Review

Parametric supplements to systems factorial analysis: Identifying interactive parallel processing using systems of accumulators☆

Gregory E. Cox [a],*, Amy H. Criss [b]

[a] *Vanderbilt University, United States*
[b] *Syracuse University, United States*

ARTICLE INFO

ABSTRACT

Systems Factorial Technology (SFT; Townsend and Nozawa (1995)) gains much of its power from finding tight nonparametric links between theory and data. But this power comes at a price: Applying SFT typically requires low error rates, many observations, and a guarantee of selective influence of experimental manipulations, conditions that cannot be satisfied in many fields of psychology. We present a set of parametric methods that, while lacking the full power of traditional SFT, allow its logic to be applied to situations that do not adhere to those conditions. These methods are based around building different parallel architectures from systems of Linear Ballistic Accumulators (Brown and Heathcote (2008)), including architectures that involve interactions between processes. The primary output of these methods is an estimate of the probabilities that a participant is best described by each of these architectures. In an example and set of simulations, we show that these methods are accurate and robust at identifying the processing architectures employed by a set of participants, may be estimated in maximum *a posteriori* or fully Bayesian fashion, and that hierarchical estimation allowing accurate identification with as few as three trials per participant per condition. We provide code that allows researchers to apply these methods to their own data at https://osf.io/m6ubq/.

© 2019 Elsevier Inc. All rights reserved.

Contents

---

1. Introduction

One of the core goals of cognitive psychology is to characterize the structures by which information is processed in living organisms. Theories in cognitive psychology must, therefore, specify how information from the environment and from the organism's memory conspire to lead to observable behavior. Just as psychophysics is concerned with the mapping function between external sensation and reported experience, cognitive psychology is concerned with the mapping between both internal and external information and observed behavior. And just as in psychophysics, to be able to address this concern in a practical way, cognitive psychologists must constrain both the information available to the organism as well as its range of behavior. While placing such constraints has the cost of making it harder to generalize results to situations where those constraints are loosened, it has the advantage of allowing for a rigorous characterization of the information processing architecture in a given domain.

One of the most powerful sets of tools for characterizing information processing architectures is Systems Factorial Technology (SFT; Townsend & Nozawa, 1995). Using a number of measures derived from response time (RT) distributions observed from a double factorial experimental design, SFT makes it possible to characterize the information processing architecture in the experimental domain in a purely qualitative way, that is, without recourse to specific parametric models of the mapping between information and behavior. This means that entire categories of theory (those that entail inappropriate processing architectures) can be falsified and subsequent work may be focused on those categories of theory that realize the appropriate architecture. Given the endless variety of potential theories in any cognitive domain, SFT is then an extremely valuable tool for both discovery (which

architectures are possible?) and navigation (where should future research be directed?).

Unfortunately, many domains of psychology – including much memory and decision research – have characteristics that make it difficult or impossible to apply the tools of SFT. Within these domains, there are typically high error rates, it is difficult to collect large amounts of data per individual, and researchers may be unable to strictly control the information that is available for a participant to process. In this article, we describe a set of parametric methods that use systems of accumulator models to realize different types of parallel cognitive architectures. These methods allow researchers in domains that are problematic for traditional SFT to estimate the probabilities with which participants are best-described by each architecture. While the parametric methods we describe can never achieve the impressive generality of SFT in its traditional non-parametric form – and are currently limited to parallel architectures – these methods can, if carefully applied and interpreted, be an effective supplement to SFT in psychological domains to which traditional non-parametric methods are difficult to apply.

1.1. Applying SFT—and its restrictions

SFT is applied in situations where multiple sources of information potentially impinge on a participant, who must use those sources to make a decision. SFT has the power to identify, in a non-parametric way, the potential relationships (the "architecture") between how those sources are processed in order to lead to that decision. This power comes from the ability to derive qualitative predictions for these architectures in a carefully controlled factorial experiment. Typically, this experiment focuses on two information sources, hence it is a "double factorial" experiment and involves separately manipulating the *strength* of two sources

of information supplied by the stimuli in the experiment, where manipulating the strength of one source of information *selectively influences* just that source—manipulating one source has no effect on the other. The participant is required to make a binary response (or at least a response that can be classified in a binary fashion) to the stimuli in the experiment on the basis of the information provided by each source.

Under the condition of selective influence (as well as other technical conditions, see Townsend & Nozawa, 1995), different processing architectures yield qualitatively different predictions regarding the distributions of response times in each condition of the experiment. These architectures can be characterized by the order of processing and by the decision rule applied to the results of such processing. The order of processing is generally divided into two types: *parallel*, in which both sources of information are processed at the same time; and *serial*, in which the two sources are processed sequentially. There are also two kinds of decision rule: *self-terminating* processes are those in which only one source needs to be finished before a response can be made; and *exhaustive* processes are those that require both sources to be processed before a decision can be made. A final processing architecture is the *coactive* architecture in which both sources of information interact and get combined into a single source prior to processing. Coactive processing can be viewed as a special case of parallel processing, though obviously there can be no distinction between self-terminating or exhaustive decision rules since only one process goes on.

Traditional SFT allows the researcher who conducts a double factorial experiment to identify which of these processing architectures each participant uses to accomplish the task, without needing to specify or estimate any parameters. This comes at a price, however. As we note below, applying the tools of SFT entails several assumptions and caveats, both practical and theoretical.

1.1.1. Form of the experiment

First, the domain of interest must allow for the implementation of a double factorial experiment: It must be possible to identify two nominally separate information sources in the domain (like hue and saturation in color perception or item and associative information in memory). And the experimental task must be such that responses can be made with a single action (like pressing a key, and unlike, e.g., speaking a sentence) that can be categorized into correct and incorrect responses. In general, though, many experimental tasks in cognitive psychology already adhere to the demands of the double factorial design, so this does not often present much of an impediment.

1.1.2. Estimating RT distributions

A significant practical hurdle is posed by the fact that SFT relies on measures that are based on RT distributions. Getting a good estimate of a RT distribution using traditional non-parametric methods like the empirical CDF requires on the order of hundreds of trials per individual per condition. Because of this need for extensive data collection, SFT is typically employed in experiments using a relatively small number of participants, each of whom contributes a large number of observations. In addition, many SFT measures make use of only *correct* responses, such that if accuracy is less than perfect, it may require even more trials to get a good estimate of the RT distributions for correct responses.

1.1.3. Errors, learning, and process mixtures

To the extent that errors reflect additional processes beyond those required to perform the task (e.g., guessing processes), imperfect performance can result in RT distributions that reflect a mixture of processing architectures rather than just those of interest to the researcher. And if a researcher must compensate by collecting a large amount of data per individual, doing so typically requires repeating individual stimuli many times over the course of the experiment. If the domain is one that is already so deeply ingrained that experimental experience only represents a small drop in a large pool (many perceptual domains would likely fall into this category), this may not be a problem. But if the domain entails novel or unusual stimuli – or the experimental task itself is sufficiently unusual – then participants are likely *learning* or *adapting* during the experiment. This means that the information processing architecture can change between the beginning and end of the experiment, such that the aggregate RT distributions obtained reflect a mixture of different processes. For example, participants in a categorization experiment may initially rely on short-term memory for recent items and transition to responding that is based on a direct stimulus–response mapping (e.g., Logan, 1988; Nosofsky, Cao, Cox, & Shiffrin, 2014; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

1.1.4. Selective influence and interacting processes

The qualitative signatures that are detectable by SFT are only diagnostic under a set of assumptions about the relationships between the underlying processes and observed behavior. Perhaps the most crucial assumption is that of *selective influence*, that the manipulation one of the two sources of information in the double factorial experiment has no affect on the *other* source (Townsend & Thomas, 1994). This assumption is unlikely to be satisfied in many domains of psychology, for example in memory (e.g., the way an event is encoded may depend on the way *other* events are encoded) or decision research (e.g., the effect of an increase in payoffs may depend on the other options that are available). Eidels, Houpt, Altieri, Pei, and Townsend (2011) identified two loci for violations of selective influence: Selective influence may fail at the level of the *inputs* to the information processing architecture, that is, that the stimuli themselves are constructed in a way that violates selective influence. It may also fail due to interactions that occur *within* the information processing architecture itself, such that the effective architecture can change depending on its internal activity (maybe when both inputs are at high strength, there is facilitation such that the system exhibits coactive behavior but if both inputs are at low strength, the facilitation is not present as the system exhibits independent parallel behavior). In situations where selective influence is likely to have been violated – in either or both of the two ways described above – SFT can offer only little guidance, and concluding which processing architectures are plausible in these situations often requires extensive simulation (Cox & Criss, 2017; Eidels et al., 2011).

Despite the difficulties they cause when trying to apply SFT, violations of selective influence that result from interactions *within* a processing architecture are, in fact, an essential characteristic many classes of architectures (Townsend & Wenger, 2004). An extreme example of such a system is a fully coactive architecture, in which all information in a task is pooled together into a single channel which is used to guide behavior. Although predictions have been made for certain classes of coactive architecture (Houpt & Townsend, 2011; Townsend & Nozawa, 1995), there is as yet no general signature that coactive architectures are known to produce that can be detected using SFT. There are also no known general signatures that are characteristic of all partially-interactive architectures, that is, those that only pool only *some* of their inputs (Cox & Criss, 2017). The potential prevalence of parallel interactive systems suggests the need for some way to detect when they may be operating, and this seems especially so for domains that involve complex sources of information (like memory, reasoning, or decision making).

1.1.5. Statistical tests

While much of the power of SFT derives from the fact that its tools are non-parametric, the qualitative nature of SFT leads to a paradox, in that it becomes difficult to quantify the degree of statistical support for any conclusions reached via SFT. When there is a massive amount of data per individual there is little need for statistics, but if uncertainty remains, either because there is not enough data or the data are extremely variable, good statistics are essential. Although non-parametric or semi-parametric methods are viable (Houpt, Blaha, McIntire, Havig, & Townsend, 2014; Houpt, Heathcote, & Eidels, 2017; Houpt, MacEachern, Peruggia, Townsend, & Van Zandt, 2016), in practice it is difficult to directly quantify the degree of statistical support for a particular architecture without recourse to a parametric model at some stage (Cox & Criss, 2017, 2017; Houpt & Fifić, 2017; Thiele, Haaf, & Rouder, 2017).

1.2. Accumulator models and SFT

As described above, SFT provides nonparametric tools for identifying information processing architectures when

1. there are sufficient observations per participant per condition to allow for good estimates of RT distributions;
2. error rates are sufficiently low and the system is stable over time, such that RT distributions provide a pure view of behavior of the processes of interest;
3. selective influence is satisfied both at the *input* level and *within* the possible cognitive architectures themselves.

To address situations with relatively few observations per participant, where error rates are non-trivial, or where there may be interactions among processes, researchers have turned to various *accumulator* models of information processing. These models have the property that specific actions (like pressing a particular response key) are associated with an "accumulator" that has, at any given time, a level of activation that represents the likelihood that a participant will take the associated action. They are called "accumulators" because their level of activation is a function of the amount of evidence or support that has been accumulated for that action at any given time. The amount of evidence/support is a function of information that flows into the system either from the environment or from a participant's memory. When the amount of accumulated activation/support for a particular action reaches a threshold (which may be different for different actions), the associated action is taken. Accumulator models of myriad types have proven widely applicable across psychology, including vision (Smith & Vickers, 1988; Vickers, 1970), memory (Cox & Shiffrin, 2017; Donkin & Nosofsky, 2012; Ratcliff, 1978), learning (Logan, 1988), categorization (Nosofsky & Palmeri, 1997), and decision making (Busemeyer & Townsend, 1993; Edwards, 1965; Link, 1975; Link & Heath, 1975).

