated to have quick non-decision times in low visual noise conditions.

The complex interactive pattern of non-decision times obtained in **Model 1** no longer appears.

By adding subject-level parameters, the current model not only provides a clearer picture of condition-level behavior of all

participants, but describes the *individual differences* of the participants modeled by the subject-level parameters, δ_{jk}, ς_{jk}, and τ_{jk}. Posterior distributions for the subject-level parameters of the easiest condition ($\pm 30°$ BR and 30% noise) are provided in the Supplemental Materials' Figure 9. Due to subject-level parameters deviating from the condition-level parameter's means, this model is able to predict within-sample data well-compared to the previous model. Percent variances explained (R^2_{pred}) of correct-RT subject medians by within-sample posterior prediction are provided in **Table 1**. **Model 2** explains at least 86.3% of median correct-RT between-subject variance in each condition.

3.3. MODEL 3: INDIVIDUAL DIFFERENCES WITH NEURAL CORRELATES

The results of **Model 2** clearly demonstrate differences between participants' cognition in the perceptual decision making task. We were further able to explain the differences in the cognitive variables using the neural data: **Model 3** was fit in a similar manner to **Model 2**, but additionally included principal components of the steady-state PLIs as regressors, as represented by the vector x_j, on the subject-level model parameters.

We generated in-sample posterior predictive distributions using condition-level parameter posterior distributions (as opposed to in-sample posterior prediction from subject-level parameters), PLI coefficient posterior distributions, and PLI variables from each subject to find principal components that best predicted correct RT distributions. A plot of in-sample unexplained median correct-RT between-subject variance as a decreasing function of number of principal component (PC) regressors included in the model is provided in Figure 10 of the Supplemental Materials. Based on this analysis, PCs 2, 4, and 7 of both the noise and signal sets were tested further to find the model that best predicted out-of-sample RT of correct responses.

Model 3 was the model that best predicted out-of-sample correct-RT distributions by using noise component 2 and signal component 7 as exogenous PLI regressors on the diffusion model parameters. It should be noted that the amount of variance of the original PLI data explained by each PC is not reflective of each PC's out-of-sample predictive power, just as the amount of variance of the original data explained by each PC is not reflective of its contribution to the model (Jolliffe, 1982). A table of percent between-subject variance of median correct-RT explained (R^2_{pred}) by out-of-sample prediction is provided in **Table 1**. Tables of percent between-subject variance of mean, 25th percentile, and 75th percentile correct-RT explained by out-of-sample prediction are provided in the Supplemental Materials. A new paricipant's correct-RT distribution in each condition can be more accurately predicted using the participant's EEG in **Model 3**'s framework than by using **Model 1**'s or **Model 2**'s framework. 31.9% of the between-subject variance of the easiest condition's median correct-RT is explained by out-of-sample prediction.

To aid in interpretation, the posterior distributions of the regression coefficients for each PC were projected into the PLI coefficient space by multiplying the matrix of PC coefficient posterior samples **G** by the inverse-weight matrix **V** from the PCA algorithm which projects the PCs into the PLI data space. The result **GV** are samples from the posterior distributions of the regression coefficients for each PLI variable. This transformation was performed once for each of the noise and signal variable sets.

The posterior distributions of the signal PLI coefficients are provided in **Figure 6** with means, medians and 95% and 99% credible intervals. From the PC coefficient and PLI coefficient posteriors, it was clear that there is a complex signal response at multiple frequencies and cortical locations on the diffusion coefficient and non-decision time. Participants with larger signal occipital 15 and 45 Hz PLIs are expected to have smaller variances

Table 1 | Percentage of between-subject variance in correct-RT medians explained by in-sample and out-of-sample prediction (R^2_{pred}) for each experimental condition.

Rotation	Noise	In-sample prediction			Out-of-sample prediction		
		M1	M2	M3	M1	M2	M3
$\pm 30°$	30%	−0.1%	94.3%	95.0%	−13.5%	−11.8%	31.9%
$\pm 35°$	30%	−0.1%	95.6%	95.8%	−12.3%	−11.7%	27.6%
$\pm 40°$	30%	−0.5%	92.2%	92.1%	−12.5%	−11.5%	19.9%
$\pm 30°$	45%	−1.2%	86.3%	87.4%	−15.1%	−11.7%	29.4%
$\pm 35°$	45%	−0.2%	92.3%	91.6%	−12.0%	−13.6%	22.8%
$\pm 40°$	45%	0.2%	92.6%	91.9%	−11.9%	−15.0%	28.0%
$\pm 30°$	60%	−0.7%	93.1%	92.8%	−12.9%	−13.0%	18.6%
$\pm 35°$	60%	−2.3%	92.5%	92.6%	−14.7%	−13.5%	13.3%
$\pm 40°$	60%	−0.6%	90.9%	91.2%	−13.8%	−18.0%	26.2%

*The in-sample predictive ability of the no-individual differences **Model 1** was unsurprisingly poor, while the in-sample predictive ability of individual differences models (with and without EEG regressors, **Model 2** and **Model 3**, respectively) explained most of the variance of correct-RT subject medians. Out-of-sample prediction was performed by using an iterative leave-one-subject-out procedure, first by obtaining posterior distribution estimates for each parameter by modeling all but one participant's behavior and EEG data and then estimating the left-out participant's correct-RT distribution using the resulting model fit and the left-out participant's EEG. Models without EEG regressors (i.e., **Model 1** and **Model 2**) are poor choices for new participant behavior prediction. The model with a noise principal component and a signal principal component of the phase-locked EEG as covariates of diffusion model parameters (**Model 3**) more accurately predicts new participants' correct-RT behavior. Negative values indicate overdispersion of the model prediction (due to posterior uncertainty) relative to the real data.*
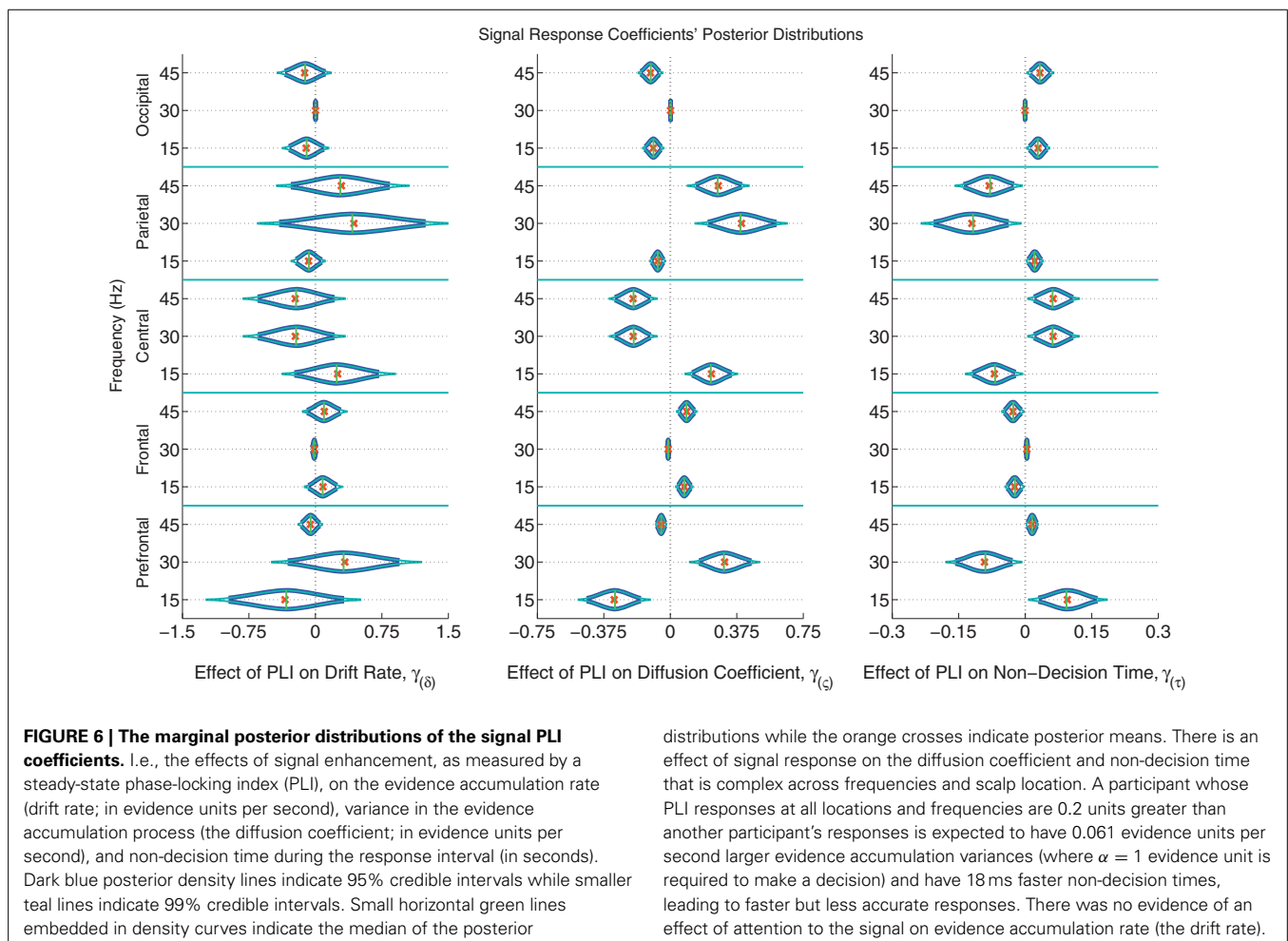
in the evidence accumulation process (diffusion coefficients) than those participants with smaller occipital signal PLIs. However, the opposite effect is found in the frontal electrodes with large 15 and 45 Hz PLIs being associated with larger evidence accumulation variances. Larger signal responses at 30 and 45 Hz in parietal electrodes is also associated with larger diffusion coefficients. The effect of signal response on non-decision time is also complex but closely related to the effect of signal response on the diffusion coefficient. No evidence of an association between participants' differences in signal response to differences in evidence accumulation rates (drift rates) was found.
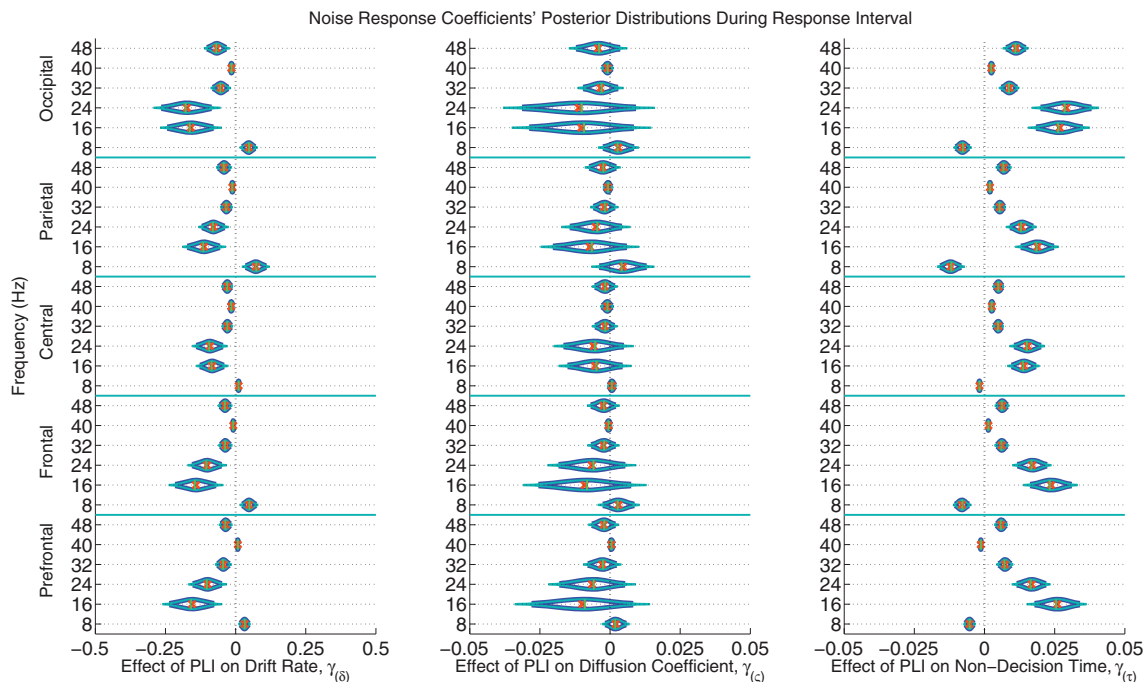
The posterior distributions of the noise PLI coefficients from the response interval are provided in **Figure 7**. The posterior distributions of the noise PLI coefficients from the noise interval are provided in the Supplemental Materials' Figure 11. In all noise harmonic frequencies during the noise interval and most harmonic frequencies (16, 24, 32, and 48 Hz) during the response interval, those subjects who had smaller PLIs at all electrode locations had faster evidence accumulation rates (drift rates). This finding suggests that those subjects who better suppressed the stimulus noise accumulated correct evidence faster. Furthermore, a similar effect was found on non-decision time. Noise suppression in the harmonic frequencies was associated with smaller

