

# **Disentangling decision models – from independence to competition**

Andrei R. Teodorescu                      and                      Marius Usher

Tel-Aviv University

Corresponding author:

Prof. Marius Usher

Department of Psychology, Tel-Aviv University, Ramat-Aviv.

E-mail: [marius@post.tau.ac.il](mailto:marius@post.tau.ac.il)

Acknowledgments:

We wish to thank James L. McClelland for very stimulating input and criticism on the design of these experiments and to Konstantinos Tsetsos & Rani Moran for computational advice and constructive discussions.

This research was partially supported by the Air Force Research Laboratory (FA9550-07-1-0537).

Simulation codes can be found at:

<https://www.dropbox.com/sh/rqm5vrtr1rnzoov/zdp8THRfg0>

## **Abstract**

A multitude of models have been proposed to account for the neural mechanism of value integration and decision making in speeded decision tasks. While most of these models account for existing data, they largely disagree on a fundamental characteristic of the choice mechanism: independent vs. different types of competitive processing. Five models, an independent race model, two types of input competition models (normalized race and Feed-Forward Inhibition [FFI]) and two types of response competition models (Max Minus Next [MMN] diffusion and Leaky Competing Accumulators [LCA]) were compared in three combined computational and experimental studies. In each study, difficulty was manipulated in a way that produced qualitatively distinct predictions from the different classes of models. When parameters were constrained by the experimental conditions to avoid mimicking, simulations demonstrated that independent models predict speedups in RT with increased difficulty, while response competition models predict the opposite. Predictions of input-competition models vary between specific models and experimental conditions. Taken together, the combined computational and empirical findings provide support for the notion that decisional processes are intrinsically competitive and that this competition is likely to kick in at a late (response) rather than early (input) processing stage.

## Disentangling decision models – from independence to competition

Decision-making is paramount in daily life, ranging from fast perceptual choices critical to survival, such as deciding whether the signal on a radar screen indicates an enemy missile or a friendly plane, to multi-dimensional, goal driven and effort-consuming decisions, such as deciding on the guilt of a defendant in a legal case. In such situations, the decision-maker is presented with samples of information and is required to decide which alternative to choose. Samples may correspond to perceptual information in the form of physical intensity values as in a sequence from a noisy visual stream (Gold & Shadlen, 2007; Glimcher, 2003), values associated with qualities of different alternatives such as consumer products, risky gambles, career choices and flat-mates (Hertwig, Barron, Weber, & Erev, 2004; Tsetsos, Usher, & Chater, 2010; Dijksterhuis, Bos, Nordgren, & van Baaren, 2006), or pieces of evidence in a legal case. The problem facing the decision maker is *how* to combine and weigh these samples towards a decision while at the same time limiting the amount of sampled information to determine *when* to terminate the process and execute a response (for further discussion, see Gold & Shadlen, 2007; Ratcliff & Smith, 2004).

The mechanism that enables humans to make such decisions has been studied extensively over the last few decades both in psychology (Bogacz, Brown, Moehlis, Holmes & Cohen, 2006; Bogacz, Usher, Zhang & McClelland, 2007; Brown & Heathcote, 2008; Laming, 1968; Link & Heath, 1975; Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004; Ratcliff & McKoon, 2008; Van Ravenzwaaij, Mulder, Tuerlinckx & Wagenmakers, in press; Stone, 1960; Usher & McClelland, 2001; Vickers, 1970) and in neuroscience (Albantakis, & Deco, 2009; Donner, Siegel, Fries, & Engel, 2009; Gold & Shadlen, 2007; Hanes & Schall, 1996; Mulder, Wagenmakers, Ratcliff, Boekel & Forstmann, in press; Purcell, Heitz, Cohen, Schall, Logan & Palmeri, 2010; Rorie, Gao, McClelland & Newsome, 2010; Rorie & Newsome, 2005; Wang, 2008), converging on the idea that multiple samples of information are translated to a value-scale (see Luce, 1959, and Anderson, N. H., 1971, for earlier precursors to the important role of value integration in decision-making, and Glimcher (2003) for a more recent discussion) and are accumulated, over time, towards a decision criterion. This *sequential sampling* principle allows the decision making mechanism to average external and internal noise over time while accounting for both response time (RT; e.g., Ratcliff & McKoon, 2008) and accuracy.

The sequential sampling principle is now considered a general decision mechanism that deals with the integration of values that fluctuate over time (Gold & Shadlen, 2001, 2002; Kiani, Hanks & Shadlen, 2008; Roe, Busmeyer & Townsend, 2001; Rorie & Newsome, 2005; Sugrue, Corrado & Newsome, 2005; Usher & McClelland, 2004). In this spirit, a number of paradigms have been developed, which use noisy perceptual stimuli (such as randomly moving dots, among which a small fraction moves coherently; Shadlen & Newsome, 1996) as a proxy to more *general* information. While certain perceptual decisions have a quite limited temporal range

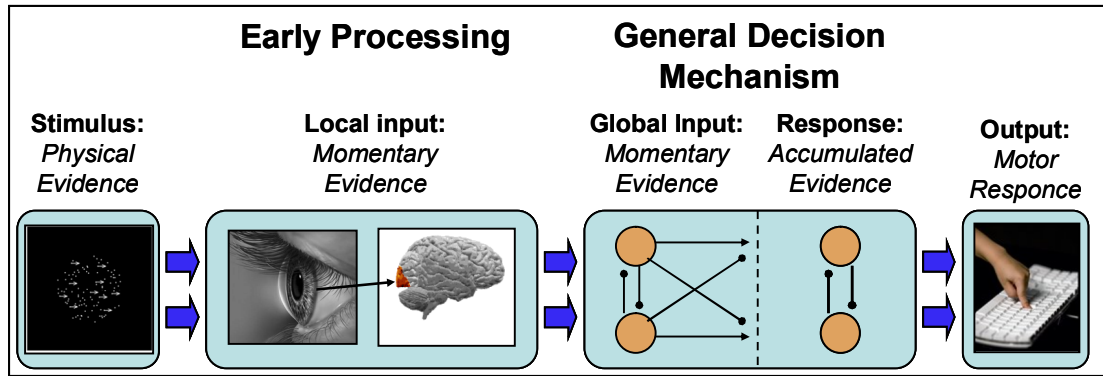
of integration (200-300ms) that is subject to stimulus specific properties, indicating a more idiosyncratic mechanism (Uchida, Kepecs & Mainen, 2006), other tasks implementing noisy evidence (e.g., the random-dots) show much longer integration times of up to a few seconds (Burr & Santoro, 2001; see also review in, Uchida, Kepecs & Mainen, 2006). The focus of our paper is to scrutinize the general decision mechanism responsible for such value integration and evaluation tasks by contrasting a variety of decision-making models.