Accumulators may be independent, in which case they form a race model—whichever accumulator "wins the race" against the others determines how a participant responds. Independent race models have a long history of being associated with SFT, in large part because they serve as a useful benchmark model; SFT's capacity measures compare the speed and/or accuracy of behavior *relative to* what would be expected from an independent race model (Townsend & Altieri, 2012). Another popular class of accumulator models are diffusion/random walk models, which have seen much use in applications of SFT to categorization (e.g., Fifić, Nosofsky, & Townsend, 2008; Little, Nosofsky, & Denton, 2011). In a random walk or diffusion model (e.g., Ratcliff, 1978), there are only two accumulators each of which is associated with one of two mutually exclusive responses (e.g., an item can be classified as belonging to category A or B). Because they are mutually exclusive,
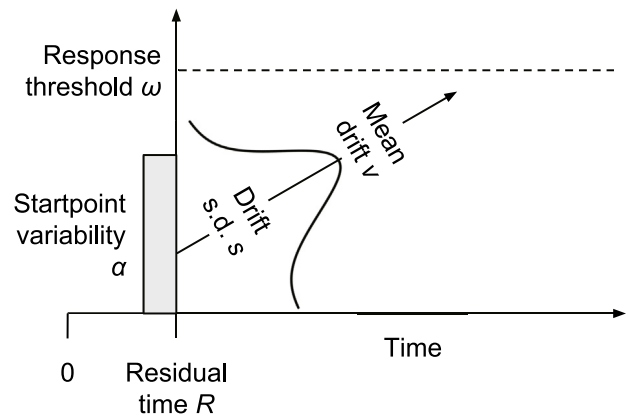

Fig. 1. A single Linear Ballistic Accumulator (LBA), illustrating the role of each parameter. See the main text for detail on each parameter.

if incoming information favors one option, in must also *disfavor* the other. Thus, a random walk model can be viewed as an *interactive* race model between two accumulators that are perfectly negatively correlated (an increase in activation for one accumulator results in an equal decrease in activation for the other). Conversely, it is possible for a parallel race model to mimic the behavior of a random walk model in many cases, so long as the inputs to the accumulators are anti-correlated (Donkin, Brown, & Heathcote, 2011; Teodorescu & Usher, 2013).[1]

Regardless of how they are set up, accumulator models yield predictions not only for which action is taken, but *how long* it takes to choose that action. In other words, they provide a formal way of deriving the RT distributions and response probabilities needed in SFT, from a set of parameters describing the inputs and thresholds for each accumulator. To the extent that these model parameters can be identified with constructs in SFT like "capacity" (Eidels, Donkin, Brown, & Heathcote, 2010a), accumulator models thus provide *parametric* ways to address similar questions about the architecture of information processing. As we shall see throughout this article, there are advantages to using parametric models in SFT, specifically, that they do not require as much data for efficient estimation and they can deal elegantly with high rates of error, process mixtures, and certain violations of selective influence. They have the disadvantage, however, of making a researcher's conclusions conditional on the choice of model, which is a distinct drawback from the full non-parametric power of SFT. However, we argue that so long as this crucial fact is kept in mind, the flexibility afforded by parametric methods can outweigh their drawbacks.

1.3. Linear ballistic accumulator models

We focus in this article on one class of parallel accumulator models that strikes a good balance between flexibility and tractability, namely, the Linear Ballistic Accumulator (LBA Brown & Heathcote, 2008). According to this model, depicted in Fig. 1, the initial level of activation/support for each accumulator is sampled from a uniform distribution between 0 and α. Variability at this stage is meant to reflect the idea that there may be activation from sources other than the stimuli to be presented on a given trial, including residual activation from previous trials, activation from other stimuli in the environment, or even "premature" activation from the stimulus itself (cf. Laming, 1968; Ratcliff & Rouder, 1998;

---

[1] This is reminiscent of the mimicry between serial and parallel processes—so long as the *inputs* to the parallel racing accumulators are set up appropriately (Townsend, 1976).

Rouder, 1996). Upon presentation of a stimulus, the accumulator's activation evolves over time in a linear fashion, with a rate sampled from a normal distribution with mean v and standard deviation s (for present purposes, s represents a scaling parameter, but see Donkin, Brown, & Heathcote, 2009). If the accumulator's activation reaches a threshold ω before any other accumulators have reached their respective thresholds, then the response associated with the winning accumulator is produced. The final response time is the time taken for the winning accumulator to reach its threshold plus a residual time R that reflects the time needed to detect and begin processing the stimulus as well as the time needed to execute the chosen response.

The apparent simplicity of the LBA belies its ability to provide good quantitative accounts of both response probability and the shapes of RT distributions across a variety of domains (e.g., Rae, Heathcote, Donkin, Averell, & Brown, 2014). Moreover, the analytical tractability of the LBA has the practical benefit that many LBA's can be "plugged together" to form entire systems of accumulators that realize particular cognitive models (Nosofsky et al., 2014; Trueblood, Brown, & Heathcote, 2014; Zandbelt, Purcell, Palmeri, Logan, & Schall, 2014). Indeed, the LBA has already been used to augment traditional SFT measures by using the mean evidence accumulation rates (the v parameters) to estimate processing capacity (Eidels et al., 2010a). Evidence accumulation rates in the LBA can also be used to model violations of selective influence *separately* from any interactions that may occur during the accumulation process. Further, it is possible to realize different processing architecture as systems of LBA's, thereby allowing researchers not just to measure processing capacity, but to infer which architecture(s) best describe the accuracy and response times of each participant in an experiment (Cox & Criss, 2017).

2. Processing architectures

As mentioned above, the goal of the methods discussed in this article is to yield an estimate of the probability for each participant that their behavior is a product of a particular type of parallel processing architecture. In this section, we define those architectures as well as the likelihood functions associated with each architecture that will allow us to estimate their relative probability.

All of the processing architectures we explore in this paper can be constructed from systems of linear ballistic accumulators. A single such accumulator was described above, and in this section we describe how multiple accumulators can be used to implement different parallel processing architectures. All of these architectures are depicted schematically in Table 1, according to the type of processing interactions and stopping rule. These different architectures cover the ways in which information from two sources, labeled source A and source B, can lead to either a positive ("yes") decision or a negative ("no") decision. This choice structure – two possible choices and two sources of information used to make that choice – mirrors that of the double factorial experimental paradigm used for systems factorial analysis, as described both above and below.

2.1. Finishing time distribution for a single accumulator

The predicted distributions of response times and response probabilities depend on the PDF and CDF of the finishing time distribution for a single accumulator, that is, the distribution that describes when (and if) the evidence level of a single accumulator reaches a threshold. For ease of exposition, we reproduce these functions from Brown and Heathcote (2008), letting $f(t|\beta, \alpha, R, v, s)$ denote the probability density function (PDF) at time t and $F(t)$ the cumulative density function (CDF) at time t, given threshold ω, startpoint variability α, residual time R, mean accumulation rate

v, and accumulation rate standard deviation s (for reference, see Fig. 1):

$$f(t|\omega, \alpha, R, v, s) = \frac{1}{\alpha}\left[-v\Phi(\xi) + s\phi(\xi) + v\Phi(\zeta) - s\phi(\zeta)\right] \quad (1)$$

$$\begin{aligned}F(t|\omega, \alpha, R, v, s) = 1 + \frac{1}{\alpha}\{&[\omega - \alpha - v(t-R)]\Phi(\xi) \\ &- [\omega - v(t-R)]\Phi(\zeta) \\ &+ s(t-R)[\phi(\xi) - \phi(\zeta)]\}\end{aligned} \quad (2)$$

where

$$\xi = \frac{\omega - \alpha - v(t-R)}{s(t-R)}$$

$$\zeta = \frac{\omega - v(t-R)}{s(t-R)}$$

and $\phi(\cdot)$ is the standard normal PDF and $\Phi(\cdot)$ is the standard normal CDF.

Note that, as in the original formulation of the LBA (Brown & Heathcote, 2008), $f(t|\omega, \alpha, R, v, s)$ allows for a non-zero (although typically trivial) probability of a negative accumulation rate, in which case evidence in a particular accumulator will never attain the threshold value ω in positive time. It is straightforward to modify the drift rate distribution to eliminate this feature of the model if needed (Heathcote & Love, 2012; Terry et al., 2015), although in practice we have found this feature useful because it acts as a kind of "gating inhibition" (Purcell, Schall, Logan, & Palmeri, 2012; Usher & McClelland, 2001) that takes an accumulator "out of the running" and allows the final response to be determined by one of the other accumulators. As we describe below, an LBA never acts alone.

2.2. Independent

When the two information sources A and B are processed independently, observed responses and response times are a function of four separate accumulators: two for each source, one of which corresponds to the accumulation of positive evidence from that source (that leads to a "yes" response) while the other corresponds to the accumulation of negative evidence from that source (that leads to a "no" response). We use subscripts to denote the parameters and finishing time distributions associated with the two accumulators from each source and superscripts to indicate whether the accumulator corresponds to positive or negative evidence, like so:

$f_A^+\left(t|\omega^+, \alpha^+, R, v^+, s^+\right), F_A^+\left(t|\omega^+, \alpha^+, R, v^+, s^+\right)$: The PDF and CDF for the accumulator corresponding to positive evidence from source A.

$f_B^+\left(t|\omega^+, \alpha^+, R, v^+, s^+\right), F_B^+\left(t|\omega^+, \alpha^+, R, v^+, s^+\right)$: The PDF and CDF for the accumulator corresponding to positive evidence from source B.

$f_A^-\left(t|\omega^-, \alpha^-, R, v^-, s^-\right), F_A^-\left(t|\omega^-, \alpha^-, R, v^-, s^-\right)$: The PDF and CDF for the accumulator corresponding to negative evidence from source A.

$f_B^-\left(t|\omega^-, \alpha^-, R, v^-, s^-\right), F_B^-\left(t|\omega^-, \alpha^-, R, v^-, s^-\right)$: The PDF and CDF for the accumulator corresponding to negative evidence from source B.

Note that we have also made several simplifying assumptions regarding the parameters that describe each accumulator, specifically that there is a single residual time R associated with all accumulators and that parameters are shared between both positive accumulators, regardless of source, and both negative accumulators, regardless of source. These assumptions do not follow from

Table 1
Schematic depictions of different processing architectures implemented as systems of linear ballistic accumulators.
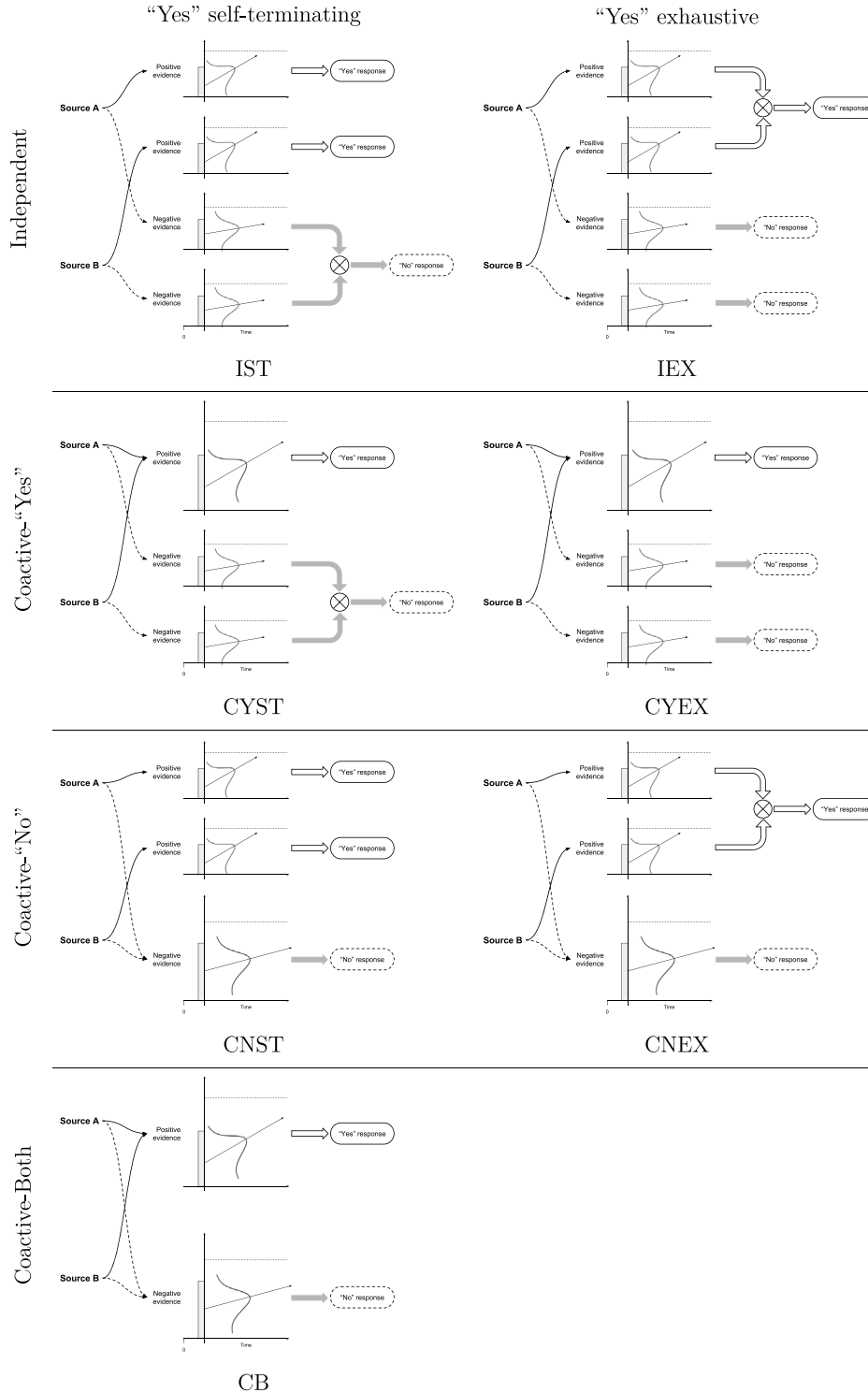
psychological theory, but rather from model identifiability: In the absence of selective influence (which we do not assume), it is not possible to know whether a given positive (or negative) response was a function of information from source A or source B, so we can only identify parameters associated with the response rather than the source. Likewise, even in situations where it is plausible that one source of information is processed earlier or faster than the other (cf. Cox & Criss, 2017; Hendrickson, Navarro, & Donkin), unless selective influence on processing time is assumed, it is not possible to know *which* source is faster or slower, so we can only

identify a single residual time. In the Discussion, we consider situations in which additional experimental manipulations can help enforce the identifiability needed for these parameters, as well as how different residual times may be used to implement serial-like architectures, but at the moment these parameter restrictions between sources must be considered a constraint of the present approach. Finally, because the evidence rate standard deviation s functions as a scaling parameter (though, again, see Donkin et al., 2009), we set $s^+ = s^- = 1$ for all accumulators.