non-decision times across subjects. However, smaller PLIs at 8 Hz were associated with slower evidence accumulation and faster non-decision times. Looking at these effects as a whole, those subjects with more suppressed responses to the noise at all frequencies had larger drift rates and smaller non-decision times leading to faster, more accurate responses. As a plausible but oversimplified example, a participant whose PLI responses at all frequencies and locations was suppressed 0.2 units more than another participant during both the noise and response intervals is expected to accumulate 0.418 evidence units per second faster than another participant and have a 70 ms faster non-decision time. There was little to no evidence of an effect of individual variation in brain responses to noise on within-trial evidence accumulation variability (the diffusion coefficient).

4. DISCUSSION

We have shown that a Bayesian diffusion model framework with hierarchical participant-level parameters is useful in describing individual differences in the rate of evidence accumulation, variance in evidence accumulation process, and preprocessing and/or motor response time in a novel perceptual decision making paradigm. Assuming the model describes the relationship between cognition and behavior sufficiently well, we are able to

FIGURE 6 | The marginal posterior distributions of the signal PLI coefficients. I.e., the effects of signal enhancement, as measured by a steady-state phase-locking index (PLI), on the evidence accumulation rate (drift rate; in evidence units per second), variance in the evidence accumulation process (the diffusion coefficient; in evidence units per second), and non-decision time during the response interval (in seconds). Dark blue posterior density lines indicate 95% credible intervals while smaller teal lines indicate 99% credible intervals. Small horizontal green lines embedded in density curves indicate the median of the posterior

distributions while the orange crosses indicate posterior means. There is an effect of signal response on the diffusion coefficient and non-decision time that is complex across frequencies and scalp location. A participant whose PLI responses at all locations and frequencies are 0.2 units greater than another participant's responses is expected to have 0.061 evidence units per second larger evidence accumulation variances (where $\alpha = 1$ evidence unit is required to make a decision) and have 18 ms faster non-decision times, leading to faster but less accurate responses. There was no evidence of an effect of attention to the signal on evidence accumulation rate (the drift rate).

FIGURE 7 | The marginal posterior distributions of the noise PLI coefficients. I.e., the effects of noise suppression, as measured by a steady-state phase-locking index (PLI), on the evidence accumulation rate (drift rate; in evidence units per second), variance in the evidence accumulation process (the diffusion coefficient; in evidence units per second), and non-decision time (in seconds) during the response interval. Dark blue lines indicate 95% credible intervals, smaller teal lines indicate 99% credible intervals, horizontal green lines indicate posterior medians, and the orange exes indicate posterior means. At noise harmonic frequencies (16, 24, 32, and 48 Hz) during the response interval, those subjects who suppressed noise had faster evidence accumulation rates; this effect was

found at all electrode groups. However, noise enhancement at 8 Hz was associated with slower evidence accumulation. Furthermore, those subjects who better suppressed noise at the same harmonic frequencies had faster non-decision times. For example, a participant whose PLI responses were suppressed 0.2 units more than another participant's responses at all locations and frequencies during the response interval is expected to accumulate 0.288 evidence units per second faster (where $\alpha = 1$ evidence unit is required to make a decision) and have 48 ms faster non-decision times, leading to faster and more correct responses. There was no evidence of an effect of attention to the visual noise on variance in evidence accumulation (the diffusion coefficient).

infer cognitive differences among participants. Furthermore, we have shown that differences in participants' attention as measured by SSVEPs relate to some of these differences in participants' cognition.

Individual differences in the rates of evidence accumulation (drift rates) were partially explained by individual differences in noise suppression as measured by SSVEPs. Participants who better suppressed noise at high frequencies during the both the preparatory period (noise interval) and the decision period (response interval) were able to accumulate correct evidence faster, which led to more accurate, faster response times. Furthermore, those individuals who better suppressed noise in the same frequency bands and locations had faster non-decision times (preprocessing and/or motor response speed). This effect on non-decision time is hypothesized to be reflective of faster preprocessing time in subjects who better suppressed noise since we do not expect noise suppression to affect motor response speed. Both findings suggest a role of noise suppression in beta and gamma EEG frequency bands on the speed of evidence accumulation and preprocessing prior to evidence accumulation in perceptual decision making tasks.

Enhancement of signal was found to describe individual variation in "randomness" of evidence accumulation within trials (as measured by the diffusion coefficient). Participants who did not properly enhance signal in occipital, central, and pre-frontal electrodes had the most variable evidence accumulation processes. There is also evidence that a participant's enhancement of signal may have affected their preprocessing time in a complex way across frequencies and cortical locations. This suggests that signal enhancement in beta and gamma EEG frequency bands affect within-trial evidence accumulation variance and preprocessing in perceptual decision making.

In summary, from the results of the modeling procedure it was found that some individual variation in evidence accumulation speed (drift rate) is explained by noise suppression, some individual variation in evidence accumulation variance (diffusion coefficient) is explained by signal enhancement, and some individual variation in non-decision time (presumably preprocessing time) is explained by both noise suppression and signal enhancement.

The usefulness of the model with SSVEP attention measures as regressors is not only in its descriptive ability, but also in its

predictive ability. New subject correct-RT behavior was not accurately described by the model without individual differences nor the model with individual differences. But by explicitly including individual differences with neural covariates in hierarchical models, the correct RT distributions of new subjects with known neural measures are more accurately predicted. We expect the addition of the phase-locking index of SSVEPs to be predictive of behavior in any perceptual decision making paradigm, especially if used in a hierarchical Bayesian framework. Theoretically the hierarchical EEG-diffusion model will also be able to predict the PLI measures of a missing participant given a participant's behavioral data. We will explore the practicality of such predictions in future studies. Possible applications of behavioral and neural data prediction include: (a) the ability to interpolate data from incomplete behavioral data sets (b) the ability to interpolate data from incomplete neural data sets (c) more powerful statistical inference through simultaneous accounting for changes in behavior and neural data.

In the future for both hypothesis testing and response-RT prediction, latent variables linearly or non-linearly related to the EEG covariates can be included with the cognitive model in a hierarchical Bayesian framework (see Vandekerckhove, 2014, for details). The benefits of such an analysis would be: to choose neural covariates maximally descriptive or predictive of the data, choose electrodes and frequencies maximally descriptive or predictive of the data, reduce the number of covariates, and reduce the multicollinearity of the covariates by assuming there exist underlying variables related to multiple EEG covariates. In the present study, the problems of multicollinearity and variable overabundance were overcome with two principal component analyses (PCAs). PCAs do not extract mixtures of the data which are most descriptive or predictive of the model parameters but instead extract mixtures of the data which are uncorrelated. A shortcoming of this study is that we did not pick frequencies and cortical locations that were maximally predictive of behavior as exogenous variables. Cortical locations naively based upon large non-focal groupings were chosen. Instead of performing a non-Bayesian PCA before submitting the neural data to the Bayesian algorithm, a linear mixture of neural data that best describes the cognitive model parameters could be extracted from the Bayesian algorithm itself, analogous to a partial least squares regression in a non-Bayesian approach (see Krishnan et al., 2013, for an example). In order to use this latent variable technique, the model must be run on a training set using a subset of the EEG data and then run on a test set to measure out-of-sample model predictive ability. This would result in a data reduction of the EEG that best predicts behavior in the context of the model.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2015.00018/abstract

REFERENCES

Batchelder, W. H. (2010). "Cognitive psychometrics: using multinomial processing tree models as measurement tools," in *Measuring Psychological Constructs: Advances in Model-Based Approaches*, ed S. Embretson (Washington, DC: American Psychological Association), 71–93. doi: 10.1037/12074-004

Bridwell, D. A., and Srinivasan, R. (2012). Distinct attention networks for feature enhancement and suppression in vision. *Psychol. Sci.* 23, 1151–1158. doi: 10.1177/0956797612440099

Bridwell, D. A., Hecker, E. A., Serences, J. T., and Srinivasan, R. (2013). Individual differences in attention strategies during detection, fine discrimination, and coarse discrimination. *J. Neurophysiol.* 110, 784–794. doi: 10.1152/jn.00520.2012

Cassey, P., Gaut, G., Steyvers, M., and Brown, S. (2014). "A simple accumulator model that simultaneously accounts for behaviour and spike-train data," in *Paper Presented at the 55th Annual Meeting of the Psychonomic Society* (Long Beach, CA).

Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34, 1443–1449. doi: 10.1016/j.neuroimage.2006.11.004

Ding, J., Sperling, G., and Srinivasan, R. (2006). Attentional modulation of SSVEP power depends on the network tagged by the flicker frequency. *Cereb. Cortex* 16, 1016–1029. doi: 10.1093/cercor/bhj044

Forstmann, B. U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E.-J., et al. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15916–15920. doi: 10.1073/pnas.1004932107

Garcia, J. O., Srinivasan, R., and Serences, J. T. (2013). Near-real-time feature-selective modulations in human cortex. *Curr. Biol.* 23, 515–522. doi: 10.1016/j.cub.2013.02.013

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7, 457–472. doi: 10.1214/ss/1177011136

Jolliffe, I. T. (1982). A note on the use of principal components in regression. *Appl. Stat.* 31, 300–303. doi: 10.2307/2348005

Krishnan, L., Kang, A., Sperling, G., and Srinivasan, R. (2013). Neural strategies for selective attention distinguish fast-action video game players. *Brain Topogr.* 26, 83–97. doi: 10.1007/s10548-012-0232-3

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychon. Bull. Rev.* 15, 1–15. doi: 10.3758/PBR.15.1.1

Lee, M. D., and Newell, B. J. (2011). Using hierarchical Bayesian methods to examine the tools of decision-making. *Judg. Decis. Making* 6, 832–842.

Link, S., and Heath, R. (1975). A sequential theory of psychological discrimination. *Psychometrika* 40, 77–105. doi: 10.1007/BF02291481

Müller, M. M., Picton, T. W., Valdes-Sosa, P., Riera, J., Teder-Sälejärvi, W. A., and Hillyard, S. A. (1998). Effects of spatial selective attention on the steady-state visual evoked potential in the 20–28 hz range. *Cogn. Brain Res.* 6, 249–261. doi: 10.1016/S0926-6410(97)00036-0

Morgan, S., Hansen, J., and Hillyard, S. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proc. Natl. Acad. Sci. U.S.A.* 93, 4770–4774. doi: 10.1073/pnas.93.10.4770

O'Connell, R. G., Dockree, P. M., and Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat. Neurosci.* 15, 1729–1735. doi: 10.1038/nn.3248

Olitsky, S., Lee, H., and Young, E. (2013). IVAC—interactive visual acuity chart. Available online at: http://www.smbs.buffalo.edu/oph/ped/IVAC/IVAC.html (Accessed December 13, 2013).

Plummer, M. (2003). "JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling," in *Proceedings of the 3rd International*

Workshop on Distributed Statistical Computing (DSC 2003), (Vienna, Austria).

Ratcliff, R., Thapar, A., and McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychol. Aging* 16:323. doi: 10.1037/0882-7974.16.2.323

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85:59. doi: 10.1037/0033-295X.85.2.59

Regan, D. (1977). Steady-state evoked potentials. *JOSA* 67, 1475–1489. doi: 10.1364/JOSA.67.001475

Rosenblum, M. G., Pikovsky, A. S., and Kurths, J. (1996). Phase synchronization of chaotic oscillators. *Phys. Rev. Lett.* 76:1804. doi: 10.1103/PhysRevLett.76.1804

Rouder, J. N., Lu, J., Speckman, P., Sun, D., and Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychon. Bull. Rev.* 12, 195–223. doi: 10.3758/BF03257252

Sazonov, A. V., Ho, C. K., Bergmans, J. W., Arends, J. B., Griep, P. A., Verbitskiy, E. A., et al. (2009). An investigation of the phase locking index for measuring of interdependency of cortical source signals recorded in the EEG. *Biol. Cybern.* 100, 129–146. doi: 10.1007/s00422-008-0283-4

Stone, M. (1960). Models for choice–reaction time. *Psychometrika* 25, 251–260. doi: 10.1007/BF02289729

Tallon-Baudry, C., Bertrand, O., Delpuech, C., and Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 hz visual responses in human. *J. Neurosci.* 16, 4240–4249.

Tuerlinckx, F., Maris, E., Ratcliff, R., and De Boeck, P. (2001). A comparison of four methods for simulating the diffusion process. *Behav. Res. Methods Instrum. Comput.* 33, 443–456. doi: 10.3758/BF03195402

Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., and Steyvers, M. (2013). A bayesian framework for simultaneously modeling neural and behavioral data. *Neuroimage* 72, 193–206. doi: 10.1016/j.neuroimage.2013.01.048

Vandekerckhove, J., Tuerlinckx, F., and Lee, M. (2008). "A Bayesian approach to diffusion process models of decision-making," in *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (Austin, TX), 1429–1434.

Vandekerckhove, J., Verheyen, S., and Tuerlinckx, F. (2010). A crossed random effects diffusion model for speeded semantic categorization decisions. *Acta Psychol. (Amst)* 133, 269–282. doi: 10.1016/j.actpsy.2009.10.009

Vandekerckhove, J., Tuerlinckx, F., and Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychol. Methods* 16, 44. doi: 10.1037/a0021765

Vandekerckhove, J. (2014). A cognitive latent variable model for the simultaneous analysis of behavioral and personality data. *J. Math. Psychol.* 60, 58–71. doi: 10.1016/j.jmp.2014.06.004

Voss, A., Rothermund, K., and Voss, J. (2004). Interpreting the parameters of the diffusion model: an empirical validation. *Mem. Cogn.* 32, 1206–1220. doi: 10.3758/BF03196893

Wabersich, D., and Vandekerckhove, J. (2013). Extending JAGS: a tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behav. Res. Methods* 46, 15–28. doi: 10.3758/s13428-013-0369-3

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., and McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *J. Mem. Lang.* 58, 140–159. doi: 10.1016/j.jml.2007.04.006

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the ratcliff diffusion model of response times and accuracy. *Eur. J. Cogn. Psychol.* 21, 641–671. doi: 10.1080/09541440802205067

Individual differences in Zhong-Yong tendency and processing capacity

Ting-Yun Chang and Cheng-Ta Yang*

Department of Psychology, National Cheng Kung University, Tainan, Taiwan

The present study investigated how an individual's Zhong-Yong tendency is related to his/her perceptual processing capacity. In two experiments, participants completed a Zhong-Yong Thinking Style Scale and performed a redundant-target detection task. Processing capacity was assessed with a non-parametric approach (systems factorial technology, SFT) and a parametric (linear ballistic accumulator model, LBA) approach. Results converged to suggest a positive correlation between Zhong-Yong tendency and processing capacity. High middle-way thinkers had larger processing capacity in multiple-signal processing compared with low middle-way thinkers, indicating that they processed information more efficiently and in an integrated fashion. Zhong-Yong tendency positively correlates with the processing capacity. These findings suggest that the individual differences in processing capacity can account for the reasons why high middle-way thinkers tend to adopt a global and flexible processing strategy to deal with the external world. Furthermore, the influence of culturally dictated thinking style on cognition can be revealed in a perception task.

Keywords: individual differences, linear ballistic accumulator model, systems factorial technology, workload capacity, Zhong-Yong

INTRODUCTION

People in different cultures differ psychologically, and they know different things, believe different things, and have different tastes. An increasing number of studies have investigated whether culture affects an individual's behavior and recent findings show that culture plays an important role in shaping human perception and cognition (Norenzayan and Nisbett, 2000; Masuda and Nisbett, 2001, 2006; Kitayama et al., 2003; Nisbett and Miyamoto, 2005; Miyamoto et al., 2006). Although it is still unclear whether this cultural influence is a result of collective unconsciousness, which is inherited through genes, or cumulative learning of the cultures, within-culture and cross-culture comparisons reveal the within- and between-cultural variation and reveal how human behavior is affected by social-cultural factors. The present study focuses on one of the most influential Chinese thinking styles, Zhong-Yong thinking style, to see how it affects the processes in perceptual decision making.

Middle-way thinking, also known as *Zhong-Yong* in Chinese, is a culturally dictated thinking style originating from Confucian philosophy. Being without inclination to either side is called *Zhong*; admitting of no change is called *Yong*. Zhong-Yong, the law of mind, was handed down from one to another in the Confucian school, until Tsze-Sze wrote a book chapter titled "*The Doctrine of the Mean*." In *The Doctrine of the Mean*, the state of "equilibrium" and the state of "harmony" are emphasized and people are encouraged to achieve these mind states. In Chapter 1, Tsze-Sze states that "*While there are no stirrings of pleasure, anger, sorrow, or joy, the mind may be said to be in the state of equilibrium. When those feelings have been stirred, and they act*

in their due degree, there ensues what may be called the state of harmony. Equilibrium is the great root from which grow all the human acting in the world, and harmony is the universal path which they all should pursue." Also, written in the *Analects of Confucius*, the cognitive style of "middle-way" is described as the rule of thumb to deal with things and get along with other people. By a simplified definition, Zhong-Yong emphasizes that one should "…consider things carefully from different perspectives, avoid going to extremes, behave in situationally appropriate ways, and maintain interpersonal harmony…" (Ji et al., 2010). Middle-way thinking is regarded as a "good" individual attribute that the Chinese praise and pursue, and it has a major impact on Chinese daily life (see Yang, 2010 for a review).

Since C.-F. Yang and C.-Y. Zhao initiated a project to study different aspects of Zhong-Yong thinking in the early 1990s, an increasing number of studies have used the Zhong-Yong Thinking Style Scale (Chiu, 2000; Wu and Lin, 2005; Huang et al., 2012) to investigate the relationship between Zhong-Yong tendency and behavior. The results of these investigations converge to suggest that high middle-way thinkers tend to adopt a more global and flexible cognitive processing strategy when interacting with the external world. For example, in Huang et al. (in press) recent study, the researchers primed the participants with a neutral word or an emotional word prior to showing them a global-local stimulus on each trial. They found that the global precedence effect was larger for the high middle-way thinkers than the low middle-way thinkers only when emotion was primed. These results suggest that the global processing strategy, i.e., stepping back to see the whole picture, characterizes a high middle-way thinker's cognitive

processing style. These results also imply that Zhong-Yong, served as an emotional regulator, affected an individual's cognitive processing strategy; this emotion regulation mechanism has not been reported in the previous models of emotion. In another study, Wang et al. (2013) examined how Zhong-Yong tendency is correlated with behavioral aspects of viewing banner ads. Participants were presented with banner ads of different levels of information complexity. The eye tracking data showed that high middle-way thinkers, compared to low middle-way thinkers, viewed banner ads of lower complexity with a larger and more distributed scan path, suggesting that they adopted a more global strategy to integrate information from all regions of the ads. In addition, high middle-way thinkers started to fixate on the banner ads of lower complexity at earlier time points. Wang et al. (2013) interpreted these findings as evidence that high middle-way thinkers were more efficient and flexible in switching from global processing (e.g., processing banner ads' gist) to local processing (e.g., processing banner ads' details).