Indeed, a number of different mathematical models, most of which implement sequential sampling as their underlying principle, have been proposed to account for general decision behavior. Such models range from *race* or *accumulator* models (Brown & Heathcote, 2008; Smith & Vickers, 1988; VanZandt, Colonius & Proctor, 2000; Vickers, 1970), to *drift-diffusion* models (Link & Heath, 1975; Ratcliff, 1978; Ratcliff & Rouder, 1998; Stone, 1960), and more recent neurocomputational models introduced in an attempt to bridge the gap between previous mathematical models and the growing understanding of brain functions. A few such examples are the neural implementation of the drift diffusion model (Mazurek, Roitman, Ditterich & Shadlen, 2003; Niwa & Ditterich, 2008; Ratcliff, Hasegawa, Hasegawa, Smith & Segraves, 2007), the *Leaky Competing Accumulator* model (LCA; Usher & McClelland, 2001) and the *attractor* model (Albantakis, & Deco, 2009; Wang, 2002, 2008).

The large corpus of currently available choice models can be classified according to a number of important principles that reflect distinct underlying mechanisms such as *competitive* vs. *independent* processing, relative vs. absolute decision criteria and perfect vs. leaky integration of evidence. Despite these clear processing differences, models of each type exist which are able to fit reasonably well most empirical data to date. Consequently, so far, only small quantitative differences in the goodness of fit have been found between models (Brown & Heathcote, 2008; Ratcliff & Smith, 2004; Usher & McClelland, 2001, VanZandt, Colonius, & Proctor, 2000) and the various models can mimic each other to a considerable degree (Donkin, Brown, Heathcote & Wagenmakers, 2011; Townsend, 1990; VanZandt & Ratcliff, 1995). Such model mimicry is a serious problem since it hinders the use of models as instruments for testing theories. Therefore, disentangling them from one another requires a more comprehensive theoretical and methodological approach to the problems of model testing, model comparison and model classification. We start with a general taxonomy of decision models followed by an outline of our model testing approach before presenting a set of experimental and computational studies set up to distinguish between classes of decision models.

### **Model Taxonomy**

Our main goal in this paper is to make clear distinctions between different types of *independent* and *competitive* models. The first step in doing so is outlining a detailed theoretical classification of perceptual choice models, which divides them into classes based on the locus of competition (or lack of it). We will begin by drawing a clear line between purely independent and competitive models. We will then proceed to distinguish four different types of competitive interactions according to their locus of action: (1) Stimulus competition; (2) Local input competition; (3) Global input competition; (4) Response competition (see Figure 1); and point out how each type of competition manifests in existing models.



**Figure 1:** The flow of visual information from the physical stimulus up to the execution of a decision. The figure illustrates the different loci where competition can take place. Information begins its journey in the stimulus as physical energy. If the energy matching one alternative is inversely dependent upon energy matching the other then it can be said that *stimulus competition* is present. Next, before entering the decision mechanism, that energy is encoded, by early processing stages, into values representing different features of the stimulus. Here, interactions can take place between values originating from spatially and temporally adjacent stimuli resulting in *local input competition*. If a decision regarding the stimulus is required, the task relevant values then enter the decision mechanism. Decision competition can take place at two loci: global input & response. *Global input competition* corresponds to interactions between early, momentary values which do not, anymore, depend on stimulus properties such as spatial vicinity. Last, interactions whose strength is proportional to the amount of accumulated, rather than momentary, values can occur resulting in *response competition*.

### The Independent Race Model

Race models have been proposed over the years to describe a variety of different processes and regularities (Brown & Heathcote, 2008; Eidels, Townsend & Algom, 2009; Kyllingsbæk, Marcussen & Bundesen, 2011; La Berge, 1962; Logan, Cowan & Davis, 1984; Mordkoff & Yantis, 1991; Morton 1964; Pike 1973; Smith & Van Zandt, 2002; Townsend, 1976; Townsend & Ashby, 1983; Usher, Olami & McClelland, 2002; VanZandt et al., 2000; Vickers, 1970). Due to their simplicity, race models can often provide analytical solutions for RT distributions and accuracies (Brown & Heathcote, 2008), thus making them appealing for both theoretical and practical reasons (VanZandt et al., 2000).