Two response rules are possible in this context: According to one rule, a "yes" response is made when *either* of the two positive accumulators reaches its threshold (and zero or one negative accumulator has reached threshold), while a "no" response can only be made when *both* negative accumulators reach their thresholds (and *both* positive accumulators are below threshold). For brevity, we label this architecture "IST", since the sources are processed **I**ndependently and "yes" responses are **S**elf-**T**erminating. The likelihoods of a "yes" or a "no" response at time t for this architecture are

$$l_{IST}(\text{"yes"}, t) = \{f_A^+\left(t|\omega^+, \alpha^+, R, v^+, s^+\right)$$
$$\times \left[1 - F_B^+\left(t|\omega^+, \alpha^+, R, v^+, s^+\right)\right]$$
$$+ f_B^+\left(t|\omega^+, \alpha^+, R, v^+, s^+\right)$$
$$\times \left[1 - F_A^+\left(t|\omega^+, \alpha^+, R, v^+, s^+\right)\right]\}$$
$$\times \left[1 - F_A^-\left(t|\omega^-, \alpha^-, R, v^-, s^-\right)\right.$$
$$\times F_B^-\left(t|\omega^-, \alpha^-, R, v^-, s^-\right)]$$

$$l_{IST}(\text{"no"}, t) = [f_A^-\left(t|\omega^-, \alpha^-, R, v^-, s^-\right) F_B^-\left(t|\omega^-, \alpha^-, R, v^-, s^-\right)$$
$$+ f_B^-\left(t|\omega^-, \alpha^-, R, v^-, s^-\right)$$
$$\times F_A^-\left(t|\omega^-, \alpha^-, R, v^-, s^-\right)]$$
$$\times \left[1 - F_A^+\left(t|\omega^+, \alpha^+, R, v^+, s^+\right)\right]$$
$$\times \left[1 - F_B^+\left(t|\omega^+, \alpha^+, R, v^+, s^+\right)\right]$$

which, because of the parameter restrictions imposed above for identifiability, can be simplified to

$$l_{IST}(\text{"yes"}, t) = 2f^+\left(t|\omega^+, \alpha^+, R, v^+, 1\right)$$
$$\times \left[1 - F^+\left(t|\omega^+, \alpha^+, R, v^+, 1\right)\right]$$
$$\times \left[1 - F^-\left(t|\omega^-, \alpha^-, R, v^-, 1\right)^2\right] \quad (3)$$

$$l_{IST}(\text{"no"}, t) = 2f^-\left(t|\omega^-, \alpha^-, R, v^-, 1\right) F^-\left(t|\omega^-, \alpha^-, R, v^-, 1\right)$$
$$\times \left[1 - F^+\left(t|\omega^+, \alpha^+, R, v^+, 1\right)\right]^2 \quad (4)$$

where we have dropped the A and B subscripts since the parameters are shared between the two sources.

Using the same simplified notation, we can express the likelihoods of a "yes" or "no" response at time t under the alternative stopping rule in which *both* positive accumulators must reach threshold to make a "yes" response (while *neither* negative accumulator has reached threshold) but only one negative accumulator needs to reach threshold for a "no" response (while zero or one positive accumulator has reached threshold). This architecture is labeled IEX since, again, processing is **I**ndependent but "yes" responses require **EX**haustive processing of the positive evidence from each source:

$$l_{IEX}(\text{"yes"}, t) = 2f^+\left(t|\omega^+, \alpha^+, R, v^+, 1\right) F^+\left(t|\omega^+, \alpha^+, R, v^+, 1\right)$$
$$\times \left[1 - F^-\left(t|\omega^-, \alpha^-, R, v^-, 1\right)\right]^2 \quad (5)$$

$$l_{IEX}(\text{"no"}, t) = 2f^-\left(t|\omega^-, \alpha^-, R, v^-, 1\right)$$
$$\times \left[1 - F^-\left(t|\omega^-, \alpha^-, R, v^-, 1\right)\right]$$
$$\times \left[1 - F^+\left(t|\omega^+, \alpha^+, R, v^+, 1\right)^2\right] \quad (6)$$

Note that, as intuition would dictate, the IEX likelihoods are just a mirror image of the IST likelihoods.

2.3. Coactive-"yes"

When accumulators for positive evidence interact with one another, we presume this takes the form of "pooling" or partial co-activation, and so we refer to such architectures as *coactive-"yes"*. We implement this partial co-activation by simple addition:

instead of two separate accumulators for positive evidence from each source, there is a single positive accumulator that *sums* the inputs associated with each source. The principal effect of this is to increase the signal-to-noise ratio of the evidence distribution, since it becomes a normal distribution with mean $2v^+$ and standard deviation $\sqrt{2}$ (since it is the *variances* that add, not the standard deviations). We presume that the parameters associated with the threshold for positive responses double as well to accommodate this stronger input, such that there is now a single accumulator for positive evidence that is described by a threshold $2\omega^+$, startpoint variability $2\alpha^+$, mean evidence rate $2v^+$, and evidence rate standard deviation $\sqrt{2}$. By increasing threshold as well as signal-to-noise ratio, this type of pooling captures the idea that coactivation involves an enhancement in the *quality* of evidence that is separate from a simple speedup (which could be modeled by varying just the drift rates v^+ in an independent model). This model is, therefore, equivalent to keeping threshold parameters constant (at ω^+ and α^+) but *averaging* the inputs between sources, such that the evidence distribution still has mean v^+ but has standard deviation $\frac{\sqrt{2}}{2}$.

Having established the form of interaction involved, we can express the likelihoods for "yes" and "no" responses at time t according to an architecture in which "no" responses are exhaustive (this architecture is labeled CYST, since processing is **C**oactive-"**Y**es" and "yes" responses are **S**elf-**T**erminating):

$$l_{CYST}(\text{"yes"}, t) = f^+\left(t|2\omega^+, 2\alpha^+, R, 2v^+, \sqrt{2}\right)$$
$$\times \left[1 - F^-(t|\omega^-, \alpha^-, R, v^-, 1)^2\right] \quad (7)$$

$$l_{CYST}(\text{"no"}, t) = 2f^-(t|\omega^-, \alpha^-, R, v^-, 1) F^-(t|\omega^-, \alpha^-, R, v^-, 1)$$
$$\times \left[1 - F^+\left(t|2\omega^+, 2\alpha^+, R, 2v^+, \sqrt{2}\right)\right] \quad (8)$$

And we can do the same for the CYEX architecture in which processing is **C**oactive-"**Y**es" but "no" responses can be made whenever a single negative accumulator reaches threshold, making "yes" responses **EX**haustive:

$$l_{CYEX}(\text{"yes"}, t) = f^+\left(t|2\omega^+, 2\alpha^+, R, 2v^+, \sqrt{2}\right)$$
$$\times \left[1 - F^-(t|\omega^-, \alpha^-, R, v^-, 1)\right]^2 \quad (9)$$

$$l_{CYEX}(\text{"no"}, t) = 2f^-(t|\omega^-, \alpha^-, R, v^-, 1)$$
$$\times \left[1 - F^-(t|\omega^-, \alpha^-, R, v^-, 1)\right]$$
$$\times \left[1 - F^+\left(t|2\omega^+, 2\alpha^+, R, 2v^+, \sqrt{2}\right)\right] \quad (10)$$

2.4. Coactive-"no"

Just as it is possible to pool positive evidence, when negative evidence from each source is pooled into a single accumulator, we refer to this form of interaction as coactive-"no". Just as above, the result is that instead of two separate accumulators for negative evidence from each source, there is a single accumulator for negative evidence with response threshold $2\omega^-$, startpoint variability $2\alpha^-$, mean evidence rate $2v^-$, and evidence rate standard deviation $\sqrt{2}$ (so, again, there is an increase in the signal-to-noise ratio for the accumulated evidence).

In the event that only a single positive accumulator is needed to make a positive response, we have a CNST (**C**oactive-"**N**o" with **S**elf-**T**erminating "yes" responses) architecture with corresponding likelihoods:

$$l_{CNST}(\text{"yes"}, t) = 2f^+(t|\omega^+, \alpha^+, R, v^+, 1)$$
$$\times \left[1 - F^+(t|\omega^+, \alpha^+, R, v^+, 1)\right]$$
$$\times \left[1 - F^-\left(t|2\omega^-, 2\alpha^-, R, 2v^-, \sqrt{2}\right)\right] \quad (11)$$

(a) Probability of making a "yes" response.

(b) Mean response time for "yes" responses.

(c) Mean response time for "no" responses.
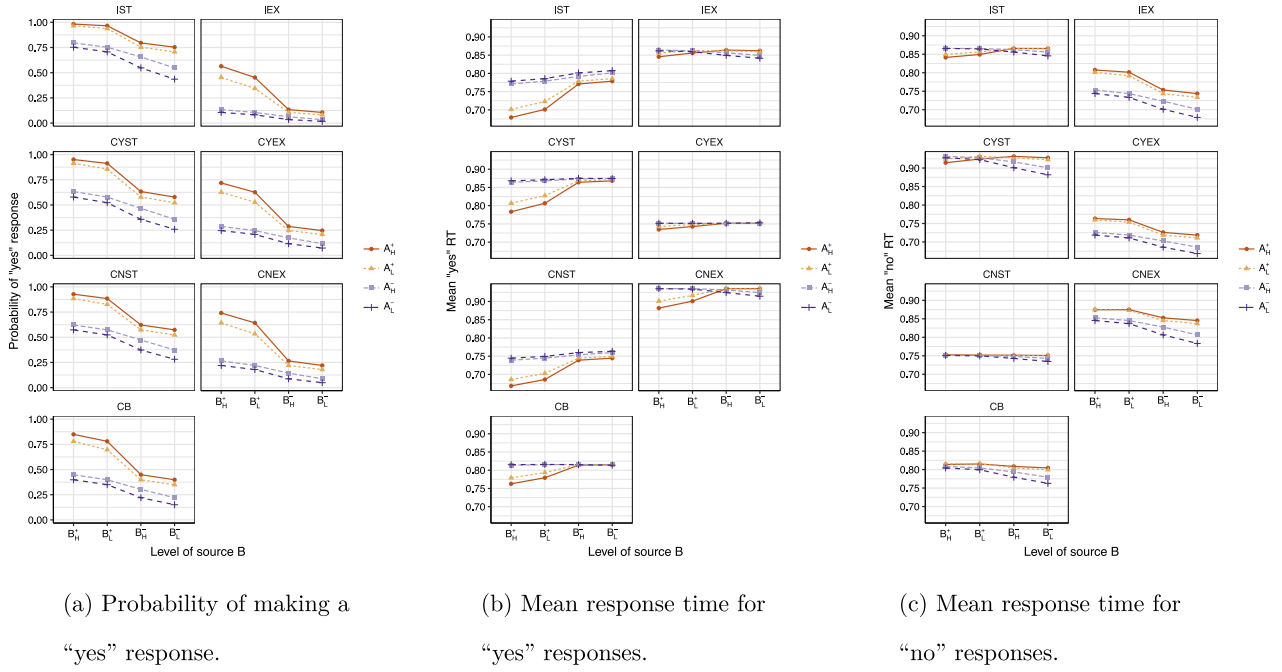
Fig. 2. Predictions across different parallel architectures using a single set of parameters for a double-factorial paradigm involving the manipulation of the strength of two sources, A and B, across four levels (A_H^+, A_L^+, A_H^-, and A_L^-, same for B) as described in the main text. Note that these predictions are for just one set of parameters and are not necessarily representative of all parameter settings. Parameter values: $\bar{\omega} = 5.6$; $v = 0.5$; $w = 0.5$; $R = 0.2$; $\beta_{ij} = 3.8$ for each i, j; δ_{ij} as defined in Table 4.

$$l_{CNST}(\text{"no"}, t) = f^- \left(t | 2\omega^-, 2\alpha^-, R, 2v^-, \sqrt{2} \right)$$
$$\times \left[1 - F^+(t | \omega^+, \alpha^+, R, v^+, 1) \right]^2 \quad (12)$$