Although the relationship between Zhong-Yong thinking style and cognitive processing style has been widely investigated, less is known about how an individual's perceptual processing capacity is related to his/her Zhong-Yong tendency. Perceptual processing capacity, also known as workload capacity, is defined as the change in processing efficiency of an information processing system that occurs as the workload (the number of to-be-processed signals) increases (Townsend and Nozawa, 1995; Wenger and Gibson, 2004; Eidels et al., 2011; Townsend and Eidels, 2011; Houpt and Townsend, 2012). Perceptual processing capacity is measured with a redundant-target detection task (Miller, 1978, 1982; Townsend and Nozawa, 1995), where participants monitor two sources of information and make a decision based on either one or both sources of information. If the processing speed of an individual channel is not affected by an increase in workload, the information processing system is defined as being unlimited in capacity; if the processing speed speeds up, the processing system is considered to have supercapacity; and lastly, if the processing speed slows down, the processing system is considered to have limited capacity. An individual's perceptual processing capacity is assumed to be independent of the way he/she processes information (Townsend and Nozawa, 1995); however, some multiple-signal processing strategies may be constrained by a system's processing capacity. For example, a coactive system usually has supercapacity, whereas the processing capacity of a standard serial system is limited (Townsend, 1972, 1974; Colonius and Townsend, 1997; Townsend and Nozawa, 1997; Wenger and Townsend, 2001; Wenger and Gibson, 2004; Eidels et al., 2011; Townsend and Eidels, 2011). In addition, a parallel system with supercapacity or limited capacity may imply that there are facilitatory or inhibitory between-channel interactions during the stage of information accumulation (Colonius and Townsend, 1997; Wenger and Gibson, 2004; Eidels et al., 2011). Thus, uncovering individual differences in perceptual processing capacity between high and low middle-way thinkers can help researchers understand the causes of differences in their cognitive processing styles.

The present study aimed to investigate the relationship between middle-way thinking style and perceptual processing capacity. In two experiments, participants completed the Zhong-Yong Thinking Style Scale (Wu and Lin, 2005) and performed a redundant-target detection task. We estimated the participants' perceptual processing capacity using a non-parametric approach (systems factorial technology, or SFT, see Townsend and Nozawa, 1995 for a review) in both experiments and a parametric approach (linear ballistic accumulator model, or LBA model, Brown and Heathcote, 2008; Eidels et al., 2010) in Experiment 2. These two approaches provide converging measures of workload capacity and have complementary advantages in the assessment (Eidels et al., 2010). We hypothesized that high middle-way thinkers tend to adopt a more global processing strategy to process information compared to low middle-way thinkers; thus, they process information in a more efficient way, especially when the workload increases, leading to supercapacity processing. On the other hand, low middle-way thinkers are more limited in perceptual processing capacity such that they are more prone to interference by information complexity.

EXPERIMENT 1

In Experiment 1, a Go/No-go version of the redundant-target detection task was conducted to measure individuals' perceptual capacity for processing an object's color and shape. We used a non-parametric approach (SFT, see Townsend and Nozawa, 1995 for a review) to estimate perceptual processing capacity. The experimental design and data analysis followed the suggestions of SFT, which will be extensively described in the *Method* Section. The participants were split into two groups according to their Zhong-Yong scores, and the capacity coefficient of each group was plotted as a function of reaction time. We expected to observe qualitatively different capacity coefficient functions between high and low middle-way thinkers.

METHODS

Participants

Fifty-seven undergraduate students (29 males and 28 females) at National Cheng Kung University participated in this experiment. All participants had normal or corrected-to-normal vision, and their mean age was 20.63 years with a standard deviation of 2.72. Prior to the experiment, each participant signed a written informed consent, which has been proved by the review board of the National Cheng Kung University, Department of Psychology.

Apparatus

A personal computer with a 2.40 G-Hz Intel Pentium IV processor controlled the display and recorded the manual responses. The display resolution was 1024×768 pixels. Stimuli were presented on a 19-inch CRT monitor with a refresh rate of 85 Hz. The experiment was programmed with E-prime 1.1 (Schneider et al., 2002). The viewing distance was 60 cm. A chin-rest was used to prevent head movements.

Questionnaire

The participants' Zhong-Yong tendency was measured with a Zhong-Yong Thinking Style Scale, which was developed by Wu and Lin (2005). The Zhong-Yong Thinking Style Scale is composed of 13 items which are divided into three subscales that

measure the three different aspects of Zhong-Yong, including diversification (i.e., considering things carefully from different aspects), integrity (i.e., integrating one's and others' perspectives), and harmony (i.e., acting in a manner for maintaining interpersonal harmony). Each item is scored on a 7-point Likert-type scale from "Strongly Disagree" (1) to "Strongly Agree" (7). An individual's Zhong-Yong score is defined as the mean score of the average scores of the three subscales. The Zhong-Yong score ranges from 1 to 7. Wu and Lin (2005) tested two samples in Studies 1 ($n = 96$) and 2 ($n = 216$) to measure the reliability and validity of the Zhong-Yong thinking style scale. They found that the coefficient of the internal consistency was 0.87 for both samples and the test-retest reliability was 0.81 ($n = 46$). The results of factor analysis showed that this scale is a single-factor scale and the factor loading for each item was greater than 0.40, suggesting that all the items are good measures of the construct of Zhong-Yong. In addition, Zhong-Yong score is positively correlated to self-consciousness, self-reflection, and inclusion of other in the self, showing high construct validity of the scale (Wu, 2006).

Design, stimuli, and procedure

In the redundant-target detection task, each test display consisted of a colored letter (X or O) presented at the center of the screen. Its color was either green ($x = 0.30$, $y = 0.60$, $luminance = 1.90$ cd/m^2) or cyan ($x = 0.33$, $y = 0.33$, $luminance = 2.71$ cd/m^2). The size of the letter was 1° × 1°. The target color was defined as green and the target shape was defined as X; the distractor color was defined as cyan and the distractor shape was defined as O. The test display consisted of both target features (i.e., a green X, redundant-target condition), either target feature (i.e., a green O or a cyan X, single-target condition), or neither target feature (i.e., a cyan O, no-target condition) (see **Figure 1A** for all the possible test trials). Each condition was equally probable and was randomly intermixed within each block such that the participants would not anticipate the presence of the redundant-target trials (Mordkoff and Yantis, 1991, 1993). There were 40 practice trials and twelve blocks of 80 formal test trials in each experiment.

The experiment was conducted in a dimly lit room. A trial began with a 500 ms fixation cross, accompanied with a 750 Hz pure tone (see **Figure 1B** for an illustration of the experimental procedure). After a blank interval ranging from 50 to 850 ms, a test display was presented. Participants were instructed to press the "/" key if they detected either target feature (color green or shape X) and they were instructed to hold their responses if they detected neither target feature. The test display disappeared after a response was made (Go trial); otherwise, it remained on the screen until 2000 ms had passed (No-go trial). The inter-trial interval (ITI) was 500 ms. Both speed and accuracy were emphasized.

Data analysis

According to SFT, the capacity coefficient $C(t)$ was computed to infer an individual's perceptual processing capacity. The capacity coefficient $C(t)$ can be expressed as follows (Townsend and Nozawa, 1995; Townsend and Eidels, 2011; Houpt and Townsend,

2012; Houpt et al., 2014):

$$C(t) = \frac{\log S_{1,2}(t)}{\log [S_1(t) \cdot S_2(t)]}, \quad (1)$$

for $t > 0$, where S_1, S_2, and $S_{1,2}$ represent the survivor functions of the two single-target conditions and the redundant-target condition, respectively. The ranges of values of $C(t)$ and their implications are as follows: if $C(t) > 1$, the system is supercapacity; if $C(t) = 1$, the system is unlimited-capacity; if $C(t) < 1$, it is limited-capacity; and if $C(t) \leqq 0.5$, the system is extremely limited in capacity.
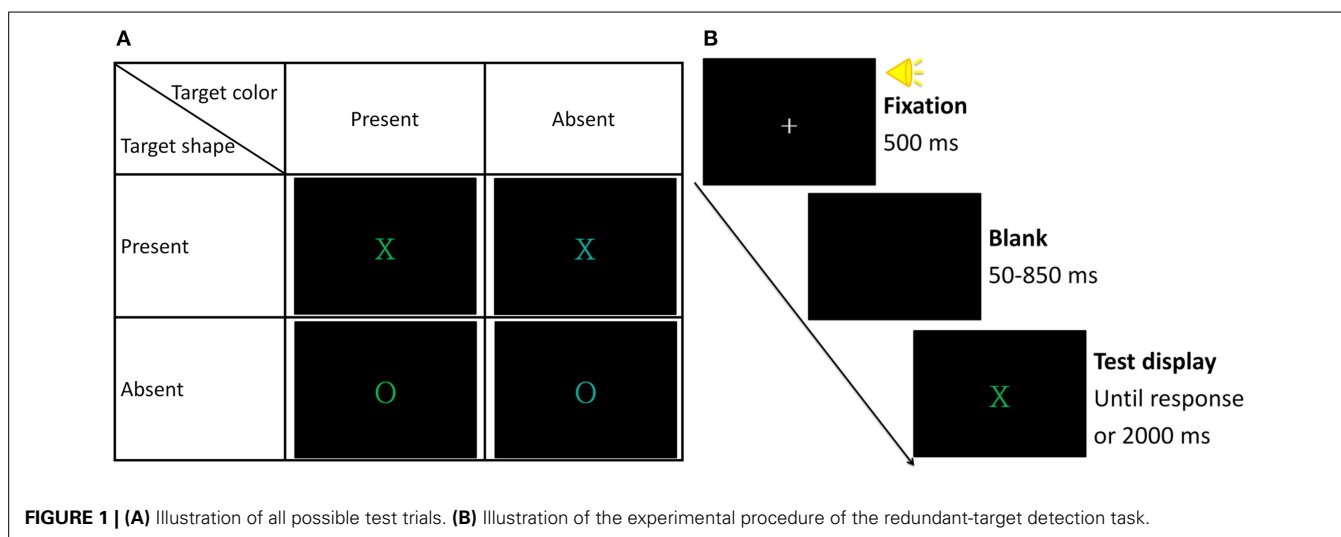
RESULTS AND DISCUSSION

We first analyzed the participants' Zhong-Yong tendency. The mean Zhong-Yong score for all of the participants was 5.80 with a standard deviation of 0.63. The participants were split into two groups according to their Zhong-Yong scores: the high middle-way thinkers ($N = 10$, $M = 6.69$, $SD = 0.17$) were the ones who scored at the top one-fifth on the Zhong-Yong scores and the low middle-way thinkers ($N = 12$, $M = 4.93$, $SD = 0.32$) were the ones who scored at the bottom one-fifth on the Zhong-Yong scores[1]. There was a significant difference in the Zhong-Yong scores between groups [$t_{(17.25)} = 16.40$, $p < 0.0001$][2].

Next, we examined the mean performance on the redundant-target detection task for each group of participants (see **Table 1**). Correct reaction times ranging from 150 to 1000 ms were extracted for further analysis. This range was chosen because simple reaction time is generally not faster than 150 ms and is not longer than 1000 ms. Under this criterion, a total of 1.4% data points were excluded from analysis. The mean accuracy was very high across conditions for both groups of participants except for the no-target conditions, suggesting a potential response bias in making a decision. We limited the remainder of our analyses to the reaction times. The mean reaction time in the redundant-target condition was faster than that in the single-target condition for the high middle-way thinkers [$t_{(9)} = 12.30$, $p < 0.0001$] and for the low middle-way thinkers [$t_{(11)} = 3.47$, $p < 0.01$], suggesting that the redundant-target effect was consistently found

---

[1]The reason why we adopted the extreme-group approach is to emphasize the differences between high and low Zhong-Yong groups since the SFT results were somewhat noisy. However, even when we used median-split to analyze the data, we still obtained a similar pattern of results.