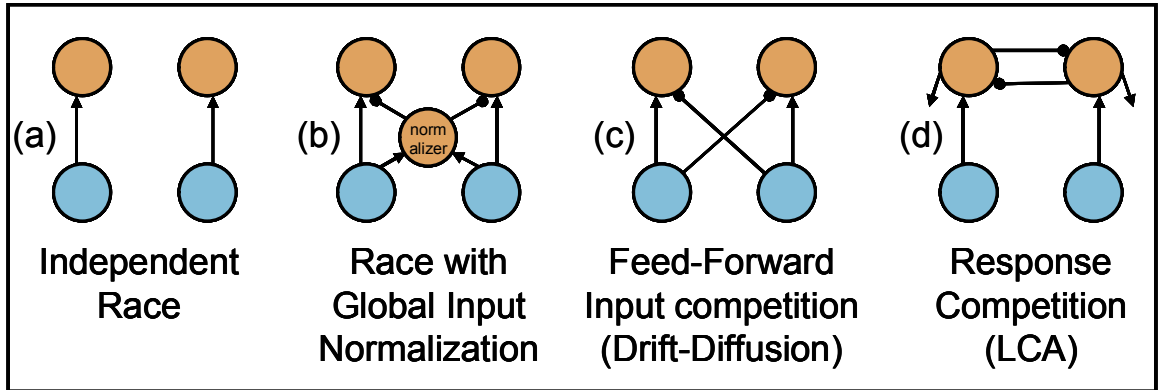
A distinguishing feature of race models is that they allocate a separate accumulator for each alternative and that these accumulators integrate inputs (samples of values) monotonically. While this is often associated with *independent* accumulation of information, not all race models are purely independent and they vary with respect to the manner in which the accumulators are correlated. For example, in Vickers' accumulator model (Vickers, 1970) the accumulators are indeed separate but at each time step only one accumulator, the one most supported by the momentary sample, receives an increment. This selective input structure introduces a correlation between the accumulators, so that the stronger the input for one alternative the slower the accumulation rate of the other alternative will be.

Other race models, however, do assume the principle of independent accumulation of information, which is characterized by the lack of interaction between separate channels at any level of processing. This approach is particularly attractive when the number of alternatives is large, since accumulating on the basis of

pairwise comparisons (as in the binary accumulator model) is unparsimonious (see section: 'Competitive models – input competition' below). Indeed, *independent* race models have been proposed in a variety of tasks that require selecting one out of  $n$ -alternatives, such as multiple-alternative perceptual choice and Hick's law (Brown & Heathcote, 2008; Pike, 1973; Usher et al., 2002; VanZandt et al., 2000), word identification (Morton, 1969; 1970), Stroop (Eidels, 2012; Eidels, Townsend & Algom, 2009), stop signal and action control (Logan & Cowan, 1984; Logan, Cowan & Davis, 1984), and saccade generation (Ludwig, Gilchrist, McSorley, & Baddeley, 2005). Figure 2a depicts a skeletal architecture for a purely independent model where no line (representing a connection) crosses the gap between channels that integrate information about different alternatives. This independent race model is described by the following equation:

$$\Delta X_i = I_i + \xi \quad (1)$$

where  $\Delta X_i$  represents the change in the total amount of accumulated input supporting alternative  $i$ ,  $I_i$  is the momentary value of the input to alternative  $i$ , and  $\xi = N(0, \sigma)$  is a Gaussian random variable corresponding to external and internal noise with mean zero and standard deviation  $\sigma$ .



**Figure 2:** Schematic representations of four choice models. (a) Independent race model: Information is transferred (in the form of neural activation) from input units (bottom) to accumulator units (top) without any interactions. (b) Race with global input normalization: Similar to (1) but in addition, information is also transferred to a global normalization unit that consequently normalizes all inputs before they enter the respective accumulators. (c) Feed-Forward input competition: input units simultaneously transmit 'positive' input to their corresponding accumulator and 'negative' input (inhibition) to the opposing accumulators such that accumulators accumulate only the difference between them. (d) Response competition: information reaches the accumulator without interacting but each accumulator applies inhibition (proportional to its activation) to the other accumulator while also losing information via decay.

### Competitive Models

Competition is believed to be important in many psychological processes. For example, inhibitory interactions in the visual system have been found as early as V1 and even at the level of retinal ganglion cells in the eye itself (Cook & McReynolds, 1998; Grossberg, Mingolla & Ross, 1997). At higher processing levels, competitive mechanisms have been suggested to underlie a variety of processes, such as decision

making (Bogacz et al., 2007; Roe, et al., 2001, Usher & McClelland, 2001; Wang, 2002) and attentional selection (Lee, Itti, Koch & Braun, 1999; Mordkoff & Yantis 1991).

Competition can be introduced into a model in many different forms yet it is sometimes unclear exactly how they differ from one another. To elucidate this matter, we will map the competition space in a functional manner according to its locus and examine its effect on model behavior. The first type of competition we discuss, stimulus competition, is often overlooked since it is not even functionally present in the brain. However, it has great influence over model assumptions and behavior as well as over the specific choice of experimental design and stimuli used to test the models. Therefore, accounting for this type of competition is instrumental in distinguishing different model architectures from one another.

### **Stimulus Competition**

Traditionally, most stimuli used to test perceptual choice theories have been one-dimensional (1D). By 1D we mean that information about the different alternatives moves along a 1D continuum from an extreme of total support for one alternative to a contrasting extreme of total support for the other. A few examples are tasks that require participants to make vertical/horizontal decisions when presented with a diagonal line at various orientations, make left vs. right motion decisions when presented with coherently rightwards/leftwards moving dots embedded in a background of randomly moving dots (Shadlen & Newsome, 1998), or make “predominant black/white” decisions when presented with black & white pixel matrices (Ratcliff & Smith, 2010).

One important property of such stimuli is that they are intrinsically competitive in the sense that increasing support for alternative A necessarily reduces support for alternative B. To illustrate, think of a 'more vertical' (V) vs. 'more horizontal' (H) orientation decision involving a line of variable orientation. The more the line's orientation is close to vertical (supporting V) the further away it is from horizontal (supporting H). From a neural perspective one can imagine an observer that decides on the basis of two orientation sensitive neural detectors (V' & H'), which respond optimally to either vertical or horizontal lines, respectively. If the stimulus corresponds to a diagonal 45 deg. line, both detectors fire equally fast. However, as we begin to tilt the line towards vertical, detector V' will begin to fire more vigorously while the firing rate of H' will gradually decrease. This will occur even if the two detectors are completely independent. Therefore, for such stimuli the evidence coming into the decision mechanism is, in a sense, competitively dependant. However, this occurs not because information was processed competitively but because it entered the system that way.