And for the CNEX architecture in which processing is **C**oactive-"**N**o" and "yes" responses are **EX**haustive, we have:

$$l_{CNEX}(\text{"yes"}, t) = 2f^+(t | \omega^+, \alpha^+, R, v^+, 1) F^+(t | \omega^+, \alpha^+, R, v^+, 1)$$
$$\times \left[1 - F^- \left(t | 2\omega^-, 2\alpha^-, R, 2v^-, \sqrt{2} \right) \right] \quad (13)$$

$$l_{CNEX}(\text{"no"}, t) = f^- \left(t | 2\omega^-, 2\alpha^-, R, 2v^-, \sqrt{2} \right)$$
$$\times \left[1 - F^+(t | \omega^+, \alpha^+, R, v^+, 1)^2 \right] \quad (14)$$

2.5. Coactive-both

The final architecture we consider is a fully coactive one (labeled CB for **C**oactive-**B**oth) in which *both* positive and negative evidence is pooled together from both sources into a single pair of accumulators. In this architecture, responding is always "self-terminating" in that the first accumulator to reach threshold – whether positive or negative – determines the response. The associated likelihoods for "yes" and "no" responses at time t are then:

$$l_{CB}(\text{"yes"}, t) = f^+ \left(t | 2\omega^+, 2\alpha^+, R, 2v^+, \sqrt{2} \right)$$
$$\times \left[1 - F^- \left(t | 2\omega^-, 2\alpha^-, R, 2v^-, \sqrt{2} \right) \right] \quad (15)$$

$$l_{CB}(\text{"no"}, t) = f^- \left(t | 2\omega^-, 2\alpha^-, R, 2v^-, \sqrt{2} \right)$$
$$\times \left[1 - F^+ \left(t | 2\omega^+, 2\alpha^+, R, 2v^+, \sqrt{2} \right) \right] \quad (16)$$

2.6. Architecture predictions

As a supplement to the interactive parallel simulations reported by Eidels et al. (2011) and Cox and Criss (2017), we report a limited set of simulations from each of the architectures described above.

These simulations come from a double factorial experiment with the same form as that described in the example and simulations below, i.e., two sources (A and B) can each take one of four strength values, producing $4 \times 4 = 16$ total conditions. As shown in Fig. 2(a), several of these architectures yield similar predictions regarding the probability of making a positive "yes" response across conditions – particularly those that share the same decision rule (ST vs. EX) – but they make divergent predictions for how response times change across conditions (Figs. 2(b) and 2(c)).

Simulations using a range of parameter values for drift rates and thresholds illustrate, however, that these architectures can yield a wide range of qualitative forms for critical measures in SFT like the survivor interaction contrast (SIC) function (Fig. 3). In particular, SIC functions produced for responses that are generated by interactive processes are capable of spanning a variety of qualitative forms, including those that look like varieties of independent processing, meaning that qualitative measures are not always sufficient to identify a single architecture (but combining constraints from multiple qualitative measures, like SIC and capacity functions, can help on this front Cox & Criss, 2017; Eidels, Donkin, Brown, & Heathcote, 2010b). In contrast, the methods laid out in this paper focus on developing *quantitative* measures of support for each architecture, which we proceed to describe.

3. Estimation and example

By specifying the LBA implementation of the model architectures described above, it is possible to simultaneously estimate not only the parameters associated with the accumulators but the *probability* with which they are arranged according to each of the seven architectures. In this section, we use a practical demonstration to illustrate how these quantities can be estimated and interpreted.

3.1. Experiment

The data for this demonstration come from a conceptual replication of the experiment reported by Cox and Criss (2017), the
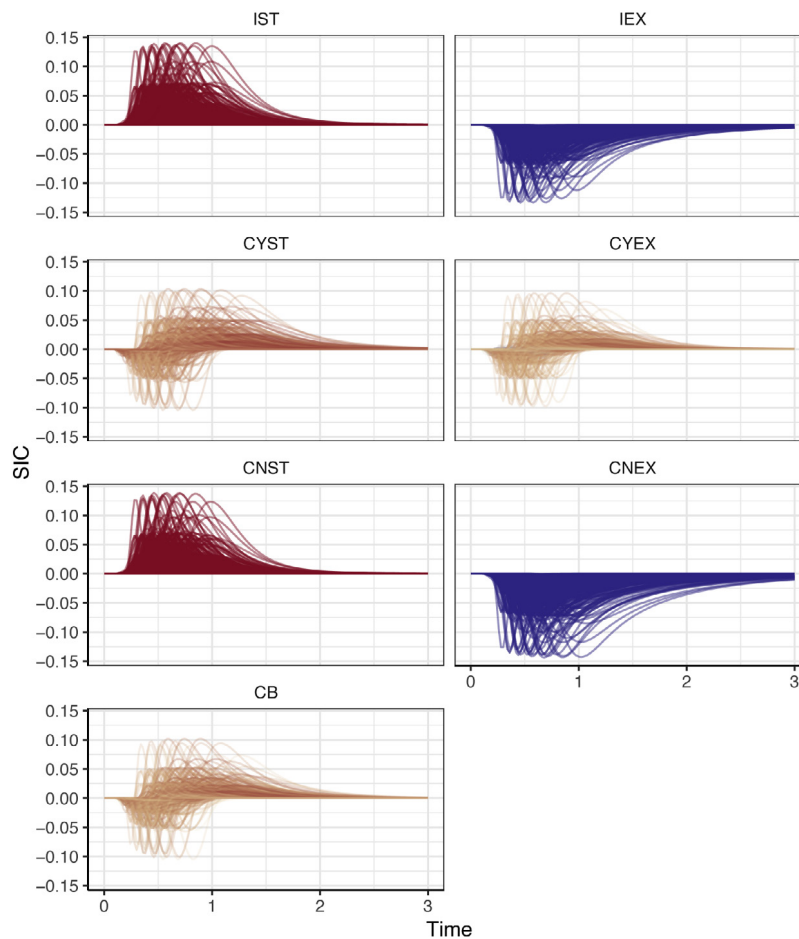
Fig. 3. Examples of predicted survivor interaction contrast (SIC) functions under different architectures under a range of different parameter settings. SIC's were computed using response time distributions for "yes" responses in the conditions in which both sources yield evidence that favors a "yes" response, i.e., $A_H^+B_H^+$, $A_H^+B_L^+$, $A_L^+B_H^+$, and $A_L^+B_L^+$.

details of which are presented in the Appendix. The goal of both the previous experiment and the partial replication presented here was to help characterize the processes by which two kinds of information are retrieved from long-term event memory: *Item* information that pertains to the perceptual and conceptual content of individual events; and *associative* information that pertains to how individual events co-occur. The experiments involved repeated blocks with two phases each, a study phase during which pairs of stimuli were presented and a recognition test phase in which pairs of items were presented and participants were required to say whether or not *that exact pairing of items* had been presented during the study phase.

Thus, the judgments on each test trial entail processing information from two sources: Retrieving item information in response to the test cue – were these items shown during the study phase? – as well as associative information—were these items *shown together* during the study phase? Both experiments implement a double factorial paradigm in which the strengths of these two kinds of information are manipulated in each test trial by either presenting the items more often during the study phase (thus boosting the strength of item information), presenting the complete pair more often during the study phase (thus boosting the strength of associative information), or both. In the original experiment reported by Cox and Criss (2017), these two repetition manipulations were carried out independently, with the consequence that repeating a whole pair during study also lead to more overall repetitions of the items within the pair. In the present conceptual replication, the strength of item information is manipulated in terms of the *total number* of times an item appears

at study, such that an item that appears in a repeated pair is shown less often by itself. The question is whether this different method of jointly manipulating item and associative strength leads to the same conclusions regarding how these two kinds of information are retrieved.

3.2. Results

The mean response probabilities and response times are given in Tables 2 and 3, respectively, and illustrate how the proportion of positive responses tends to increase with the strength of either source of information. In addition, correct response times are faster for targets and slower for distractors as the strength of either source of information is increased.

Table 2
Mean probability of giving a positive response to each pair type (standard deviation in parentheses).

	Associative strength			
	A_H^+	A_L^+	A_H^-	A_L^-
I_H^+	0.59 (0.25)	0.53 (0.23)	0.34 (0.24)	0.37 (0.25)
I_L^+	0.58 (0.25)	0.51 (0.25)	0.33 (0.20)	0.39 (0.21)
I_H^-	0.26 (0.16)	0.23 (0.21)	0.26 (0.20)	0.26 (0.19)
I_L^-	0.27 (0.20)	0.22 (0.18)	0.21 (0.17)	0.22 (0.20)

(Item strength)

3.3. Estimating architecture probability

Despite using the double factorial design necessary for SFT, a typical systems factorial analysis cannot be performed on these data for several reasons. First, as is clear from Table 2, error rates are high in this task, as is typical of many long-term recognition memory tasks. High error rates are not necessarily a problem for traditional SFT, where error rates can play a role in measuring capacity (Townsend & Altieri, 2012), but a high proportion of errors reduces the number of trials available for estimating the distributions of correct response times. Indeed, this is the second reason why traditional SFT methods cannot be used on these data, that is, there relatively few trials per participant per condition. Moreover, the number of trials varies between participants since it depends on how many study/test blocks they had time to complete. Finally, it is not clear whether our repetition manipulation selectively influences only item or only associative memory.

Our analysis estimates two kinds of quantities:

1. The LBA parameters that describe the evidence accumulation rates, response thresholds, and residual time across each condition in the double factorial experiment.
2. The probabilities with which those LBA's are arranged according to each of the architectures described above.

Further, we estimate these quantities *hierarchically*, that is, simultaneously at the individual level as well as the group level. Hierarchical estimation has a long history in statistics – particularly Bayesian statistics – and is especially useful in situations where there is relatively little data at the individual level because it allows for so-called "partial pooling" of information across individuals while still allowing for individual variability. In the simulations below, we explore the consequences of partial pooling across individuals for our estimation procedure. And in this example, we illustrate the relationship between maximum likelihood estimation and Bayesian estimation.

3.3.1. Likelihood function

Above, we specified how each LBA architecture assigns a likelihood to a particular response ("yes" or "no", denoted by x) at a particular time t. We assume for simplicity that each response is independent and is generated by the same set of processes, such that the likelihood of a set of N responses $\{(x_1, t_1), (x_2, t_2), \ldots, (x_N, t_N)\}$ generated by architecture M (one of the seven defined above) is just the product of the individual likelihoods, $l_M(x_1, t_1) \times l_M(x_2, t_2) \times \cdots \times l_M(x_N, t_N)$. Let $\theta_{IST,k}, \theta_{IEX,k}, \theta_{CYST,k}, \theta_{CYEX,k}, \theta_{CNST,k}, \theta_{CNEX,k}, \theta_{CB,k}$ denote the probabilities that participant k adheres to each of the seven architectures described above. Then the complete specification of the likelihood λ_k of the N_k responses produced by this participant is a mixture:

$$\lambda_k = \theta_{IST,k} \times \prod_{i=1}^{N_k} l_{IST}(x_i, t_i)$$

$$+ \theta_{IEX,k} \times \prod_{i=1}^{N_k} l_{IEX}(x_i, t_i)$$

$$+ \theta_{CYST,k} \times \prod_{i=1}^{N_k} l_{CYST}(x_i, t_i)$$

$$+ \theta_{CYEX,k} \times \prod_{i=1}^{N_k} l_{CYEX}(x_i, t_i)$$

$$+ \theta_{CNST,k} \times \prod_{i=1}^{N_k} l_{CNST}(x_i, t_i)$$

Table 3
Mean correct response time (in seconds) for each pair type (standard deviation in parentheses).

		Associative strength		
	A_H^+	A_L^+	A_H^-	A_L^-
I_H^+	1.16 (0.37)	1.20 (0.40)	1.21 (0.48)	1.27 (0.38)
I_L^+	1.19 (0.41)	1.37 (0.77)	1.12 (0.35)	1.14 (0.44)
I_H^-	1.21 (0.36)	1.18 (0.37)	1.17 (0.31)	1.22 (0.40)
I_L^-	1.24 (0.43)	1.24 (0.49)	1.18 (0.38)	1.16 (0.33)

(Item strength labels the rows)

$$+ \theta_{CNEX,k} \times \prod_{i=1}^{N_k} l_{CNEX}(x_i, t_i)$$

$$+ \theta_{CB,k} \times \prod_{i=1}^{N_k} l_{CB}(x_i, t_i). \tag{17}$$

As discussed further below, the mixture probabilities θ represent the relative probabilities that each architecture under consideration best accounts for (i.e., assigns the highest likelihood too) the data produced by participant k, similar to a set of Bayes factors, and corresponds to a hierarchical computation of relative probability, as described in Chapter 10 of Kruschke (2015).