[2]We thank the anonymous reviewer for raising this question: both high and low Zhong-Yong groups show Zhong-Yong tendency even though there are significant differences in their Zhong-Yong scores. Unfortunately, there is no norm for the Zhong-Yong Thinking Style Scale. Therefore, we used the data reported in Wu and Lin (2005) to estimate the mean and standard deviation of their participants' Zhong-Yong scores ($n = 216$ in Study 2). The mean is 5.44 and the standard deviation is 0.32. Compared to our current findings [high middle-way thinkers: $M = 6.69$, $SD = 0.17$; low middle-way thinkers: $M = 4.93$, $SD = 0.32$], our high/low Zhong-Yong group had the score significantly higher/lower than the average score reported in the original study. Therefore, we can claim that the high and low Zhong-Yong groups in the current study had different Zhong-Yong tendency than the average of the Taiwanese population, although it is still possible that all the Taiwanese participants have a stronger Zhong-Yong tendency than other people from different culture backgrounds. Future studies are required to explore the cross-cultural variation.

FIGURE 1 | (A) Illustration of all possible test trials. **(B)** Illustration of the experimental procedure of the redundant-target detection task.

Table 1 | Mean performance of the redundant-target detection task for each group of participants in Experiment 1.

Group	Accuracy			Reaction time (ms)		
	RT	*ST*	*NT*	*RT*	*ST*	*RG*
High	0.99	0.99	0.86	397.79	441.64	43.85
Low	0.99	0.96	0.81	399.10	438.05	38.95

"High" and "Low" denote the high and low middle-way thinkers, respectively. "RT," "ST," and "NT" represent the redundant-target, single-target, and no-target conditions, respectively. Redundancy gain (RG) is defined as the difference in mean reaction times between the redundant-target and single-target conditions. Note that mean reaction time of the no-target condition was not shown because in Experiment 1 any response in this condition is incorrect for the Go/No-go version of the redundant-target detection task.

in both groups of participants. In addition, the redundancy gain was not significantly different between the groups [$t_{(13.16)} = 0.42$, $p = 0.68$].

We then computed $C(t)$ for each participant and plotted the estimated $C(t)$ by group. **Figure 2A** shows $C(t)$ as a function of reaction time for each group. From visual inspection, the results showed that for most high middle-way thinkers, $C(t)$ was larger than 1 for the faster reaction times, suggesting supercapacity processing. By contrast, for most low middle-way thinkers, $C(t)$ was less than 1 for all times t and a few values of $C(t)$ were hovering between ~0 and 0.5, suggesting limited-capacity to extremely limited-capacity processing. To verify these observations, we adopted a non-parametric bootstrapping method to simulate 1000 samples for each condition and to construct the 95% confidence interval for $C(t)$ individually (Van Zandt, 2000). If the 95% confidence interval for $C(t)$ exceeds 1 at some times t, we conclude that the participant adopts supercapacity processing to process multiple signals. Otherwise, we conclude that the participant adopts unlimited-capacity or limited-capacity processing. **Table 2** presents the classification results of the inferences based on the simulated data for each group. Results showed that 4 out of 10 high middle-way thinkers adopted

supercapacity processing; in contrast, only 1 (out of 12) low middle-way thinkers showed this pattern of results. When applying Fisher's exact test to test whether processing capacity and Zhong-Yong tendency are independent, the results, however, did not reach the significance level ($p = 0.14$). It is perhaps due to the small sample size that we did not obtain a significant result. Though, there is a trend showing that more high middle-way thinkers had a supercapacity system than low middle-way thinkers.

The results of Experiment 1 were consistent with our expectations. The high middle-way thinkers had systems with larger perceptual processing capacity than the low middle-way thinkers. The high middle-way thinkers generally exhibited supercapacity processing, suggesting that they adopted coactive processing to process multiple sources of information or that there were facilitatory between-channel cross-talks during the stage of information accumulation (Eidels et al., 2011). In contrast, the low middle-way thinkers exhibited limited-capacity or extremely limited-capacity processing when processing multiple signals, suggesting that they processed information in sequence or that there were inhibitory interactions between channels (Eidels et al., 2011). Therefore, the current findings provided empirical support for the notion that the high middle-way thinkers process redundant information more efficiently and in an integrative fashion, and the low middle-way thinkers were much more limited in capacity such that they serially processed multiple sources of information and were prone to interference as the workload increased.

EXPERIMENT 2

In Experiment 1, we adopted a non-parametric approach (SFT) to estimate perceptual processing capacity, and the results of the visual inspection showed that the high middle-way thinkers had larger perceptual processing capacity than the low middle-way thinkers. However, there are a few limitations in Experiment 1. First, we only used correct reaction times for capacity estimation while ignoring the incorrect reaction times. Second, the lower accuracy in the no-target condition may reflect a
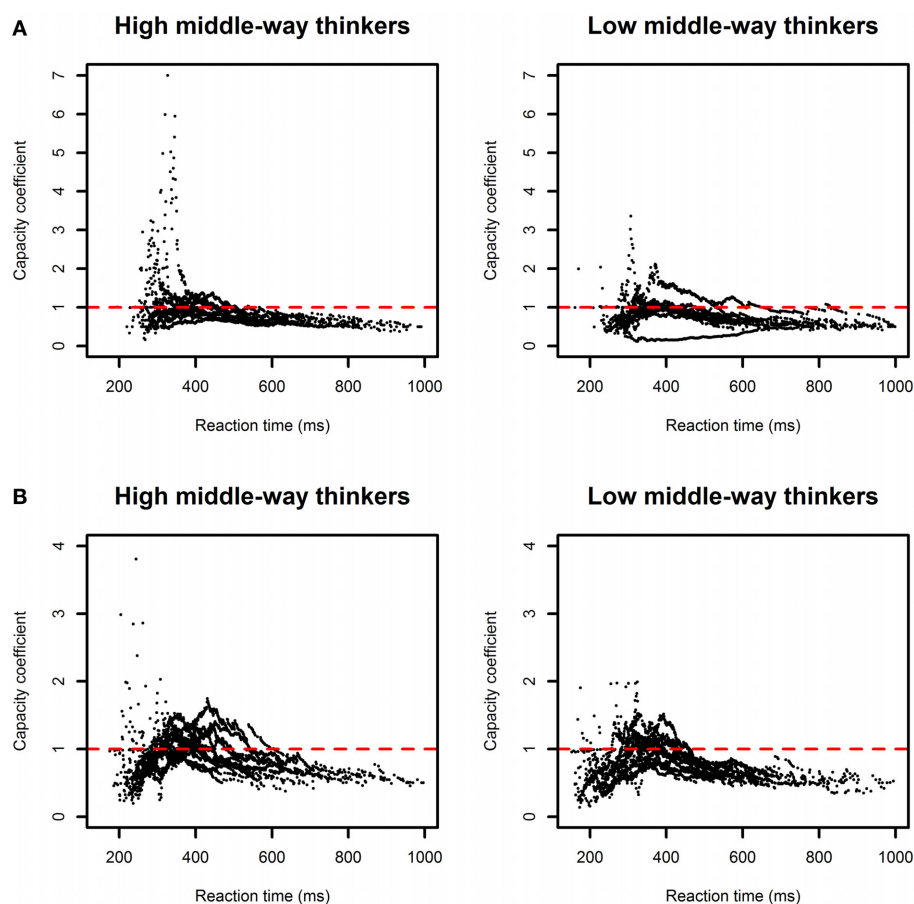
FIGURE 2 | (A) Plots of the capacity coefficient *C(t)* for the high and low middle-way thinkers in Experiment 1. **(B)** Plots of the capacity coefficient *C(t)* for the high and low middle-way thinkers in Experiment 2.

Table 2 | The classification results (frequency) of the inferences based on the simulated data for each group in Experiments 1 and 2.

Group	Experiment 1		Experiment 2	
	Supercapacity	Non-supercapacity	Supercapacity	Non-supercapacity
High	4	6	6	7
Low	1	11	2	13

"High" and "Low" denote the high and low middle-way thinkers, respectively.

potential response bias in target detection. Third, the extreme-group approach adopted in Experiment 1 only provides a discrete distinction between the high and low middle-way thinkers. It is unclear whether there is a linear relationship between Zhong-Yong tendency and perceptual processing capacity. Hence, a parametric approach, LBA model (Brown and Heathcote, 2008; Eidels et al., 2010), was adopted in Experiment 2 to estimate perceptual processing capacity in order to obtain a continuous measurement of the relationship between the Zhong-Yong tendency and perceptual processing capacity. This approach also provides researchers with a parametric testing tool to identify the perceptual processing capacity of a system. To implement the LBA model in this experiment, a yes/no version of the

redundant-target detection task was used instead of a Go/No-go version of the redundant-target detection task because the analysis required reaction time data in both the target-present condition and the target-absent condition. We expected that the relationship between Zhong-Yong tendency and perceptual processing capacity observed in Experiment 1 would generalize to the choice reaction time experiment.

METHODS

Participants

Seventy-three undergraduate students (27 males and 46 females) at National Cheng Kung University who had not participated in Experiment 1 participated in this experiment. All of the

participants had normal or corrected-to-normal vision, and their mean age was 19.27 years with a standard deviation of 1.34. Prior to the experiment, each participant signed a written informed consent, which has been proved by the review board of the National Cheng Kung University, Department of Psychology.

Design, stimuli, and procedure

The stimuli, design, and procedure used in the redundant-target detection task were the same as those in Experiment 1, except that the participants were instructed to make a yes/no response for target detection. When the participants detected either target feature, they had to press "/" key; otherwise, they had to press "z" key.

Data analysis

We used both a non-parametric approach (SFT, Townsend and Nozawa, 1995) as in Experiment 1 and a parametric approach (LBA model, Brown and Heathcote, 2008; Eidels et al., 2010) to estimate the participants' perceptual processing capacity. First, the estimated $C(t)$ for the high and low middle-way thinkers were plotted separately and a non-parametric bootstrapping method was used to construct each participant's 95% confidence interval for $C(t)$ to infer the perceptual processing capacity. Second, we computed the Pearson's product-moment correlation coefficient (r) between the LBA-based capacity and Zhong-Yong score to verify the relationship between the two measurements.

The following is a brief description of the LBA model (Brown and Heathcote, 2008; Eidels et al., 2010). The LBA model takes both correct and incorrect reaction times in the target-present and the target-absent conditions into consideration in the analysis. In a redundant-target detection task, four parallel accumulators are assumed to accumulate evidence independently and simultaneously about the presence of the target color (C), the absence of the target color ($\sim$C), the presence of the target shape (S), and the absence of the target shape ($\sim$S), respectively. Each accumulator starts to accumulate evidence from a random initial starting point, which is distributed as a uniform distribution in [0, A]. Evidence is accumulated linearly at a drift rate that is drawn from a normal distribution with a mean v and a standard deviation s. Accumulation is terminated and a decision is made when the amount of evidence reaches a threshold b. The reaction time is the decision time (i.e., the time for the accumulation reaching the threshold) plus the base time t_0 (i.e., the time for the perceptual processing and motor execution).