Competitive stimuli of this sort are not ideal to test competition in the decision mechanism as they introduce competition even before the physical energy of the stimuli impinges on the senses. Thus, any attempt to distinguish models on the basis of differences in competitive interactions while using 1D stimuli is likely to fail since observed behavior would reflect both the competitiveness of the stimulus and any competitive processing that might be present in the decision mechanism, rendering them effectively indistinguishable. To illustrate how this problem could be circumvented imagine a similar 'orientation discrimination' task involving two, instead of one, spatially separated lines of variable orientation. In contrast to the 1D

example, instead of making a 'more vertical' vs. 'more horizontal' decision, the task is now to determine which of the two lines, right (R) or left (L), is more horizontal. As before, we can think of an observer deciding on the basis of two orientation sensitive neural detectors (R' & L'), both responding optimally to horizontal stimuli but at different retinotopic locations (R' & L' have non-overlapping receptive fields). Unlike in the 1D version, we can now increase the decision relevant, perceptual quality of the R choice by making the line on the right more horizontal while keeping the orientation of the line on the left constant. Importantly, this will lead to stronger activation of R' without any change in the activation of L' thus allowing for the independent manipulation of the two alternatives<sup>1</sup>.

Furthermore, the use of competitive stimuli in experimental paradigms also imposes certain limitations on model assumptions. For example, in order to accommodate such stimuli, some models have used an input *normalization* assumption (Brown & Heathcote, 2008; Usher & McClelland, 2001), which states that the sum of the inputs to the different accumulators is kept constant (typically

$$\sum_{i=1}^n I_i = 1; \text{ but see Donkin, Brown \& Heathcote, 2009). In this case, the normalization}$$

assumption reflects the nature of the stimulus and can correspond to any model (independent or competitive) that faces 1-D stimuli of the type discussed above. However, such an assumption becomes unnecessary for stimuli that allow for the independent manipulation of value for each different alternative.

The main aim of this paper is to distinguish between independent and competitive models of value integration (see runner metaphor in section: 'Qualitative predictions to probe choice-competition' below). Consequently, it was imperative for us to eliminate any *external* factors, such as intrinsically competitive stimuli, that might mimic or mask the effects of *internal* competitive interactions. For this distinction to be made, it is crucial for the experimenter to be able to manipulate the value of the task relevant, perceptual quality of each alternative independently of the others. This constraint will play a principal role in our choice of stimuli, which will, among other things, avoid 1-D type stimuli that do not allow for such independent manipulations.

### Input Competition

Input competition is the first level at which competition can act within the brain. By input competition we refer to any competition that occurs from the moment the physical energy of the stimulus has been transformed into neural code and up to, but not including, the information accumulation stage of the decision mechanism (see Figure 1). A segregation of the flow of information into a sensory stage and a later decision stage has been proposed by Smith and Ratcliff (2009) as part of an integrated theory of attention and decision making. In their theory a *sensory response function* that is dependent "*on stimulus contrast and on the properties of the early spatiotemporal filters that encode the stimulus*" (p. 287), provides input to a visual short term memory (VSTM) module which in turn feeds evidence to the decision module. In their study, the VSTM component was necessary due to the use of briefly

---

<sup>1</sup> The decision can still, of course, be made on the basis of a 1D variable (for example the difference in firing rate between R' & L'). However, this would now be part of the properties of the decision mechanism rather than the properties of the stimulus and as such would need to be accounted for within the framework of the model.



presented masked stimuli. However, the equivalent of a *sensory response function* can be thought of as an intermediate stage, mediating between the physical stimulus and the higher level, distinct, decision mechanism. Building on this framework, we suggest a distinction between two different types of input competition, one occurring during the early sensory encoding stage and another during the *entry* into the decision stage.

In support of this perceptual/decision distinction, Philiastides, Ratcliff & Sajda (2006; see also Philiastides & Sajda, 2006) found evidence for a late post-sensory decision component and a separate, early perceptual component in EEG neural activity of observers performing either a face/car categorization or a red/green discrimination, according to a pre-stimulus cue. Stimuli for both conditions were the same and constituted face or car pictures that could be either green or red. Philiastides et al. (2006), first reported an early component (N170) that discriminated between faces and cars (higher for faces), but was equally strong in both the face/car categorization task and in the red/green discrimination task; note that in the latter task, the content of the picture (a face or a car) was irrelevant to the decision task. Therefore, the authors concluded that "*this component represents an early perceptual event and is not directly linked to the actual decision*" (p. 8972). Such a component could be mediated by an early perceptual area sensitive to *local* features.

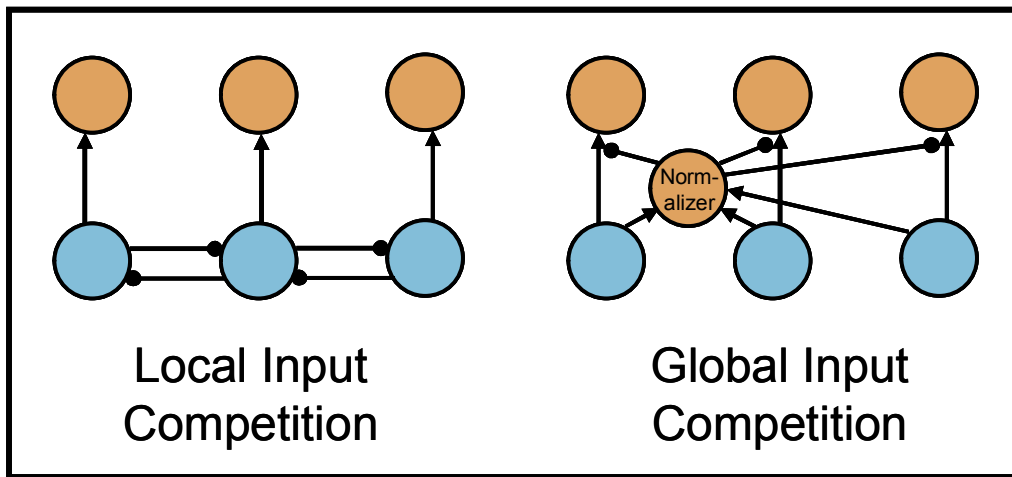
In the same study, Philiastides et al., (2006) also found a later component (300 to 450 ms after stimulus) that discriminated between faces and cars (stronger for faces). Unlike its faster counterpart, this component was much stronger in the face/car task than in the green/red task indicating a link to the decision mechanism itself. Furthermore, the late component was also (i) highly predictive of behavioral accuracy and (ii) its strength strongly correlated with drift rates as computed from fits of the diffusion model to the behavioral data. These observations led the authors to conclude that "*the late component represents the post-sensory evidence that is fed into the diffusion process that ultimately determines the decision*"(p. 8973). Philiastides et al., (2006) also show that "*because the late component is stimulus locked and does not persist until the response, it does not predict the trial-by-trial RT distribution within a coherence level*" (p. 8973). Therefore, this component is compatible with the inputs to the decision process but is unlikely to reflect the accumulation stage itself.