3.3.2. Parameterization

In addition to the seven architecture mixture probabilities (only six of which are free since they must sum to one), each participant k is associated with parameters that describe the evidence accumulation rates for each response in each condition as well as parameters for their residual time and response thresholds. To reduce correlations between the quantities to be estimated, we estimate a parameter $\bar{\omega}_k$ ($\bar{\omega}_k > 0$) that reflects the sum of the response thresholds, a parameter v_k ($0 < v_k < 1$) that reflects how much of the threshold is subject to startpoint variability, a parameter w_k ($0 < w_k < 1$) that represents how much of the total threshold is given to the negative accumulator(s), and a parameter R_k ($0 < R_k < \min RT_k$) for the residual time that is bounded above by the minimum observed response time for participant k. Thus, the amount of startpoint variability is $\alpha_k^+ = \alpha_k^- = v_k \bar{\omega}_k$, the response threshold for the positive accumulator(s) is $\omega_k^+ = v_k \bar{\omega}_k + (1 - v_k)(1 - w_k)\bar{\omega}_k$, and the response threshold for the negative accumulator(s) is $\omega_k^- = v_k \bar{\omega}_k + (1 - v_k)w_k\bar{\omega}_k$. Finally, for each of the sixteen conditions, we also estimate parameters β_{ijk} and δ_{ijk} that represent the sum and difference, respectively, of evidence rates for the positive and negative accumulators in the condition with item strength i and associative strength j.

Owing to the restricted range of each parameter, we estimate $\log \bar{\omega}_k$, $\text{logit } v_k$, $\text{logit } w_k$, $\text{logit } R_k$, and $\text{logit } \theta_{m,k}$. Thus, each participant is described by a vector of 42 parameters: the four (transformed) LBA parameters; sixteen β and sixteen δ parameters; and six (transformed) architecture mixture probabilities (since the seventh is not free to vary). These are estimated hierarchically according to the model depicted in Fig. 4, which explicitly accounts for potential correlations between these 42 participant parameters.
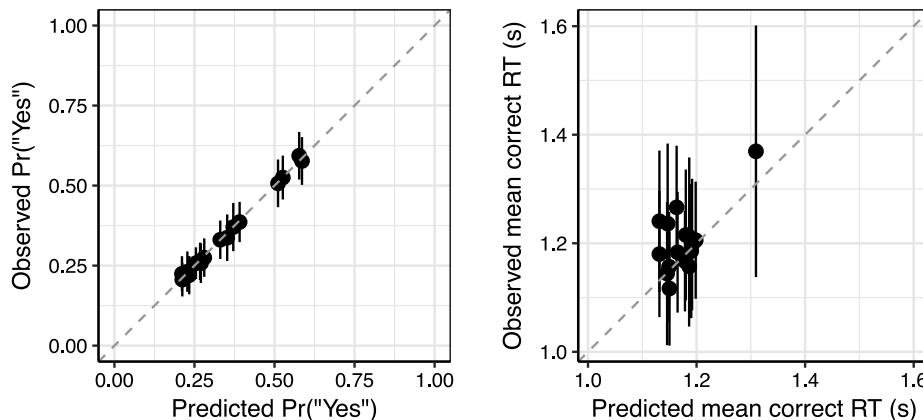
3.3.3. Maximum a posteriori estimation

We implemented this model in Stan (Carpenter, Gelman, Hoffman, Lee, Goodrich, & Betancourt, 2017), with the code and data available via the Open Science Framework at https://osf.io/m6ubq/. To begin, we found the joint set of group/individual parameters that maximized the posterior probability (maximum *a posteriori*,
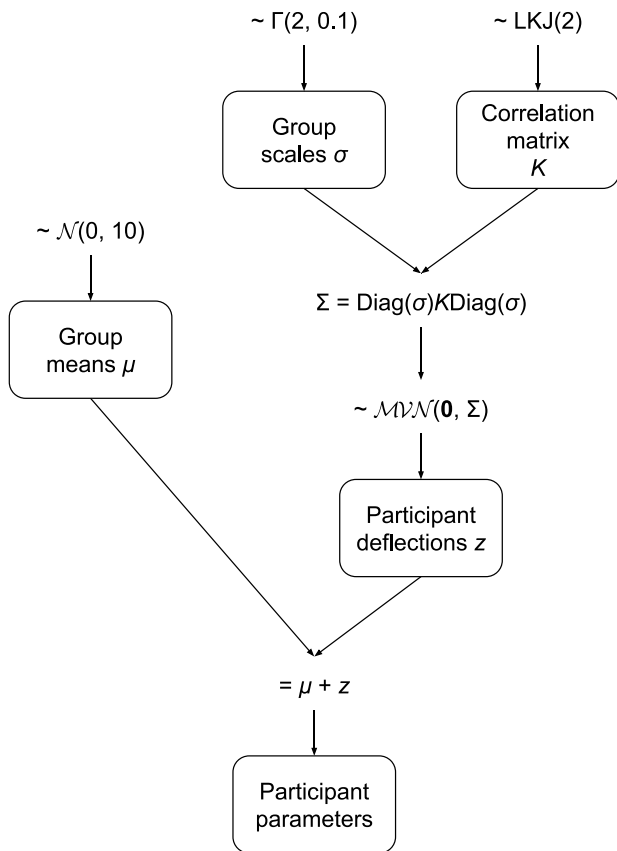
Fig. 4. Schematic depiction of the way individual parameters are estimated, where individual participant deflections are constrained to sum to zero across participants such that each participant's parameters is the sum of the group mean and their particular deflection. The "LKJ" prior over correlation matrices is as defined by Lewandowski et al. (2009). Note that the top-level prior parameters are depicted as "default" weakly informative, but may be adjusted in light of reasonable prior information.

and all of these accumulators race against one another to determine the final response. Note that the CYEX architecture helps to explain the long response times in the $I_L^+ A_L^+$ condition: An exhaustive decision rule for "yes" responses would be expected to result in longer RT's for all $I^+ A^+$ stimuli, but the other target conditions ($I_H^+ A_L^+$, $I_L^+ A_H^+$, and $I_H^+ A_H^+$) are sped up by the facilitatory influence of having at least one high-strength source.

3.3.4. Bayesian estimation

We also obtained an estimate of the Bayesian posterior distribution over these parameters (the analyses by Cox & Criss, 2017, were also fully Bayesian), again using Stan (Carpenter et al., 2017), obtaining 1000 samples each from 5 Monte Carlo chains following 1000 steps of warm-up (for 5000 total posterior samples). Across the complete set of 42 group mean parameters and $42 \times 42 = 1764$ participant deflection parameters (1806 total parameters, see Fig. 4), the natural logarithm of the $\hat{R}$ statistics was 0.016 ± 0.050 (values near zero indicate good convergence; Gelman & Rubin, 1992) and the natural logarithm of the signal-to-noise ratio (absolute value of the posterior mean divided by the Monte Carlo standard error, where larger values indicate stronger signal; Kass, Carlin, Gelman, & Neal, 1998) was 2.32 ± 1.40, indicating generally good convergence of these chains to a representative sample of the posterior distribution.

Marginalizing over the posterior samples, the resulting posterior architecture probabilities for each participant are shown in Fig. 7. These estimates reflect the uncertainty around the MAP estimates and are far less "committal". In particular, it reflects uncertainty about whether all participants are, indeed, accounted for by CYEX processing (marginalizing over participants, this architecture receives a probability of 50%) or IEX (which receives a marginal probability of 35%), that is, whether or not there are interactions that facilitate making positive responses. Although the balance of evidence favors such interactions, the value of full Bayesian estimation lies in its ability to quantify the residual uncertainty about this conclusion. As we shall see in the simulations below, these two architectures tend to be harder to tell apart than other pairs of architectures. There are two possibilities for reducing this uncertainty: First, as shown in the simulations below, additional observations can resolve the uncertainty between these two architectures. Second, we could incorporate prior information from the results of Cox and Criss (2017), which found a high posterior probability for the CYEX architecture in a similar experiment. This second option, adhering to the Bayesian dictum that "yesterday's posterior is today's prior", would shift the balance of evidence in favor of the CYEX architecture, but of course could also be misleading if this prior is inapplicable to the new experiment and set of participants.

MAP). As shown in Fig. 5, this provides a good quantitative account of the present data, an important pre-condition for interpreting these parameters. The critical parameters for present interest are the architecture mixture probabilities (θ) which are shown in Fig. 6. Consistent with the results of Cox and Criss (2017), most participants are best described by the CYEX architecture, i.e., one in which positive evidence is pooled into a single holistic accumulator but negative evidence is processed in two separate accumulators,
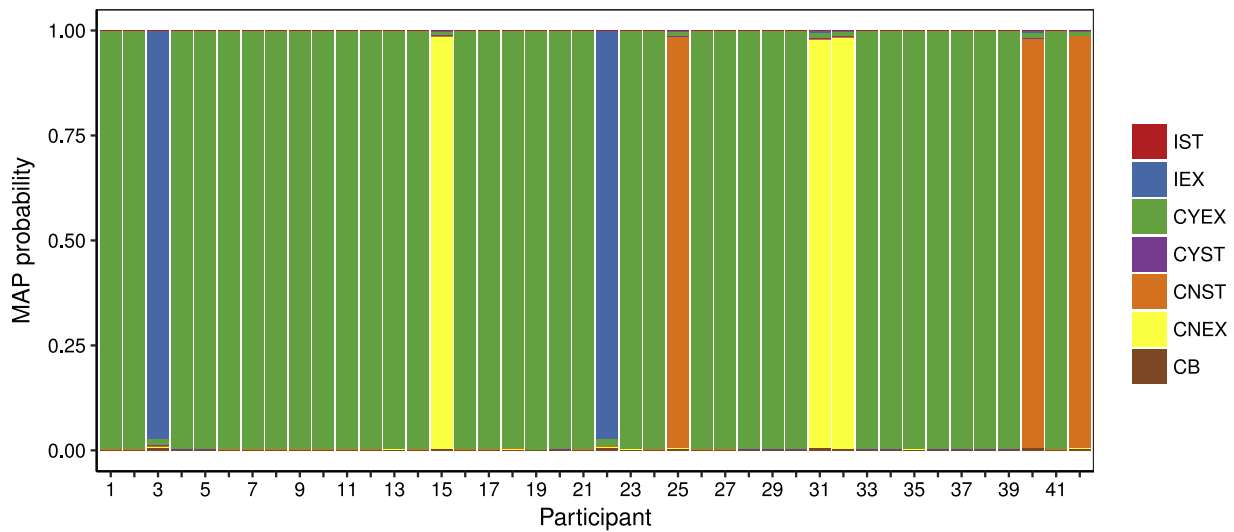

Fig. 5. Predicted mean response probability (left) and correct response time (right) versus observed mean response probability and correct response time. Bars denote 95% confidence intervals about the mean.

Fig. 6. Maximum *a posteriori* estimates of the probabilities with which each participant is best described by one of the seven architectures defined in the main text. Participants are in the same order as in Fig. 7.
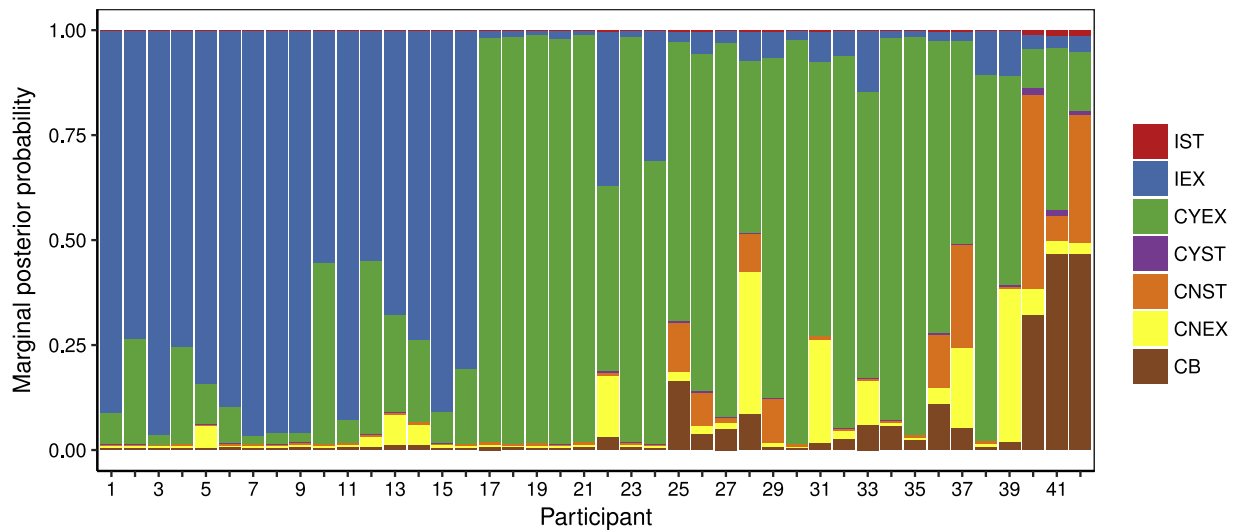
Fig. 7. Marginal posterior estimates of the probabilities with which each participant is best described by one of the seven architectures defined in the main text. Participants are in the same order as in Fig. 6.