In a redundant-target detection task, either of the yes/no responses can be made on each trial: "YES" for the presence of either target feature, and "NO" for the absence of both target features. Specifically, a "YES" response occurs when accumulator C reaches the threshold but accumulator S has not reached the threshold or when accumulator S reaches the threshold but accumulator C has not reached the threshold. The overall likelihood of a "YES" response occurring at time t is expressed as

$$L(YES, t) = [1 - F_{\sim C}(t) \cdot F_{\sim S}(t)] \cdot [f_C(t) \cdot S_S(t) + f_S(t) \cdot S_C(t)], \tag{2}$$

where F, f, and S denote the cumulative distribution function, density function and survivor function for each accumulator, respectively. Similarly, a "NO" response occurs when accumulators $\sim$C and $\sim$S reach the threshold before accumulators C and S have not reached the threshold. The overall likelihood of a "NO" response occurring at time t is expressed as

$$L(NO, t) = S_C(t) \cdot S_S(t) \cdot [f_{\sim C}(t) \cdot F_{\sim S}(t) + f_{\sim S}(t) \cdot F_{\sim C}(t)]. \tag{3}$$

Likelihood functions, L(YES,t) and L(NO,t), were used to obtain the maximum likelihood estimates of the parameters for each accumulator given the correct and incorrect reaction times. The initial starting point A was fixed across conditions, and the standard deviation s was set as 0.25 in reference to Donkin et al. (2009). We assumed two decision threshold parameters for the target-present condition (b_T) and target-absent condition (b_{NT}) because the participants may set different criteria for making "YES" and "NO" responses due to the unequal presentation probability across the two conditions. However, b_T was assumed to not vary across the redundant-target condition and the two single-target conditions because changes in the boundary parameter were unlikely to occur when all target-present conditions were randomly intermixed within a block (Ratcliff, 1978). Base times for the redundant-target accumulator (t_{0RT}), the single-target accumulator (t_{0ST}), and the no-target accumulator (t_{0NT}) were estimated separately because sensory encoding time may vary as a function of the number of signals to be processed.

Drift rate estimation is the most important part of the estimation of the LBA-based capacity measure. When the target was present, we assumed three drift rate parameters for the redundant-target accumulator (v_{RT}), the single-target accumulator (v_{ST}), and the no-target accumulator (v_{NT}). When the target was absent, we assumed two drift rate parameters for the no-target accumulator ($v_{\sim NT}$) and the target accumulator ($v_{\sim T}$). Note that there are 16 possible drift rate parameters (see **Table 3**), but we only estimated five of them because we assumed that the drift rates for accumulator C and accumulator S were the same and the drift rates for accumulator $\sim$C and accumulator $\sim$S were also the same. These two assumptions need not to be true; however, similar pattern of results was observed when we allowed the variation between all the 16 drift rate parameters. Therefore, a total of 11 free parameters (A, b_T, b_{NT}, t_{0RT}, t_{0ST}, t_{0NT}, v_{RT}, v_{ST}, v_{NT}, $v_{\sim T}$, $v_{\sim NT}$) were estimated for each participant.

The LBA-based capacity is defined as the relative magnitudes between drift rates in the redundant-target condition and the single-target condition, which can be expressed as

$$v_{diff} = v_{RT} - v_{ST}. \tag{4}$$

If $v_{diff} > 0$, the system is supercapacity processing; if $v_{diff} = 0$, the system is unlimited-capacity processing; if $v_{diff} < 0$, the system is limited-capacity processing.

Table 3 | The simplified set of five drift rate parameters (right-hand side) used in the LBA model and their corresponding drift rates of all accumulators (left-hand side) in the redundant-target task.

		Target color	
		Present (C)	Absent (∼C)
Target shape	Present (S)	$v_{C\|CS} = v_{RT}$	$v_{C\|\sim CS} = v_{\sim T}$
		$v_{S\|CS} = v_{RT}$	$v_{S\|\sim CS} = v_{ST}$
		$v_{\sim C\|CS} = v_{NT}$	$v_{\sim C\|\sim CS} = v_{\sim NT}$
		$v_{\sim S\|CS} = v_{NT}$	$v_{\sim S\|\sim CS} = v_{NT}$
	Absent (∼S)	$v_{C\|C\sim S} = v_{ST}$	$v_{C\|\sim C\sim S} = v_{\sim T}$
		$v_{S\|C\sim S} = v_{\sim T}$	$v_{S\|\sim C\sim S} = v_{\sim T}$
		$v_{\sim C\|C\sim S} = v_{NT}$	$v_{\sim C\|\sim C\sim S} = v_{\sim NT}$
		$v_{\sim S\|C\sim S} = v_{\sim NT}$	$v_{\sim S\|\sim C\sim S} = v_{\sim NT}$

Subscripts for the simplified set of five drift rates are described in the Data Analysis section of Experiment 2. Subscripts for the full set of 16 drift rate parameters denote the drift rate for a specific accumulator given any of the four test trials. For instance, $v_{C|CS}$ represents the drift rate for accumulator C when both the target color and shape are present and is mapped to the drift rate for the redundant-target accumulator v_{RT}.

RESULTS AND DISCUSSION

Data from two participants were excluded because they were unable to follow the experimental instructions. The mean Zhong-Yong score for all of the participants was 5.72 with a standard deviation of 0.70. We used an extreme-group approach, as we did in Experiment 1. The participants who scored at the top one-fifth on the Zhong-Yong score were regarded as high middle-way thinkers ($N = 13$, $M = 6.56$, $SD = 0.16$), and the participants who scored at the bottom one-fifth on the Zhong-Yong score were considered as low middle-way thinkers ($N = 15$, $M = 4.67$, $SD = 0.58$). There was a significant difference in the Zhong-Yong scores between groups [$t_{(16.31)} = 12.13$, $p < 0.0001$].

Next, we examined the mean performance of the redundant-target detection task for each group of participants (see **Table 4**). Using the same criterion as Experiment 1, a total of 6.1% reaction time data of the redundant-target detection task was excluded from further analysis. Similar to Experiment 1, accuracy was lower in the no-target conation than the other conditions, suggesting a potential response bias in target detection. Although the mean performance in this experiment was worse than that in Experiment 1 [accuracy: $t_{(105.40)} = 2.06$, $p < 0.05$; reaction time: $t_{(114.70)} = 10.89$, $p < 0.0001$], we still observed the redundant-target effect for both the high middle-way thinkers [$t_{(12)} = 10.76$, $p < 0.0001$] and the low middle-way thinkers [$t_{(14)} = 10.04$, $p < 0.0001$]. In addition, the redundancy gain was not significantly different between the groups [$t_{(25.33)} = 1.14$, $p = 0.27$].

As in Experiment 1, we computed $C(t)$ and constructed the 95% confidence interval for $C(t)$ for each participant to infer the perceptual processing capacity. **Figure 2B** plots the results of $C(t)$ for each group of participants. The results of the non-parametric measures of capacity replicated what we found in Experiment 1; that is, $C(t)$ was generally larger for the high middle-way thinkers

Table 4 | Mean performance of the redundant-target detection task for each group of participants in Experiment 2.

Group	Accuracy			Reaction time (ms)			
	RT	ST	NT	RT	ST	NT	RG
High	0.99	0.96	0.88	348.81	396.14	457.98	47.33
Low	0.99	0.96	0.83	362.99	403.53	477.95	40.54

"High" and "Low" denote the high and low middle-way thinkers. "RT," "ST," and "NT" represent the redundant-target, single-target, and no-target conditions, respectively. Redundancy gain (RG) is defined as the difference in the mean reaction times between the redundant-target and single-target conditions.

than for the low middle-way thinkers. Based on the simulated data (see **Table 2**), we inferred that 6 out of 13 high middle-way thinkers had a system of supercapacity processing, while only 2 out of 15 low middle-way thinkers showed this pattern of results. Note that a few low middle-way thinkers had $C(t)$ that was greater than 1 at early time points (see **Figure 2B**); however, compared to high middle-way thinkers, the values of $C(t)$ were relatively small, suggesting that low middle-way thinkers were less efficient in processing multiple sources of information. We then conducted a Fisher's exact test to test whether processing capacity and Zhong-Yong tendency are independent. The result still did not reach the significance level ($p = 0.10$) although there is a trend showing that more high middle-way thinkers were classified in the supercapacity category than low middle-way thinkers and less high-middle-way thinkers were classified in the non-supercapacity category than low middle-way thinkers. Nevertheless, when we combined the data of Experiments 1 and 2 to increase the sample size, the result of the Fisher's exact test was significant ($p < 0.05$), verifying that Zhong-Yong tendency and processing capacity are dependent on each other.
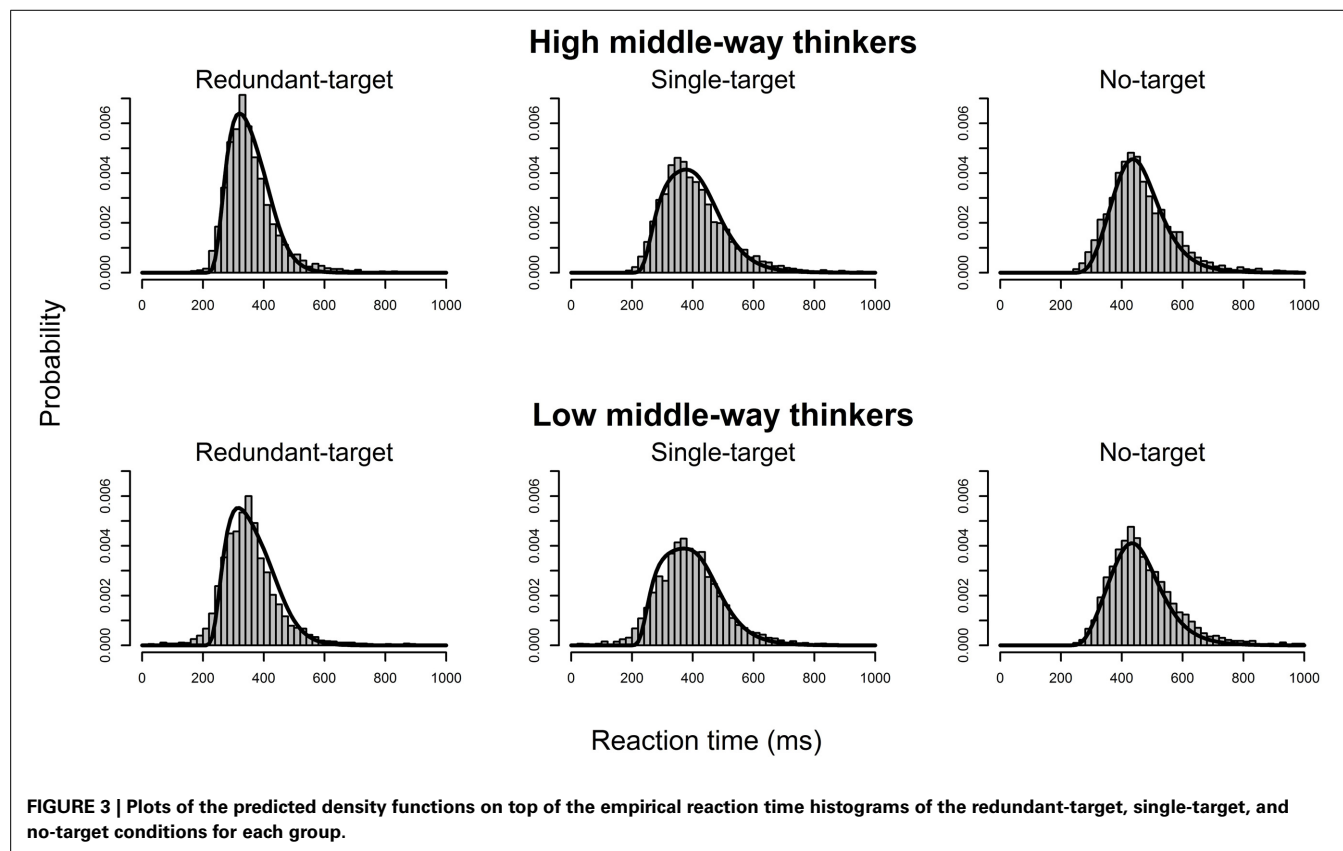
Next, we adopted the LBA model to analyze the reaction time data to estimate a set of parameters that maximized the likelihood function described in the *Method* Section for each participant. **Table 5** presents the average of 11 estimated parameters for each group. None of the parameters differed between high and low middle-way thinkers ($ps > 0.12$). We then used the average of the estimated parameters to generate model predictions from the LBA model and plotted the empirical histograms for correct responses along with corresponding model predictions (see **Figure 3**). The results showed that the LBA model successfully captured the underlying distributions of the reaction time data, suggesting that the LBA model fit the participants' reaction time data well.