Using this conceptualization, we propose that *input-competition* can occur at two possible hierarchical loci according to the level and extent of processing. The first level where input competition can manifest is a lower sensory encoding level which can be thought of as corresponding to the *sensory response function* or the early N170 component mentioned above. Due to the dependency of early perceptual encoding on stimulus properties we assume competition at this level to be *stimulus-specific* in the sense that different stimuli evoke different amounts of competitive interactions. In the visual domain, for example, interactions at low levels of *visual* processing have been shown to occur only among neighboring cells or within the same receptive fields (Luck, Chelazzi, Hillyard & Desimone, 1997; Moran & Desimone, 1985). We thus address this type of interaction as *local* input competition.

The second level where input competition can manifest is at a higher processing level that might correspond to the VSTM module of Smith and Ratcliff (2009) or the late component found in Philiastides et al., (2006). We assume competition at this level is not directly contingent on specific stimulus properties but rather depends on

the structure of the decision mechanism where all available evidence is incorporated. Thus, any interactions occurring after the initial sensory encoding stage but prior to the evidence accumulation stage and are not contingent on particular properties of the sensory modality or of the stimulus itself are, henceforth, referred to as *global* input competition.<sup>2</sup>

Global input competition is therefore directly relevant to the theoretical distinction between decision theories since it is part of the decision process. On the other hand, local input competition and stimulus competition are only mediators between the object of interest (the stimulus) and the decision process, which are not directly addressed by decision theories. In this study, we are interested in properties of the *general* decision making mechanism, which can incorporate both *global* input competition and *response* competition (see section: 'Response competition' below). Thus, in our experimental manipulations we aimed at minimizing both stimulus competition as well as any *local*, momentary interactions in order to avoid confounding them with higher competitive processes of the kind discussed below.



**Figure 3:** Global vs. Local Input competition. Local input competition (left) occurs at early processing stages and is only effective between spatially adjacent visual stimuli. Global input competition (right) occurs at higher 'decision' stages and takes into account *all* decision relevant stimuli regardless of their spatial arrangement.

One noteworthy example of a *global* input competition model is Vickers' accumulator model (Vickers, 1970). Here, at each time step, inputs (samples that match a perceptual hypothesis) are compared either to each other or to a common criterion (depending on the task). This comparison outputs a pre-accumulation (post-perception) decision indicating which accumulator is best supported by the momentary evidence. As a result, for that sample, only the winning accumulator receives any input (in this case a normally distributed random variable with a fixed mean). Alternatively, the comparison unit in Vickers' *accumulator* model could be replaced by a normalization unit so that all inputs are normalized before entering their

2 It is not clear that this distinction between local and global input competition is, de facto, present in the brain. While there is ample evidence for the presence of local competition in the visual system, the concept of global input competition is still unsupported by empirical studies (but see Piliastides et al., 2006). However, we find this distinction useful from a theoretical point of view since prior attempts at modeling perceptual choice have implemented competitive mechanisms that can be interpreted as either local or global input competition.

respective accumulators (see Figure 3(right)). This can also be considered a variant of a *global* input competition model as long as the normalization process does not depend on specific stimulus attributes. Other examples for global input competition models include the recruitment model (La Berge, 1962) and the Feed Forward Inhibition (FFI; Figure 2c) variant of the diffusion model (Mazurek et al., 2003; Niwa & Ditterich, 2008).

As competition is introduced at increasingly higher processing levels, another issue arises. Competitive models, unlike the independent ones, are not as straightforward to extend from binary to multiple alternatives. Again Vickers' accumulator model is a good example. Comparing two inputs in the accumulator model is straightforward and requires only one comparison, comparing three inputs requires three comparisons, and this scales up combinatorically.

Since we intend to use both two-alternative and multi-alternative choice in our experiments, it was important to use only models that can be extended to any number of alternatives. For this reason, and in addition to the normalized race model, we consider a close relative of the classical *Drift Diffusion Model* (DDM; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998), the *Feed Forward Inhibition* model (FFI; Figure 2c; see Mazurek et al., 2003 for two alternatives; Niwa & Ditterich, 2008 for three alternatives; Roe et al 2001 for multi-alternative, multi-attribute decisions), as a representative of the input-competition model category. This model allocates separate accumulators for the different alternatives, which then race each other towards a common decision boundary. Each accumulator  $X_i$  receives positive activation from the input to its corresponding alternative ( $I_i$ ) and negative activation equal to the average input to the other alternatives ( $\frac{1}{n-1} \sum_{j \neq i} I_j$ ). Note that for  $n=2$ , the FFI is equivalent to the DDM, even though it employs two parallel diffusion processes ( $I_1 - I_2$  &  $I_2 - I_1$ ) racing towards a common threshold rather than one diffusion process with two (upper & lower) thresholds, as in the classic model.

The equation describing the FFI's accumulation for the  $i$ 'th accumulator  $X_i$  in an  $n$ -alternative choice task is:

$$\left\{ \Delta X_i = I_i - \frac{1}{n-1} \sum_{j \neq i} I_j + \xi \right. \quad (2)$$

As in Equation 1 before,  $\xi = N(0, \sigma)$  is a Gaussian noise parameter with mean zero and standard deviation  $\sigma$ . As we can see, in the FFI model inputs compete by inhibiting each other directly.