3.3.5. Interpreting mixture probabilities

Because the mixture probabilities θ are computed with respect to the set of architectures under consideration, they should be interpreted at most in relative as opposed to absolute terms. In other words, $\theta_{M,k}$ should not be conceived as expressing the literal probability that architecture M is "true" about participant k, hence we describe these probabilities as indicating which of the architectures under consideration "best describes" a given participant. In the same way that a Bayes factor indicates support for one model *relative to* another, the probabilities output by our analyses represent the relative quality of each architecture as a description of the behavior of each participant.

3.4. Summary

In this section, we walked through a complete application of our methods for estimating the probabilities with which individuals are best described by particular processing architectures, from experimental design through to the estimated probabilities. We thus showed how to apply our methods to a double factorial experiment which would otherwise be unsuitable for traditional SFT. We also demonstrated the different properties of MAP and Bayesian estimation, with Bayesian estimation allowing for better estimates of uncertainty regarding the probabilities with which each participant engages in a certain kind of processing. In the next section, we use simulated data to explore various facets of our methods that can guide researchers who wish to apply them to their own data.

4. Simulations

We conducted a series of simulations to assess how well it is possible to correctly identify the LBA instantiations of each of the parallel processing architectures described above. We wished to address several questions pertinent to the researcher looking for a practical way of identifying such architectures:

1. Does the form of any particular architecture allow it to mimic the predictions of another and thereby get assigned a higher probability in a spurious manner?

Table 4
A double factorial design in which two information sources, A and B, can appear in either target (superscript $+$) or distractor (superscript $-$) form at two levels of strength, high (H) or low (L). Values in each cell indicate the mean value of δ for each condition in our simulations.

	A_H^+	A_L^+	A_H^-	A_L^-
B_H^+	2	1.5	-0.25	-0.5
B_L^+	1.5	1	-0.5	-0.75
B_H^-	-0.25	-0.5	-1	-1.5
B_L^-	-0.5	-0.75	-1.5	-2

2. How much data is needed per participant to achieve reasonable identification accuracy and does this depend on the amount of heterogeneity *within* a participant, that is, the degree to which each individual adheres to a single architecture?

3. What is the effect of heterogeneity *between* participants and how does this interact with partial pooling that results from using hierarchical estimation?

4.1. Methodology

4.1.1. Double factorial paradigm

All of our simulations are of performance in a double factorial experiment diagrammed in Table 4, specifically, this is a double factorial design that includes distractors and with equal rate of presentation of each stimulus (Houpt et al., 2014). We do not enforce a particular "correct" decision rule (e.g., OR vs. AND) since the decision rule is a component of the processing architecture. As a result, we only classify responses as "positive" or "negative" rather than "correct" or "incorrect".

4.1.2. Individual participant parameters

As described above, there are four parameters related to the accumulators that are separate from the inputs to those accumulators. These are boundary separation ($\bar{\omega}_k$), bias (w_k), startpoint variability (ν_k), and residual time (R_k). For each simulated participant k, these four parameters were sampled from distributions that were informed by our previous applications, to ensure that simulated behavior was within the realm of typical human performance but sufficiently variable that the simulations could address the ability to generalize across a range of parameters:

$$\log \bar{\omega}_k \sim \mathcal{N}(1.6, 0.49)$$
$$\text{logit } w_k \sim \mathcal{N}(0.024, 0.11)$$
$$\text{logit } \nu_k \sim \mathcal{N}(0.047, 1.04)$$
$$\log R_k \sim \mathcal{N}(-1.92, 0.80)$$

Specifically, the parameters of these distribution were derived from the (transformed) means and standard deviations for the corresponding parameters in the experiment analyzed by Cox and Criss (2017). For each of the $4 \times 4 = 16$ conditions, β_{ijk} was sampled independently from a normal distribution with mean of 3.80 and standard deviation of 1.28 while the δ_{ijk} parameters were sampled independently from normal distributions with standard deviation 0.71 and means specified in Table 4. Although our analysis model explicitly allows for correlations between these parameters across individuals (see Fig. 4), we did not simulate any across-participant correlations.

4.1.3. Individual and group consistency

To vary how difficult it is to identify a participant's processing architecture, we vary two kinds of consistency: Individual consistency C_I is the probability with which each trial produced by a participant adheres to the modal architecture for that simulated participant. Thus, if $C_I = 1$, a participant uses the same architecture on every trial and if $C_I = 0$, a different architecture is chosen randomly at uniform from the set of seven we focus on in this article. By varying C_I, we gain insight into how robust the analysis is to a) mis-fitting of the modal architecture; and b) violations of the assumption that there is a single best-fitting architecture for each participant. Group consistency C_G is the probability with which each *participant* has the same modal architecture, where $C_G = 1$ means every participant has the same modal architecture and $C_G = 0$ means that the modal architecture for each participant is sampled randomly at uniform from the set of seven. Varying group consistency yields insights about how hierarchical estimation can either help or mislead by allowing sharing of information across participants.

4.1.4. Selecting a modal architecture

For each simulation, one of the seven architectures is assigned to be the modal architecture and the probability that a participant shares that architecture is $C_G + (1 - C_G)\frac{1}{7}$, i.e., either it is chosen directly (with probability C_G) or at chance (with probability $(1 - C_G)\frac{1}{7}$). Once each participant's parameters and modal architecture are selected, we simulate a number of trials per participant per condition. As described above, the probability that each trial is produced by the participant's modal architecture is $C_I + (1 - C_I)\frac{1}{7}$, i.e., either the trial is consistent with the modal architecture (with probability C_I) or is chosen at chance (with probability $(1 - C_I)\frac{1}{7}$).

4.1.5. What is varied across simulations

Each simulated dataset to which the analysis is applied is characterized by

- the number of participants N_P, one of 7, 21, or 63;
- the total number of trials per condition to be divided among each participant, N_T, one of 189 or 630;
- the overall modal architecture, one of the seven we focus on in this article;
- group consistency C_G, one of 0, $\frac{1}{3}$, $\frac{2}{3}$, or 1; and
- individual consistency C_I, one of 0, $\frac{1}{3}$, $\frac{2}{3}$, or 1.

Obviously, when there is no individual consistency, it is meaningless to speak of different amounts of group consistency, so we do not bother varying the modal architecture or C_G when $C_I = 0$.

4.2. Results

We obtained MAP estimates of each simulated participant's LBA parameters and architecture probability mixture weights for each set of simulated data, using the same Stan model described above. We visualize the main results of our simulations in Fig. 8, which illustrates the average (over participants and simulations) probability estimated for each architecture in the form of a set of confusion matrices. The cell in row i, column j of each matrix depicts the estimated probability for architecture i for those participants whose true modal architecture was j, where darker cells indicate higher probability. Thus, perfect identification of each participant's generating architecture would be illustrated as a set of black squares along the diagonal. Overall quality of the fit between the MAP estimates and simulated data was quite high across all simulated datasets, with the MAP model explaining an average of 98% of variance in response probabilities (with standard deviation
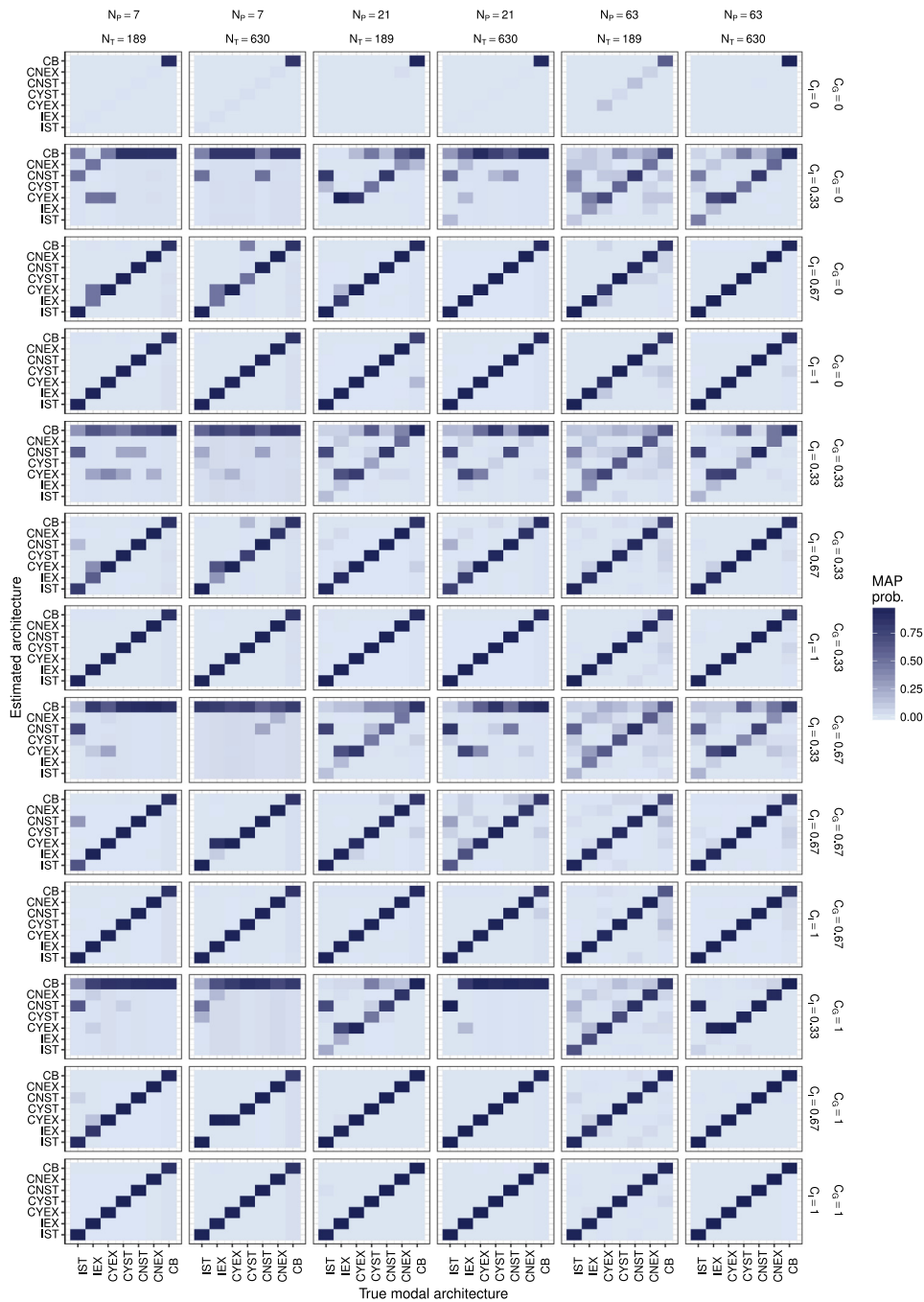
Fig. 8. Confusion matrices representing the average probabilities with which participants are identified as being best-described by a particular processing architecture. N_P: number of simulated participants; N_T: number of total trials per participant; C_G: group consistency; C_I: individual consistency. Note that when $C_I = 0$ (top row), there is no "true modal architecture", so those matrices simply depict the estimated architecture probability along their main diagonal, illustrating that the CB architecture is generally assigned the highest estimated probability. See main text for additional details.

2%), 92% of variance in mean "yes" RT (with standard deviation 4%), and 92% of variance in mean "no" RT (with standard deviation 5%) across simulations. We now discuss how these simulations address each of the questions set out at the beginning of this section in turn.

4.2.1. *A priori* preference

In our simulations with individual consistency $C_I = 0$, we have the ideal situation for testing what architecture is selected when, in fact, there is no consistent architecture that describes any one participant or group of participants. As shown in the top row of Fig. 8, the fully coactive architecture is preferred in these situations.

But this model does indeed provide an excellent quantitative account of these data, explaining an average of 98% of the variance in response probabilities (with standard deviation 2%), 94% of variance in mean "yes" RT (with standard deviation 3%), and 92% of variance in mean "no" RT (with standard deviation 3%). Recall that these simulated data represent a mixture of not only independent processing, but of various kinds of partial and complete interaction as well. Only the coactive architecture can capture the various types of information pooling that these simulated data represent, so it is sensible that when a single architecture must be selected to describe the complete set of data, the coactive one is preferred (regardless of the number of participants or observations). For

practical purposes, this indicates that a high probability assigned to the coactive architecture might indicate that, in fact, a mixture of interactive processes is actually occurring.