We then computed the LBA-based capacity (v_{diff}) for each group (see **Table 5**). The results showed that the drift difference for the high middle-way thinkers ($M = 0.07$, $SD = 0.17$) was larger than that of the low middle-way thinkers ($M = -0.04$, $SD = 0.16$) [$t_{(24.40)} = 1.87$, $p < 0.05$]. Lastly, we computed the Pearson's product-moment correlation (r) between the LBA-based capacity and the Zhong-Yong score, and we found a significant positive correlation between the two measurements [$r = 0.35$, $p < 0.01$, 95% CI = (0.13, 0.54)] (**Figure 4**), suggesting that the perceptual processing capacity monotonically increases as Zhong-Yong tendency increases.

Table 5 | The average values of 11 estimated parameters and the LBA-based capacity (v_{diff}) for the high and low middle-way thinkers.

Group	Estimated parameters										
	A	b_T	b_{NT}	t_{0RT}	t_{0ST}	t_{0NT}	v_{RT}	v_{ST}	v_{NT}	$v_{\sim T}$	$v_{\sim NT}$	v_{diff}
High	288.06	507.53	581.60	112.59	101.32	79.45	1.29	1.22	0.66	0.37	1.34	0.07
Low	313.75	493.60	576.45	120.25	114.32	78.51	1.17	1.21	0.64	0.35	1.31	-0.04

"High" and "Low" denote the high and low middle-way thinkers.

FIGURE 3 | Plots of the predicted density functions on top of the empirical reaction time histograms of the redundant-target, single-target, and no-target conditions for each group.
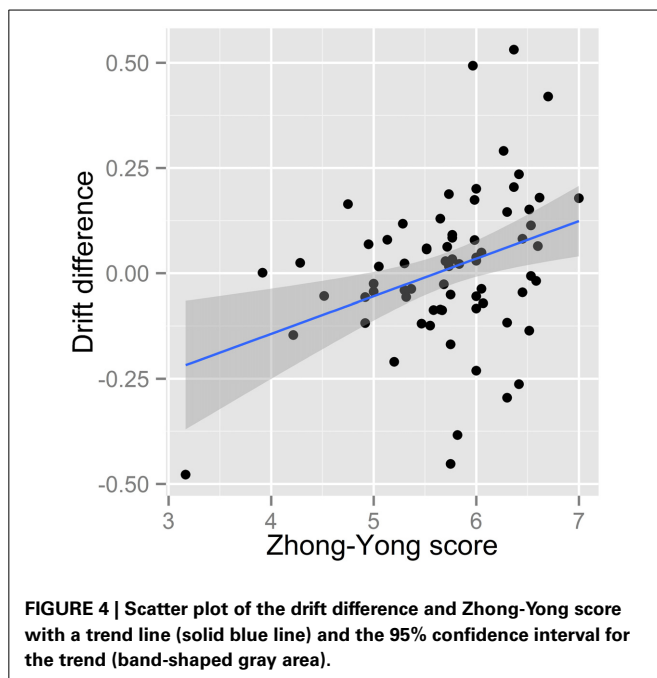
GENERAL DISCUSSION

In the present study, two experiments were conducted to investigate how an individual's Zhong-Yong tendency is related to his/her perceptual processing capacity. The Zhong-Yong Thinking Style Scale (Wu and Lin, 2005) was used to assess the participant's Zhong-Yong tendency. The redundant-target detection task was adopted to infer the participants' perceptual processing capacity in a non-parametric manner (SFT in Experiments 1 and 2) as well as in a parametric manner (LBA model in Experiment 2). The results from the non-parametric and parametric analyses converged to suggest that participants with a strong Zhong-Yong tendency had larger perceptual capacity in processing redundant information for decision making. High middle-way thinkers had an unlimited-capacity to supercapacity processing system, suggesting that the processing time of an individual channel was unaffected or even sped up when workload increased. In contrast, low middle-way thinkers had a limited-capacity processing system, suggesting that the

individual-channel processing time slowed down as a result of the increasing workload.

ZHONG-YONG TENDENCY AND PERCEPTUAL PROCESSING CAPACITY

The current results were consistent with our expectation that high middle-way thinkers have larger perceptual processing capacity and process multiple signals more efficiently as workload increases. Two possible accounts may explain the reasons why the high middle-way thinkers had larger perceptual processing capacity than the low middle-way thinkers. First, it is worthwhile to note that although the processing architecture (i.e., the way that redundant information is processed) and the processing capacity (i.e., the variation in the efficiency of a system as a function of workload) are independent measures of information processing (Townsend and Nozawa, 1995), processing capacity may constrain the processing order of multiple signals. For example, a coactive system is commonly assumed to have supercapacity, while a standard serial model is assumed to be

FIGURE 4 | Scatter plot of the drift difference and Zhong-Yong score with a trend line (solid blue line) and the 95% confidence interval for the trend (band-shaped gray area).

limited in capacity, although the standard serial model and the unlimited-capacity parallel model can mimic each other theoretically(Townsend, 1972, 1974; Colonius and Townsend, 1997; Townsend and Nozawa, 1997; Wenger and Townsend, 2001; Wenger and Gibson, 2004; Eidels et al., 2011; Townsend and Eidels, 2011). Our results showed that the high middle-way thinkers had supercapacity processing, implying that they tended to process redundant information in a coactive fashion. That is, multiple signals are processed in parallel and simultaneously, and separate activations from multiple channels are accumulated and summed into a single accumulator. A decision is made when the accumulated evidence reaches the decision criterion. By contrast, the low middle-way thinkers exhibited limited-capacity processing, implying that they had less capacity for multiple-signal processing such that they may process redundant information in a serial fashion. Namely, one of the target features is processed first, and if the information is sufficient for decision making, the other processing is terminated as predicted by a serial self-terminating model.

However, individual differences in perceptual processing capacity do not necessarily mean that high and low middle-way thinkers adopt different processing strategies. Assuming that multiple signals are processed in a parallel fashion for all participants, differences in processing capacity may suggest differences in the way multiple processes interact with each other during information accumulation. According to Eidels et al. (2011), different types of between-channel interactions explain the variation in the processing efficiency of an individual channel as workload increases. They simulated a parallel model with different levels of between-channel interactions and found that a parallel model with supercapacity processing suggests that there are facilitatory (positive) interactions between channels during information accumulation, while a parallel model with limited-capacity

processing suggests that there are inhibitory (negative) between-channel cross-talks. Accordingly, high middle-way thinkers can integrate multiple signals more efficiently with positive between-channel interactions; by contrast, low middle-way thinkers are more prone to interference by information complexity due to negative between-channel interactions that result in mutual inhibitions between each process.

Future studies are required to further examine the possibility that high and low middle-way thinkers may adopt different multiple-signal processing strategies for decision making. An ongoing study has been designed following Townsend and Nozawa (1995) suggestions to use a standard double factorial paradigm in which nine test stimuli with simultaneous manipulation of the target feature and the target intensity are used to directly test the processing architecture adopted by high and low middle-way thinkers. In addition, this study may also enable us to uncover differences in between-channel interactions during information accumulation.

ZHONG-YONG TENDENCY AND COGNITIVE PROCESSING STYLE

Many researchers are interested in understanding how culture shapes behavior. In regard to middle-way thinking, or Zhong-Yong, Chinese culture has long regarded middle-way thinking as one of the most important meta-cognitive factors that regulate one's emotions and attitudes (Ji et al., 2010; Yang, 2010). People who have a strong Zhong-Yong tendency can be characterized by their global and flexible cognitive processing styles (Wang et al., 2013; Huang et al., in press). In addition, a recent study showed that Zhong-Yong can moderate the relationship between perceived creativity and innovation behavior in Chinese companies (Yao et al., 2010).

The present study, which tested individual differences in perceptual processing capacity, can offer further insights into aspects of how Chinese culture influences individuals' behavior. First, individual differences can be observed in a relatively fundamental perceptual task (i.e., the color-shape detection task used in the present study). These findings are in line with previous research on cross-cultural comparisons between East Asian and West Caucasian (Norenzayan and Nisbett, 2000; Masuda and Nisbett, 2001, 2006; Kitayama et al., 2003; Nisbett and Miyamoto, 2005; Miyamoto et al., 2006). One distinction that has been revealed in cross-cultural research is the contrast between individualist cultures (Western culture) and collectivist cultures (Eastern culture) (see Triandis, 1995). Individualists emphasize individual achievements and goal; collectivists emphasize group membership and value group cohesion and success above personal achievement. Nisbett and colleagues conducted a large body of research, which suggests that members of individualist and collectivist cultures tend to have measurably different cognitive processing styles. That is, East Asians (collectivist) are field-dependent, and they process information more holistically, seeing the relation between things; by contrast, West Caucasians (individualist) are field-independent and they process information analytically, focusing on individual objects. The cultural variation in cognition and perception allows us to challenge the idea that the rules used in thought are fixed by a hard-wired mental logic and provides empirical supports for the top-down influence on perception.