### Response Competition

Response competition refers to any competitive interactions that occur at the accumulator level (i.e., the integrated values) and whose strength is proportional to the activation level of the accumulators themselves. Thus, unlike input competition models, here, the competition does not involve momentary values, but rather depends on the *total* amount of integrated value. The accumulators' activations, in sequential sampling models can be conceptualized as representing, the degree of belief in each hypothesis or the current tendency towards executing a certain response (hence the

name 'response competition'). Examples of response competition models are the classical Drift Diffusion Model (Ratcliff, 1978; Ratcliff & McKoon, 2008), the Max Minus Next variant of the diffusion model (MMN; McMillen & Holmes, 2006) the Leaky Competing Accumulator model (LCA; Figure 2d Bogacz et al., 2007; Usher & McClelland, 2001), the attractor model (Albantakis, & Deco, 2009; Wang, 2002; Wong & Wang, 2006) the Ballistic Accumulator (BA; Brown & Heathcote, 2005) and a variety of Bayesian decision models (Bogacz, 2009; Ditterich, 2010). In the LCA model, for example, lateral inhibition and neural leak (i.e., decay of integrated values or activations) is applied to separate accumulators thus interpolating between the benefits of both race and diffusion models.

In the following computational studies, this category will be represented by both the LCA and the MMN models. In the LCA each alternative is assigned a separate accumulator. Lateral inhibition between the accumulators results in response competition which is then balanced by leakage of accumulated activation from each accumulator. The activation level of accumulator  $i$  ( $X_i$ ) in a choice involving  $n$  alternatives is updated with each time step according to the formula:

$$\begin{cases} \Delta X_i = -\lambda X_i + I_i - \beta \sum_{j \neq i} X_j + \xi \\ X_i(t+1) = \max(X_i(t) + \Delta X_i, 0) \end{cases} \quad (3)$$

Where,  $I_i$  are the inputs,  $0 < \lambda < 1$  is the leak,  $0 < \beta < 1$  corresponds to the inhibition and  $\xi = N(0, \sigma)$  reflects the noise in the integration process assumed to be Gaussian with zero mean and an SD of  $\sigma$ . The Max function in the bottom equation reflects a non-linearity imposed on the activations (a reflecting boundary), which is motivated by the fact that neural activity is bounded from below (for a more detailed discussion see Usher & McClelland, 2001, p.14 and Appendix A; as well as Bogacz, et al., 2007). This neural property is approximated by maintaining  $X_i \geq 0$  so that when activation becomes negative it is immediately truncated to zero. Note that for the special case where  $\lambda = \beta = 0$  the model reduces to the purely independent model described in Equation 1. Furthermore, when  $0 < \lambda = \beta < 1$  the model is said to be balanced and the 2-alternative version of this model (minus the non-linearity) can be thought of as equivalent to the classical drift diffusion model (Bogacz et al. 2006).

The MMN, on the other hand, can be regarded as an independent race model with a competitive stopping rule. While the independent race model stops integrating evidence and executes a decision when one of the accumulators reaches a predetermined decision criteria, the MMN stops only when the *difference* between the largest and the second largest accumulators reaches a predetermined decision criteria (hence the name Max Minus Next). Therefore, the accumulation process for the MMN can be described by equation 1, supplemented by a competitive termination rule:

Decide in favor of alternative  $m$  if:

$$\begin{cases} \max \{X_i^n\} = X_m \\ [X_m - \max \{X_{j \neq m}\}] > \theta \end{cases} \quad (4)$$

Note that for  $n=2$ , the MMN is practically equivalent to the standard DDM and the FFI. The only difference between them lies in the order of accumulation and subtraction. While the FFI first subtracts the momentary inputs and only then accumulates them, the MMN first accumulates and then subtracts (see study 2 for an illustration of where this difference becomes substantial).

### *Auxiliary Model Assumptions*

To account for empirical data, *all* the models discussed above require additional assumptions about external sources of between trial variability. These could manifest as starting point variability, drift-rate variability, variability in decision criteria, variability in the non-decision component of RT as well as the specific choice of distributions from which these random variables are sampled (Dyrholm, Kyllingsbæk, Espeseth, Bundesen, 2011; Ratcliff & Smith, 2004). While essential for dealing with certain aspects of the data like fast and slow errors, the skewness of RT distributions and bounded asymptotic accuracy, such assumptions affect neither the underlying mechanism of the models nor their affiliation with one class of competition or another. Discussion of these assumptions will, therefore, be deferred to the section on data fitting below.

To conclude, we examined three distinct classes of decision models: independent, input competition and response competition. These classes of models represent different mechanistic theories of value integration and decision (see Table 1 for a more exhaustive classification of popular models according to the taxonomy outlined above). In order to discriminate between these classes we must draw clear predictions from each one and test them against empirical data. How to do this best, however, is not always straightforward when dealing with stochastic, computational models due to the complexities of using data fits for model comparison and the limitations of conclusions derived from such methods (for more detailed discussions on this topic see Pitt & Myung, 2002; Pitt, Myung & Zhang, 2002; Roberts & Pashler, 2000; Jacobs & Grainger, 1994).