4.2.2. Amount of data and individual heterogeneity

Rows 2–4 of Fig. 8 address this question in a situation where group consistency $C_G = 0$, that is, there is no modal architecture across the group of participants but there is one for each individual participant. Accurate identification of individual participant architectures when $C_G = 0$ is less a function of the amount of data (or even number of participants) than it is the consistency with which each participant adheres to a particular architecture (C_I). When $C_I = \frac{1}{3}$ and the majority of trials from a participant do not adhere to the modal architecture for that participant, identification is not much better than chance and the coactive architecture continues to be preferred by default. When $C_I = \frac{2}{3}$, the majority of trials are consistent with a single modal architecture and it becomes possible to assign reasonably high probability to that architecture. Finally, when $C_I = 1$, accuracy is quite high and our method is able to correctly identify the architecture that generated each participant's data.

4.2.3. Group heterogeneity and partial pooling

Accurate identification at the individual level depends more on consistency within the individual than on the number of observations of that individual in each condition (at least up to the number allowed in our simulations). However, by using hierarchical estimation, our methods can partially overcome individual inconsistency as long as there is some degree of consistency across individuals, that is, so long as there is a modal architecture that is more prevalent than others in the sample. This is illustrated in rows 5–13 of Fig. 8, those where $C_G > 0$. As we have seen, when individual consistency is low ($C_I = \frac{1}{3}$), a coactive architecture tends to be preferred since it is the only one capable of encompassing the various forms of interaction mixed together, but this tendency is mitigated when there are enough participants ($N_P = 63$). This shows the power of partial pooling in hierarchical estimation— even if it is difficult to identify the architecture corresponding to any one individual, consistency across individuals makes it possible to pool knowledge across the sample and converge on a "best guess" as to the modal architecture for each individual. When $C_I = \frac{2}{3}$ or 1, identification of the modal architecture is nearly perfect regardless of the level of group consistency or the number of participants or trials.

4.2.4. Some confusions are easier than others

We have already seen that these simulations indicate that, when individual participants do not strictly adhere to a single modal architecture and there is not enough pull from the hierarchical structure of the data, a coactive architecture tends to be assigned much of the probability "by default" because it is the only single architecture that can capture both facilitatory and inhibitory interactions (which are mixed in by the structure of the simulations). However, it is worth noting a few other architectures that seem to be more easily confused with one another—these appear on the off-diagonals of the confusion matrices in Fig. 8. First, when data was generated by a IST architecture, the CNST architecture can still be considered more likely; second, when data was generated by a IEX architecture, the CYEX architecture is sometimes considered more likely. These two types of confusion – which only occur when $C_I < 1$ – are somewhat sensible in that they entail the same decision rules and differ only in their degree of processing interaction. This is reminiscent of the non-committal results of the Bayesian estimation example above which favored CYEX on average but still assigned reasonable probability to IEX. Our simulations show that these confusions can be resolved by additional data either by collecting more trials per participant or increasing the number of participants.

4.3. Summary

We summarize our simulation results into two key points which should prove useful to researchers using our techniques to help identify processing architectures:

1. Accurate identification is possible even with as few as 3 trials per condition per participant (when $N_P = 63$ and $N_T = 189$) so long as

    (a) most of each participant's trials (in our simulations, at least two thirds of them) were generated by the same architecture; and/or
    (b) most of the participants in the sample share the same modal architecture.

2. A coactive architecture may be identified as the most probable when behavior is in fact a mixture of both interactive and non-interactive architectures. Because this arises from the constraint that each individual be described by a single architecture, we discuss below the possibility of identifying mixtures of processing within an individual, which can help resolve this ambiguity.

3. Some architectures with the same decision rule but different kinds of interactivity can be confused with one another, but this confusion can be alleviated either by increasing the number of participants or the number of trials per participant per condition.

Unfortunately, we can only offer some rules-of-thumb to check for conditions 1a and 1b in practical settings. In general, careful inspection of data from each participant should be made prior to analysis to detect whether there are individuals that show qualitative departures from one another or within their own data, such as by looking for multimodal distributions of accuracy and/or response times both within each participant and across participants. But, of course, if data really are as sparse as 3 trials per condition per participant, it may not be possible to check for these features, in which case the analysis must be taken with a grain of salt. In all cases, but especially those with sparse data, full Bayesian estimation is advisable in order to get a handle on the degree of uncertainty in the results, which can serve as a guide to how much "salt" may be needed.

5. Discussion

We have described, applied, and investigated a method for using hierarchical estimation and systems of LBA's to help researchers identify the type of processing architecture employed by participants in a double factorial experiment. We have provided code in Stan (Carpenter et al., 2017) that researchers can use for their own purposes. These methods cannot supplant the full power of SFT, but they can help to supplement it, particularly in situations where it is not possible to apply traditional SFT methods. These situations include cases with high error, relatively few trials per participant, and lack of selective influence. In addition, these methods provide a set of probabilities for each participant that directly assess the degree of statistical support for different processing architectures.

5.1. Extensions to other architectures

Although we explored seven varieties of parallel architectures that run the gamut from complete independence (IST and IEX) to complete coactivity (CB), many other architectures may be specified using systems of LBA's. Although the distributions used in the original formulation of the LBA have proven sufficient for the applications we have explored so far, as mentioned in the

model description it is possible to formulate accumulators using different distributions of evidence accumulation rates and/or start-point variability (e.g., Heathcote & Love, 2012; Terry et al., 2015). Another straightforward extension would be to allow for more than two sets of accumulators, as might happen in an experiment where three or more sources of information were manipulated (e.g., a triple, quadruple, etc. factorial design). Many measures from SFT and other approaches have been extended to these situations (Yang, Fifić, & Townsend, 2014; Zhang & Dzhafarov, 2015).

5.1.1. Estimating the number of information sources

There may be situations in which an experimenter seeks to manipulate sources of information that could be further decomposed, in the way that color might be decomposed into separate channels for hue and saturation. In such a setting, it could be possible, in principle, to *estimate* the number of such sources of information involved in the same way that we estimate the probability with which each participant adheres to a particular processing architecture. This comes with a caveat: Just as we saw a coactive architecture tended to be assigned a high probability even when behavior was actually produced by a mixture of interactive and independent architectures, allowing too much freedom in specifying different architectures can lead to a situation in which an architecture is preferred simply due to its flexibility. In other words, because our methods involve *parametric* models whose predictions are constrained in form, any discrepancy between the model and behavior can lead to mis-identification. Thus, while we remark on this possibility, we caution researchers that this route can be a dangerous one to travel.

5.1.2. Serial and serial-like architectures

An obvious shortcoming of the present set of analyses is that we did not consider serial models. While systems of LBA's lend themselves naturally to specifying parallel models, the convolution involved in serial likelihoods means that numerical methods must be used in that case. In addition, as is clear from Fig. 3, traditional signatures of serial processing like a flat SIC can be produced by interactive parallel architectures, increasing the well-known potential for mimicry between these model classes (Townsend, 1976; Townsend & Thomas, 1994). Finally, it is not clear how to specify interactions for serial models, except perhaps as a kind of feed-forward interaction where the output of an earlier stage is fed into a subsequent stage in the manner of a "cascade" or "continuous-flow" system (Ashby, 1982; McClelland, 1979; Townsend & Fikes, 1995).

One way to capture some aspect of serial processing, alluded to in the model specifications above, would be to allow different sets of LBA's to have different residual times, implying that they are staggered in time (Hendrickson et al.). As we noted above, however, this introduces an identification problem in that it may not be clear *which* accumulators are earlier or later, but this could be solved either by a theoretical assumption (enforcing a relative processing order) or by experimental manipulation. For example, one might withhold presentation of one information source until after the other to enforce a particular order, or including conditions that require speeded responding may allow a researcher to infer the relative processing order by examining whether speeded responses indicate that only one source is processed prior to responding (e.g., Donkin & Little, 2014).

5.2. Other experimental paradigms

Throughout this article, we have focused on a double factorial experimental design, particularly one with four levels of strength for each factor. This is largely due to the fact that our methods are designed as a supplement to SFT, and this is the paradigm for which most of its measures are designed. In principle, however, our analyses can be applied to data from partial-factorial designs (i.e., those with missing cells) as well as designs with more or fewer than the three or four levels per factor typically used in SFT. The reason for this flexibility is because our methods do not depend on having particular stochastic dominance relationships that are prerequisites for some SFT measures (like the SIC) and which are the intended outcome of the double-factorial manipulation. But while our methods do not depend on this qualitative feature of the data, our methods do require that strength manipulations produce a sufficiently wide range of effects to allow for different architectures to make different predictions.

In addition to double-factorial designs, we have focused on tasks that require discrimination between two types of stimuli, i.e., those that require a positive ("yes") response versus those that require a negative ("no") response. A detection paradigm (as discussed in Donkin & Little, 2014) may only entail a positive response upon detection of any kind of target stimulus, with no negative responses at all. One way to apply the present approach to such a paradigm would be to simply eliminate the accumulators associated with a negative response, such that only three architectures remain in the two-source case (IST, IEX, and coactive), but this would require ensuring that accumulation rates in no-target trials were extremely low, otherwise a response would be made on every trial. An alternative would allow for an implicit set of "no response" accumulators that would allow for a response to be withheld if they won the race. While it seems cognitively plausible that there would be a "no response" accumulator for each information source – like the "no" response accumulators in the models above – in practice it might be difficult to infer whether those accumulators interact since, by definition, their finishing times are not directly observed (cf. Logan, Van Zandt, & Verbruggen, 2014).

5.3. Within-participant process mixtures

Our methods have shared the assumption of traditional SFT that each individual's performance in a given task can be characterized as having been generated by a single processing architecture. But we noted in the Introduction the possibility that individuals represent a mixture of processes and saw in our simulations that, in the event that a participant involves a mixture between interactive and independent processing, the interactive architecture tends to be preferred. But this preference is conditional on the assumption that each individual adheres to a single architecture throughout. Just as we estimate the probability with which each individual is best described by any one architecture, it is entirely possible to extend this idea to estimate the probability with which *each trial* from each participant is best described by a particular architecture, as in multinomial process tree models (e.g., Klauer & Kellen, 2018; Riefer & Batchelder, 1988). The downside of this approach is that the additional variability associated with individual-level mixtures propagates toward all estimates, requiring more observations per individual to attain a reasonable degree of certainty.

5.4. Bridging the gap between large- and small-N

The recent and ongoing "replication crisis" in the psychological sciences has often been attributed, in part, to the use of samples that are too small to reliably detect the effects of interest (e.g., Open Science Collaboration, 2015). However, many of the most robust results in psychology, particularly in psychophysics, have come from experiments with very few or even a single participant (Smith & Little, 2018). What distinguishes these robust small-*N* results from fickle large-*N* results is that the small-*N* experiments began from a set of well-specified *theories* about how different perceptual and cognitive processes yield particular patterns of behavior.

Indeed, this is the chief power of SFT, that there is a strong and clear relationship between theory, experimental design, and data. While our use of parametric models like the LBA restricts the kinds of theories that can be tested using our methods, it enables estimation to be done in a *hierarchical* manner. Hierarchical modeling bridges the gap between large- and small-*N* by respecting individual differences while letting individual estimates be informed by the rest of the group.

5.5. Utility for online experiments

Hierarchical estimation, coupled with strong theory, is why successful identification using our methods is possible even with few trials or few participants. This fact becomes especially important when considering the growth of online data collection in many fields of psychology. While this enables larger and more diverse samples than are possible via traditional laboratory testing, often online data collection is constrained to collect relatively few trials per participant (to avoid attrition) and evinces greater variability between participants due to their different environments and backgrounds. Hierarchical estimation and strong theory – of which our methods are a prime example – can enable researchers to extract robust conclusions, at least at the group level, from such potentially variable data.

5.6. SFT for all?

The primary motivation behind the work discussed in this article was to allow researchers working in domains that are normally unsuited to SFT – domains with high error rates, relatively few observations per participant, and no guarantee of selective influence – to apply *some* of the logic of systems factorial technology. While the methods discussed in this article are not perfect and lack the generality of traditional SFT, we believe these are reasonable prices to pay in exchange for the ability to draw principled conclusions about the structure of psychological processes in these domains. It is, however, critical for researchers who apply our methods to understand that they are only a guide and not a final answer: they assume that the LBA provides a good quantitative account of their data, that only parallel architectures are worth considering, and that the various kinds of interactions that can occur during processing can be captured by channel summation (and these methods can be fooled if participants in fact represent a mixture of independent and interactive processes). The results of our methods should always be checked against other methods (other analyses or simulations) if possible, as in our own prior work (Cox & Criss, 2017). Still, we believe that the methods we describe, limited as they are, can help advance psychological science more generally by bringing the power of SFT – rooted in a strong connection between theory and data – to regions it would otherwise be unable to explore.

Appendix. Experimental methods

This Appendix presents the details of the experiment the data from which were analyzed for example purposes in the main text.