Second, the current findings oppose the argument proposed by a few Zhong-Yong studies that the mechanism of Zhong-Yong thinking, the wisdom of "middle way," is akin to the mechanism of Western wisdom, and its influence can be revealed only when conflicts, dilemmas, or affections are raised (Grossmann et al., 2010, 2013). This argument was empirically supported by Huang et al. (in press), in which differences were found in the global precedence effect between high and low middle-way thinkers only when participants' emotions were primed. Nonetheless, in the present study, we found individual differences in a perceptual task without manipulating emotions. One possibility to explain the inconsistent findings is the difference between the scales used in the current study and Huang et al.'s study. In the current study, we used the scale developed by Wu and Lin (2005) which measures three aspects of Zhong-Yong; by contrast, Huang et al. used the Zhong-Yong Belief-Value Scale developed by Huang et al. (2012) which emphasizes the harmony dimension of Zhong-Yong. Therefore, we suggest that the influence of Zhong-Yong can be context independent in terms of the way Zhong-Yong tendency is assessed. The culturally induced wisdom or thinking style is a stable meta-cognitive factor that regulates one's behavior and is not specific to any context. Perceptual processing capacity may play an important role in mediating the influence of Zhong-Yong thinking on cognitive processing style. Future investigations are required to verify the mediating role of perceptual capacity in dealing with complex cognitive tasks.

ADVANTAGES AND LIMITATIONS OF THE PRESENT STUDY

The present study adopted both parametric (LBA model) and non-parametric (SFT) mathematical modeling approaches to study individual differences in perceptual processing capacity, and both levels of analyses showed similar patterns of results. Compared to previous research that tested mean reaction time by aggregating the data of each group (Wang et al., 2013; Huang et al., in press), this study considered the reaction time distribution and inferred the information processing characteristics individually. In addition, SFT and the LBA model have compensatory advantages in analyzing reaction time distributions (Eidels et al., 2010). SFT only considers correct reaction time data but allows researchers to examine the processing architecture (serial vs. parallel vs. coactive), the decisional stopping rule (self-terminating vs. exhaustive), and the processing capacity (limited-capacity vs. unlimited-capacity vs. supercapacity) (Townsend and Nozawa, 1995). By contrast, the LBA model assumes that two processes occur in a parallel fashion, but it incorporates reaction time and accuracy data into the analysis (Brown and Heathcote, 2008; Eidels et al., 2010). In addition, the LBA model provides a statistical basis for making inferences about the perceptual processing capacity of an information processing system (Eidels et al., 2010).

However, testing the processing capacity does not directly test the processing order of multiple-signal processing, given that the perceptual capacity and the processing architecture are two independent measures of information processing (Townsend and Nozawa, 1995). To further understand how middle-way thinking influences information processing strategies, a standard double factorial paradigm (Townsend and Nozawa, 1995) is required, as stated in the previous section. With a closer examination of the variation of the processing characteristics of information processing, we can further our understanding of cultural differences in cognitive processing.

CONCLUSION

The present study is the first study to elucidate the relationship between Zhong-Yong tendency and perceptual processing capacity. We found that individual differences in perceptual processing capacity are predicted well by an individual's Zhong-Yong tendency. Specifically, participants with stronger Zhong-Yong tendencies had larger perceptual processing capacities. These individual differences provide insight into the reasons why high middle-way thinkers are more flexible and efficient in processing multiple sources of information in an integrative fashion. These results emphasize that culture can shape an individual's cognitive processing style, and that the cultural shaping of cognitive style can be revealed in a fundamental perceptual task.

ACKNOWLEDGMENTS

REFERENCES

Brown, S. D., and Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cogn. Psychol.* 57, 153–178. doi: 10.1016/j.cogpsych.2007.12.002

Chiu, C.-Y. (2000). Assessment of Zhong-Yong (dialectic) thinking - preliminary findings from a cross-regional study. *Hong Kong J. Soc. Sci.* 18, 33–55.

Colonius, H., and Townsend, J. T. (1997). "Activation-state representation of models for the redundant signals effect," in *Choice, Decision, and Measurement: Essays in honor of R. Duncan Luce*, ed A. A. J. Marley (Mahwah, NJ: Lawrence Erlbaum Associates), 245–254.

Donkin, C., Brown, S. D., and Heathcote, A. (2009). The overconstraint of response time models: rethinking the scaling problem. *Psychon. Bull. Rev.* 16, 1129–1135. doi: 10.3758/PBR.16.6.1129

Eidels, A., Donkin, C., Brown, S. D., and Heathcote, A. (2010). Converging measures of workload capacity. *Psychon. Bull. Rev.* 17, 763–771. doi: 10.3758/PBR.17.6.763

Eidels, A., Houpt, J. W., Altieri, N., Pei, L., and Townsend, J. T. (2011). Nice guys finish fast and bad guys finish last: facilitatory vs. inhibitory interaction in parallel systems. *J. Math. Psychol.* 55, 176–190. doi: 10.1016/j.jmp.2010.11.003

Grossmann, I., Na, J., Varnum, M. E. W., Kitayama, S., and Nisbett, R. E. (2013). A route to well-being: intelligence versus wise reasoning. *J. Exp. Psychol. Gen.* 142, 944–953. doi: 10.1037/a0029560

Grossmann, I., Na, J., Varnum, M. E. W., Park, D. C., Kitayama, S., and Nisbett, R. E. (2010). Reasoning about social conflicts improves into old age. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7246–7250. doi: 10.1073/pnas.1001715107

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., and Townsend, J. T. (2014). Systems factorial technology with R. *Behav. Res. Methods* 46, 307–330. doi: 10.3758/s13428-013-0377-3

Houpt, J. W., and Townsend, J. T. (2012). Statistical measures for workload capacity analysis. *J. Math. Psychol.* 56, 341–355. doi: 10.1016/j.jmp.2012.05.004

Huang, C.-L., Chung, Y.-C., and Lin, Y.-C. (in press). "Forest before trees when needed: global processing and Zhong-Yong," in *Chinese Social Psychological Review*, Vol. 8, eds Z.-F. Yang and Q.-W. Wei (Beijing: Social Sciences Academic Press).

Huang, C.-L., Lin, Y.-C., and Yang, C.-F. (2012). Revision of the ZhongYong belief-value scale. *Indigenous Psychol. Res. Chin. Soc.* 38, 3–41. doi: 10.6254/2012.38.3

Ji, L. J., Lee, A., and Guo, T. (2010). "The thinking styles of Chinese people," in *The Handbook of Chinese Psychology, 2nd Edn.*, ed M. Bond (Oxford, NY: Oxford University Press), 155–167.

Kitayama, S., Duffy, S., Kawamura, T., and Larsen, J. T. (2003). Perceiving an object and its context in different cultures: a cultural look at new look. *Psychol. Sci.* 14, 201–206. doi: 10.1111/1467-9280.02432

Masuda, T., and Nisbett, R. E. (2001). Attending holistically vs. analytically: comparing the context sensitivity of japanese and americans. *J. Pers. Soc. Psychol.* 81, 922–934. doi: 10.1037/0022-3514.81.5.922

Masuda, T., and Nisbett, R. E. (2006). Culture and change blindness. *Cogn. Sci.* 30, 381–399. doi: 10.1207/s15516709cog0000_63

Miller, J. (1978). Multidimensional same–different judgments: evidence against independent comparisons of dimensions. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 411–422.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cogn. Psychol.* 14, 247–279.

Miyamoto, Y., Nisbett, R. E., and Masuda, T. (2006). Culture and the physical environment. Holistic versus analytic perceptual affordances. *Psychol. Sci.* 17, 113–119. doi: 10.1111/j.1467-9280.2006.01673.x

Mordkoff, J. T., and Yantis, S. (1991). An interactive race model of divided attention. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 520–538.

Mordkoff, J. T., and Yantis, S. (1993). Dividing attention between color and shape: evidence of coactivation. *Percept. Psychophys.* 53, 357–366.

Nisbett, R. E., and Miyamoto, Y. (2005). The influence of culture: holistic versus analytic perception. *Trends Cogn. Sci.* 9, 467–473. doi: 10.1016/j.tics.2005.08.004

Norenzayan, A., and Nisbett, R. E. (2000). Culture and causal cognition. *Curr. Dir. Psychol. Sci.* 9, 132–135. doi: 10.1111/1467-8721.00077

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime: User's Guide.* Pittsburgh: Psychology Software Tools Inc.

Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *Br. J. Math. Stat. Psychol.* 25, 168–199.

Townsend, J. T. (1974). "Issues and models concerning the processing of a finite number of inputs," in *Human Information Processing: Tutorials in Performance and Cognition*, ed B. H. Kantowitz (Hillsdale, NJ: Erlbaum Press), 133–168.

Townsend, J. T., and Eidels, A. (2011). Workload capacity spaces: a unified methodology for response time measures of efficiency as workload is varied. *Psychon. Bull. Rev.* 18, 659–681. doi: 10.3758/s13423-011-0106-9

Townsend, J. T., and Nozawa, G. (1995). Spatio-temporal properties of elementary perception: an investigation of parallel, serial, and coactive theories. *J. Math. Psychol.* 39, 321–359. doi: 10.1006/jmps.1995.1033

Townsend, J. T., and Nozawa, G. (1997). Serial exhaustive models can violate the race model inequality: implications for architecture and capacity. *Psychol. Rev.* 104, 595–602.

Triandis, H. C. (1995). *Individualism & Collectivism.* Boulder, CO: Westview Press.

Van Zandt, T. (2000). How to fit a response time distribution. *Psychon. Bull. Rev.* 7, 424–465. doi: 10.3758/BF03214357

Wang, M.-Y., Tang, D.-L., Kao, C.-T., and Sun, V. C. (2013). "Banner evaluation predicted by eye tracking performance and the median thinking style," in *Design, User Experience, and Usability: Health, Learning, Playing, Cultural, and Cross-Cultural User Experience*, ed A. Marcus (Las Vegas, NV: Springer Berlin Heidelberg), 129–138.

Wenger, M. J., and Gibson, B. S. (2004). Using hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 708–719. doi: 10.1037/0096-1523.30.4.708

Wenger, M. J., and Townsend, J. T. (2001). "Faces as gestalt stimuli: process characteristics," in *Computational, Geometric, and Process Perspectives on Facial Cognition: Contexts and Challenges*, ed M. J. W. A. J. T. Townsend (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 229–284.

Wu, C.-H. (2006). Chong-Yong makes my life better: the effect of Chong-Yong thinking on life satisfaction. *J. Psychol. Chin. Soc.* 7, 163–176.

Wu, C.-H., and Lin, Y.-C. (2005). Development of a Zhong-Yong thinking style scale. *Indigenous Psychol. Res. Chin. Soc.* 24, 247–300. doi: 10.6254/2005.24.247

Yang, C. F. (2010). Multiplicity of Zhong Yong studies. *Indigenous Psychol. Res. Chin. Soc.* 34, 3–96. doi: 10.6254/2010.34.3

Yao, X., Yang, Q., Dong, N., and Wang, L. (2010). Moderating effect of Zhong Yong on the relationship between creativity and innovation behaviour. *Asian J. Soc. Psychol.* 13, 53–57. doi: 10.1111/j.1467-839X.2010.01300.x

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION
Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

SUPPORT
By our Swiss-based editorial team

IMPACT METRICS
Advanced metrics track your article's impact

GLOBAL SPREAD
5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK
Our network increases readership for your article

Find us on