Category Model	Independent	Input Competition	Response Competition
Independent Race	+		
Poisson counter	+		
LBA	+		
Normalized Race		+	
Recruitment		+	
Accumulator		+	
Diffusion – Wiener		+	+
Diffusion – OU		+	+
Diffusion – FFI		+	
Diffusion – MMN			+
BA			+
LCA			+
Leaky Accumulator (LA)	+		
LA – Relative Criteria		+	+
Attractor			+
Basal Ganglia			+

**Table 1:** model taxonomy of various models in the literature with regards to type of competition. Independent race (Logan et al., 1984), Poisson counter (VanZandt et al., 2000), LBA (Brown & Heathcote 2008), Normalized race (LBA with input normalization; Brown & Heathcote 2008), Recruitment (LaBerge, 1962), Accumulator (Vickers, 1970), Diffusion – Wiener (Ratcliff 1978, 1988), Diffusion – OU (Ratcliff & Smith, 2004, Usher & McClelland, 2001), Diffusion – FFI (Niwa & Ditterich, 2008), Diffusion – MMN (McMillen & Holmes, 2006), BA (Brown & Heathcote, 2005), LCA (Usher & McClelland, 2001), Leaky Accumulator (LA; Ratcliff & Smith, 2004), LA – Relative Criteria (Ratcliff & Smith, 2004), Attractor (Wang, 2002), Basal Ganglia (Bogacz & Gurney, 2007).

To deal with this complexity, we undertake a comprehensive approach to theory testing, also inspired by Roberts & Pashler (2000), which is based on strong inference experiments (Jewett, 2005; Platt, 1964) where stimulus manipulations are specifically chosen to probe narrow, non-overlapping predictions about specific measures of behavior. To this effect, and in addition to model fits, we also employ a specialized graphical display that is explicit about a wide range of possible, as well as impossible, model predictions. This is realized by randomly varying model parameters and plotting predictions on a two dimensional plot<sup>3</sup> with the axes representing the relevant behavioral measures. To allow for the simultaneous evaluation of both the quality of the data and the amount of support it provides for the theory, we also plot, in the same figure, the individual empirical observations with

3 Note that, due to the 2D restriction, the use of this graphical display is critically dependent on first making narrow predictions relating to the interactions of no more than two observable measures of behavior at a time. That is because, when making simultaneous predictions about complex interactions between multiple behavioral measures, as is common in studies that fit models to data, a display that captures predictions for entire parameter spaces is difficult to manufacture.

error bars representing the variability on the relevant measurement dimension, (for illustrations see Figure 26.5 Roberts & Sternberg, 1993; Figure 9, Tsetsos, Usher & McClelland, 2011; and Figure 10 & 16, this paper). Finally, running a statistical analysis to test if the predicted effect is significant provides an additional, more precise, estimation of the quality of the data.

### **Qualitative predictions to probe choice-competition.**

We begin with an informal description of the rational behind the predictions that distinguish between the classes of models before we continue to the experimental and computational studies that test them.

#### *1. Independent vs. competitive models.*

The mathematical properties of independent processes have been studied extensively and have been shown to be distinct from those of interactive or co-activation models (Townsend, 1972), resulting in unique predictions for certain stimulus manipulations. Independent models are characterized by the absence of interactions between parallel processing channels. For such processes, a reduction in the amount of input (i.e. lower drift rate) to at least one of the channels is a necessary condition for an increase in the termination time of the decision process (assuming of course that stimuli are intermixed such that stopping criteria do not change between conditions; Ratcliff & Smith, 2004; Ratcliff & McKoon, 2008; VanZandt et al., 2000). This leads to the unique prediction that an increase in the value of the incorrect alternative will *not* slow down, and might even speed up, the termination time of the process. To see this, consider the following metaphoric illustration of the difference between an independent and a negatively interactive (i.e. competitive) model.

*The runner metaphor:* Imagine a race between two runners that can not assist or hinder each other in any way and, for that matter, are not even aware of each others' position at any given time. Now, consider two such races: *race-1*: a race between a fast runner (F) and a slow runner (S); *race-2*: a race between the same fast runner (F) and a medium runner (M). On average, finishing times for *race-2* would be faster than for *race-1*. This happens since runner (F) is just as fast in both races but runner (M) is faster than runner (S). So, runner (F) loses more of his slower runs to runner (M) in *race-2* than to runner (S) in *race-1*, resulting in a speedup of overall finishing times. This phenomenon is aptly named statistical facilitation (Luce, 1986; Raab, 1962; Townsend & Nozawa, 1995) and is most evident in independent processes. Let us now introduce competitive interactions into this situation. Assume that, as the runner who is behind gets closer to the leader, her ability to slow the first runner down (say, by pulling on her shirt), improves. In contrast to the independent race, now (with competition on) *race-2* will result in slower finishing times since the medium runner (M) in *race-2* has more opportunity (compared with the slow runner (S) in *race-1*) to hinder the fast runner (F).

Following from this simplified analogy, one can predict that manipulating task difficulty by increasing the momentary, task relevant, value of the weak alternative (i.e., replacing the slow runner with a medium runner) should provide a way to distinguish purely independent models from competitive ones. This metaphor will be

the guiding principle in the computational and experimental investigations we present here and will help us distinguish independent from competitive models.

## 2. *Input vs. response competition models.*

The key characteristic of response competition models is that competitive interactions are proportional to the amount of accumulated evidence. This leads to the unique prediction that an increase in the starting point of one of the non-target accumulators would also increase the total amount of competition in the system and therefore slow down overall RTs. To illustrate how this can be used to distinguish choice models let's consider a second metaphor, which involves two hypothetical academic contests (1 & 2) between two competing scientists.

*Competing scientists' metaphor:* In both contests the first scientist to reach N publications on a given topic wins a substantial prize. Now, imagine that the rivaling scientists are: i) a fast publisher (F), ii) a slow publisher (S), both keen on winning the prize. For the sake of the example let's assume that both scientists have no previous publications in the field, they always review each others papers and that they always give each other negative reviews<sup>4</sup>. Let's also assume that in *contest-1* the (negative) opinion of a reviewer is weighted proportionally to the number of papers she has published on the relevant subject (a form of response competition). Under these circumstances, it is clear that in *contest-1*, the more publications a researcher has accumulated the more power she has to thwart the others' publication efforts and hence slow down the rate with which her opponent accumulates them.