A.1. Participants

57 undergraduate students at Syracuse University participated in this study in exchange for course credit, in accord with local Institutional Review Board policy.

A.2. Materials

The stimulus materials were identical to those used by Cox and Criss (2017), namely, images of indoor and outdoor scenes used by Goh, Siong, Park, Gutchess, Hebrank, and Chee (2004) and Konkle, Brady, Alvarez, and Oliva (2010). 512 images were

Table A.1
Design of study lists that separately manipulate item strength (total number of item presentations) and associative strength (number of whole pair repetitions) with numbers referring to a specific image half.

Paired items	Number of isolated item repetitions	Number of whole pair repetitions	Total number of item presentations
1, 2	2	3	5
3, 4	4	1	5
5, 6	2	3	5
7, 8	2	3	5
9, 10	4	1	5
11, 12	4	1	5
13, 14	0	3	3
15, 16	2	1	3
17, 18	0	3	3
19, 20	0	3	3
21, 22	2	1	3
23, 24	2	1	3
25, 26	2	3	5
27, 28	4	1	5
29, 30	0	3	3
31, 32	2	1	3

Table A.2
Implementation of a double factorial design that separately manipulates item strength (I) and associative strength (A). Numbers refer to the items shown at study, with apostrophes denoting the half of each image item that was *not* shown at study.

		Associative strength			
		A_H^+	A_L^+	A_H^-	A_L^-
Item strength	I_H^+	1, 2	3, 4	5, 8	9, 12
	I_L^+	13, 14	15, 16	17, 20	21, 24
	I_H^-	25', 26'	27', 28'	7', 6'	11', 10'
	I_L^-	29', 30'	31', 32'	19', 18'	23', 22'

selected from these sets to serve as a pool of stimuli, according to the similarity criteria described by Cox and Criss (2017).

A.3. Design

We use the same general double factorial design as in Cox and Criss (2017), just with a different way of manipulating item and associative strength. Within the study phase (see Table A.1), high associative strength pairs were shown three times over the course of the study list while low associative strength pairs were shown only once. Item strength was determined by the *total* number of appearances of the item during study, including appearances that occur as part of a pair. To increase this number and thereby increase the strength of item information in memory, we presented each item of a pair in isolation in the study list, with the number of such repetitions depending on the number of times the whole pair containing the item had been presented, as given in Table A.1.

Test lists consisted of 16 pairs of items as described in Table A.2. The design of the test list matches a double factorial paradigm with distractors (see Houpt et al., 2016), such that there are four types of pair: I^+A^+ pairs in which both the item (I) and associative (A) information in the test pair match what had been studied (i.e., they are target items and should receive a "yes" response). I^+A^- pairs are formed by rearranging the left and right items between two studied pairs with the same level of item and associative strength; the result is a test pair in which the item information matches what had been studied, but the associative information does not. I^-A^+ pairs are formed by taking a studied pair and replacing the each image in the pair with the *unseen half* of that image, with the

result that the item information no longer matches what had been studied but the associative information does. Finally, I^-A^- pairs are formed by applying both of these operations, first rearranging the left and right items of a studied pair then replacing those items with the unstudied half of each image.

A.4. Procedure

After providing informed consent, each participant engaged in between 5 and 11 blocks (depending on how much they could complete in the allotted time) consisting of a study and test phase as described above. Each block began with a screen that instructed participants that they would be shown a sequence of image pairs and that they should try to remember not only the images in each pair, but which images appeared together. In addition, participants were informed that they would only need to remember the most recent set of pairs, since items would never appear in more than one block (which was ensured by the randomization procedure). After a minimum of 15 s, participants could proceed to the study phase in which each pair was presented for 1.5 s with a 0.5 s blank interval in between. The two image pairs were horizontally separated on the screen by a small blank space and had a slight random vertical offset from one another, with the aim of encouraging participants to process the images as separate entities (rather than as two parts of the same image) as well as minimizing visual masking between trials. After all study stimuli were presented, participants were shown an instruction screen that told them to respond "yes" (by hitting either the F or J key, randomly determined for each participant) only if a test pair exactly matched a pair they had just studied, otherwise they should respond "no" as rapidly and accurately as possible. After a minimum of 15 s, participants could proceed to the test phase. Each test trial began with a fixation cross in the center of the screen for 0.5 s, followed by a 0.5 s blank screen, then the presentation of the test pair (which was, as at study, horizontally centered but with a small random vertical offset between the two images). After responding, participants received feedback about whether they were correct or incorrect for a minimum of 1 s if they were correct or 1.5 s if they were incorrect. In addition, if their response took longer than 4 s from the onset of the test pair, feedback included an instruction to try to respond more quickly; if their response took less than 0.3 s, feedback included an instruction to try to take more time to respond and feedback was shown for an additional 3 s. A random interval between 0.25 and 0.75 s preceded the onset of the next test trial.

A.5. Exclusion criteria

Prior to analysis, we excluded participants who gave more positive responses to lure items than target items as well as any participant who did not produce at least one correct response in each cell of the double factorial design, under the assumption that such participants were not following instructions. We also excluded any trial with response time less than 200 ms (since these could not reflect processing of the test pair) or greater than 10 s (since these were likely to be contaminated by other processes).

References

Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review*, 89(5), 599–607.

Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive Psychology*, 57, 153–178.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic–cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432–459.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., & Betancourt, A. (2017). Stan: a probabilistic programming language. *Journal of Statistical Software*, 76(1).

Cox, G. E., & Criss, A. H. (2017). Parallel interactive retrieval of item and associative information from event memory. *Cognitive Psychology*, 97, 31–61.

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review*, 124(6), 795–860.

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: rethinking the scaling problem. *Psychonomic Bulletin & Review*, 16(6), 1129–1135.

Donkin, C., Brown, S., & Heathcote, E.-J. (2011). Diffusion versus linear ballistic accumulation: different models but the same conclusions about psychological processes?. *Psychonomic Bulletin & Review*, 55, 140–151.

Donkin, C., & Little, J. W. (2014). Assessing the speed-accuracy trade-off effect on the capacity of information processing. *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 1183–1202.

Donkin, C., & Nosofsky, R. M. (2012). The structure of short-term memory scanning: an investigation using response time distribution models. *Psychonomic Bulletin & Review*, 19, 363–394.

Edwards, W. (1965). Optimal strategies for seeking information: models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, 2, 312–329.

Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010a). Converging measures of workload capacity. *Psychonomic Bulletin & Review*, 17(6), 763–771.

Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010b). Converging measures of workload capacity. *Psychonomic Bulletin & Review*, 17(6), 763–771.

Eidels, A., Houpt, J. W., Altieri, N., Pei, L., & Townsend, J. T. (2011). Nice guys finish fast and bad guys finish last: facilitatory vs. inhibitory interaction in parallel systems. *Journal of Mathematical Psychology*, 55, 176–190.

Fifić, M., Nosofsky, R. M., & Townsend, J. T. (2008). Information-processing architectures in multidimensional classification: a validation test of the systems factorial technology. *Journal of Experimental Psychology: Human Perception and Performance*, 34(2), 356–375.

Gelman, A., & Rubin, B. D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–511.

Goh, J. O. S., Siong, S. C., Park, D., Gutchess, A., Hebrank, A., & Chee, M. W. L. (2004). Cortical areas involved in object, background, and object-background processing revealed with functional magnetic resonance adaptation. *The Journal of Neuroscience*, 24(45), 10223–10228.

Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology*, 3(292), 1–19.

Hendrickson, A. T., Navarro, D. J., & Donkin, C. (n.d.). Evidence accumulation in same-different judgments: Integrating featural similarity with structural knowledge using a linear ballistic accumulator (Unpublished manuscript).

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., & Townsend, J. T. (2014). Systems factorial technology with R. *Behavior Research Methods*, 46, 307–330.

Houpt, J. W., & Fifić, M. (2017). Approach to distinguishing serial and parallel processing a hierarchical bayesian approach to distinguishing serial and parallel processing. *Journal of Mathematical Psychology*, 79, 13–22.

Houpt, J. W., Heathcote, A., & Eidels, A. (2017). Analyses of cognitive architecture bayesian analyses of cognitive architecture. *Psychological Methods*, 22(2), 288–303.

Houpt, J. W., MacEachern, S. N., Peruggia, M., Townsend, J. T., & Van Zandt, T. (2016). Approaches to systems factorial technology semiparametric bayesian approaches to systems factorial technology. *Journal of Mathematical Psychology*, 75, 68–85.

Houpt, J. W., & Townsend, J. T. (2011). An extension of sic predictions to the wiener coactive model. *Journal of Mathematical Psychology*, 55, 267–270.

Kass, R. E., Carlin, B. P., Gelman, A., & Neal, R. M. (1998). Markov chain monte carlo in practice: a roundtable discussion. *The American Statistician*, 52(2), 93–100.

Klauer, K. C., & Kellen, D. (2018). Rt-mpts : process models for response-time distributions based on multinomial processing trees with applications to recognition memory. *Journal of Mathematical Psychology*, 82, 111–130.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, 139(3), 558–578.

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2 edn.). London: Academic Press.

Laming, D. R. J. (1968). *Information theory of choice-reaction times Information theory of choice-reaction times*. London: Academic Press.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100, 1989–2001.

Link, S. W. (1975). The relative judgment theory of two choice response time. *Journal of Mathematical Psychology*, 12, 114–135.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40, 77–105.

Little, D. R., Nosofsky, R. M., & Denton, S. E. (2011). Response-time tests of logical-rule models of categorization. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 37(1), 1–27.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95(4), 492–527.

Logan, G. D., Van Zandt, T., & Verbruggen, E.-J. (2014). On the ability to inhibit thought and action: general and special theories of an act of control. *Psychological Review*, *121*(1), 66–95.

McClelland, J. L. (1979). On the time relations of mental processes: an examination of systems of processes in cascade. *Psychological Review*, *86*(4), 287–330.

Nosofsky, R. M., Cao, R., Cox, G. E., & Shiffrin, R. M. (2014). Familiarity and categorization processes in memory search. *Cognitive Psychology*, *75*, 97–129.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349*, 6251.

Purcell, B. A., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2012). From salience to saccades: multiple-alternative gated stochastic accumulator model of visual search. *The Journal of Neuroscience*, *32*(10), 3433–3446.

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*(5), 1226–1243.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318–339.

Rouder, J. N. (1996). Premature sampling in random walks. *Journal of Mathematical Psychology*, *40*, 287–296.

Schneider, W., & Shiffrin, R. M. (1977). Detection, search, and attention controlled and automatic human information processing: i detection, search, and attention. *Psychological Review*, *84*(1), 1–66.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: ii perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*(2), 127–190.

Smith, P. L., & Little, D. R. (2018). Small is beautiful: in defense of the small-*n* design. *Psychonomic Bulletin and Review*, http://dx.doi.org/10.3758/s13423-018-1451-8.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice decision. *Journal of Mathematical Psychology*, *32*, 135–168.

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: from independence to competition. *Psychological Review*, *120*(1), 1–38.

Terry, A., Marley, A. A. J., Barnwal, A., Wagenmakers, E.-J., Heathcote, A., & Brown, S. D. (2015). Generalising the drift rate distribution for linear ballistic accumulators. *Journal of Mathematical Psychology*, *68–69*, 49–58.

Thiele, J. E., Haaf, J. M., & Rouder, J. N. (2017). Is there variation across individuals in processing? bayesian analysis for systems factorial technology. *Journal of Mathematical Psychology*, *81*, 40–54.

Townsend, J. T. (1976). Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology*, *14*, 219–238.

Townsend, J. T., & Altieri, N. (2012). An accuracy-response time capacity assessment function that measures performance against standard parallel predictions. *Psychologial Review*, *119*(3), 500–516.

Townsend, J. T., & Fikes, T. (1995). A beginning quantitative taxonomy of cognitive activation systems and application to continuous flow processes. *Tech. Rep. No. 131*, Bloomington: IN Indiana University Cognitive Science Program.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: an investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, *39*, 321–359.

Townsend, J. T., & Thomas, R. D. (1994). Stochastic dependencies in parallel and serial models: effects on systems factorial interactions. *Journal of Mathematical Psychology*, *38*, 1–34.

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: new capacity measures and predictions for a response time inequality series. *Psychologial Review*, *111*(4), 1003–1035.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, *121*(2), 179–205.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*(1), 37–58.

Yang, H., Fifić, M., & Townsend, J. T. (2014). Survivor interaction contrast wiggle predictions of parallel and serial models for an arbitrary number of processes. *Journal of Mathematical Psychology*, *58*, 21–32.

Zandbelt, B., Purcell, B. A., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2014). Response times from ensembles of accumulators. *Proceedings of the National Academy of Sciences*, *111*(7), 2848–2853.

Zhang, R., & Dzhafarov, E. N. (2015). Noncontextuality with marginal selectivity in reconstructing mental architectures. *Frontiers in Psychology*, *6*(735), 1–12.