Now, instead of replacing (S) with a medium publisher (M) (like we did in the previous 'runner' metaphor), let's observe what happens when at the beginning of the contest (S) has a head start of several published papers in the field (we denote the slow publisher with a head start by (S+)). Thanks to the head start, the (negative) opinion of (S+) is now more valued in the reviewing comity than that of (S) and she can thus slow down (F) more than (S) could have. Therefore when (F) goes up against (S+) overall finishing times would be slower than when she faces (S).

The critical comparison here is between the former *contest-1* and a similar *contest-2* where everything is the same except that a reviewer's opinion is weighted not by her accumulated publications in the field but rather by the time that has passed since her last publication (sort of a momentary 'publication drift rate'). This is a form of input competition because the faster a publisher is the more she would be able to slow down her opponent. However this ability now does not take into account the total amount of publications accumulated by the scientist in the field of interest. Consequently, in *contest-2* the ability of (S+) to slow down (F) does not differ from that of (S) despite the fact (S+) has had a starting advantage of several published papers to begin with. Thus, no slowdown in finishing times should be observed for the closer competition ((S+) vs. (F)). Moreover, (S+) will, on average, reach the goal of N publications faster than (S) would have, thanks to the head start. As a result, and in contrast to *contest-1* where replacing (S) with (S+) resulted in increased finishing times, here overall finishing times would decrease due to statistical facilitation. This example encapsulates the underlying rationale behind study 2.

To formally test these predictions, we will now present three combined computational and experimental studies. In the first one, we manipulate the task

---

4 Fortunately, this story is imaginary and does not correspond to real situations in our field.



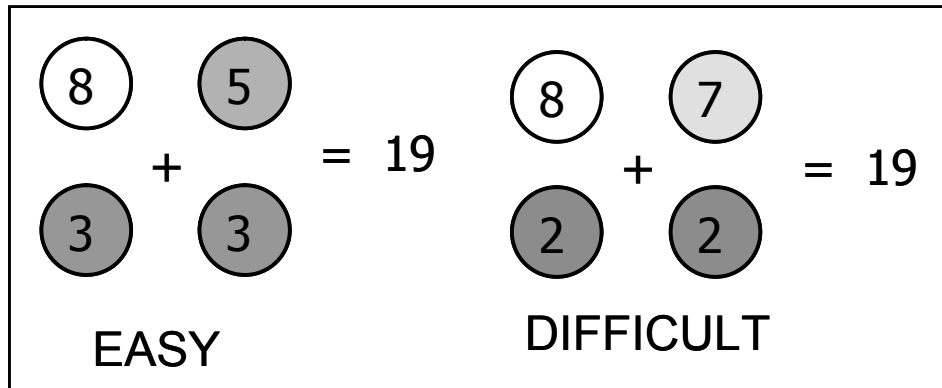
relevant values of the non-target alternatives in a multi-alternative choice task in such a way that it results in qualitatively different predictions for competitive and independent models. In the second study, we probe the level at which the competition is implemented by the use of pre-cues that affect the priors of the various alternatives. In the third, we control the momentary values of the non-target alternative, either at the beginning, or throughout the duration of a trial, while also directly probing for normalization of input strength.

To summarize, our approach can be segmented into several core components: (1) theoretical taxonomy – which model belongs to which category with regard to the *central* assumption we are going to scrutinize; (2) generation of diverging predictions – how can specific manipulations of inputs (independent variables) discriminate between the models with regards to particular measures of behavior (dependent variables); (3) design of strong inference experiments – how can the theoretical manipulations be translated into an experimental design; (4) choosing appropriate stimuli – which stimuli have quantifiable informational contents that are both compatible with the experimental design, and allow us to use the physical magnitudes of the stimulus alternatives to constrain model inputs and the momentary, task relevant, perceptual values that underlie them (for similar constraints see Niwa & Ditterich, 2008; Palmer, Huk & Shadlen, 2005); (5) data collection, model fits, analysis of predicted main effects and comparison of prediction spaces. In doing so we hope to provide narrow but conclusive results relating to the *central* theory assumptions of independent vs. competitive processing. Our goal, however, is not to support or reject any *particular* model. Rather, we focus on entire categories (i.e. mechanisms or theories) as per the taxonomy presented above. The models we test here are merely exemplars of each category and are used for illustrative purposes. This should help us begin to disentangle the tight cluster of flexible and resilient choice models.

### **Study 1: Independent Manipulation of the Evidence**

To examine competition in value integration, we use a paradigm in which at each time frame the stimuli provide independent values for four alternatives. Since we are interested in the general decision mechanism, we also aimed to minimize local interactions that are stimulus-specific. For this reason we choose our stimuli according to three guiding principles: (1) large spatial separation – visual stimuli that are well separated in space are unlikely to interact during low level processing (Luck et al., 1997; Moran & Desimone, 1985); (2) temporal separation - stimuli that do not overlap in time are less likely to be processed together and therefore less likely to interact at the perceptual level; (3) processing simplicity - receptive field size is known to increase with stimulus complexity and local interactions are likely to be present in neurons that share receptive fields (Desimone & Ungerleider 1989; Kastner & Ungerleider, 2000). Therefore, in addition to keeping the stimuli spatially, and in Experiment 1b also temporally, separated we chose our task relevant perceptual dimension with the goal of minimizing receptive field size in mind. To this end, we use a brightness discrimination task for Experiment 1a and a flicker rate discrimination task for Experiment 1b that are assumed to tap into only the most basic and least processed perceptual information for which interactions were found to be very localized (Burr, Ross, & Morrone, 1985).

Importantly, for brightness stimuli, there is a simple monotonic correspondence between the spatially distinct, momentary, physical intensities of the stimuli and their representations as neural activations in separate retinotopic areas of the primary visual cortex. The activation representing the brightness value of a given stimulus alternative is thus considered as the momentary, task relevant, value of (or input to) that alternative which can then be processed and accumulated towards a decision<sup>5</sup>. The task in Experiment 1a, then, is to detect which of four round gray patches of fluctuating brightness, is the brightest overall (Caspi, Beutter & Eckstein, 2004; Ludwig, et al., 2005; for illustration see Figure 4 & 7).

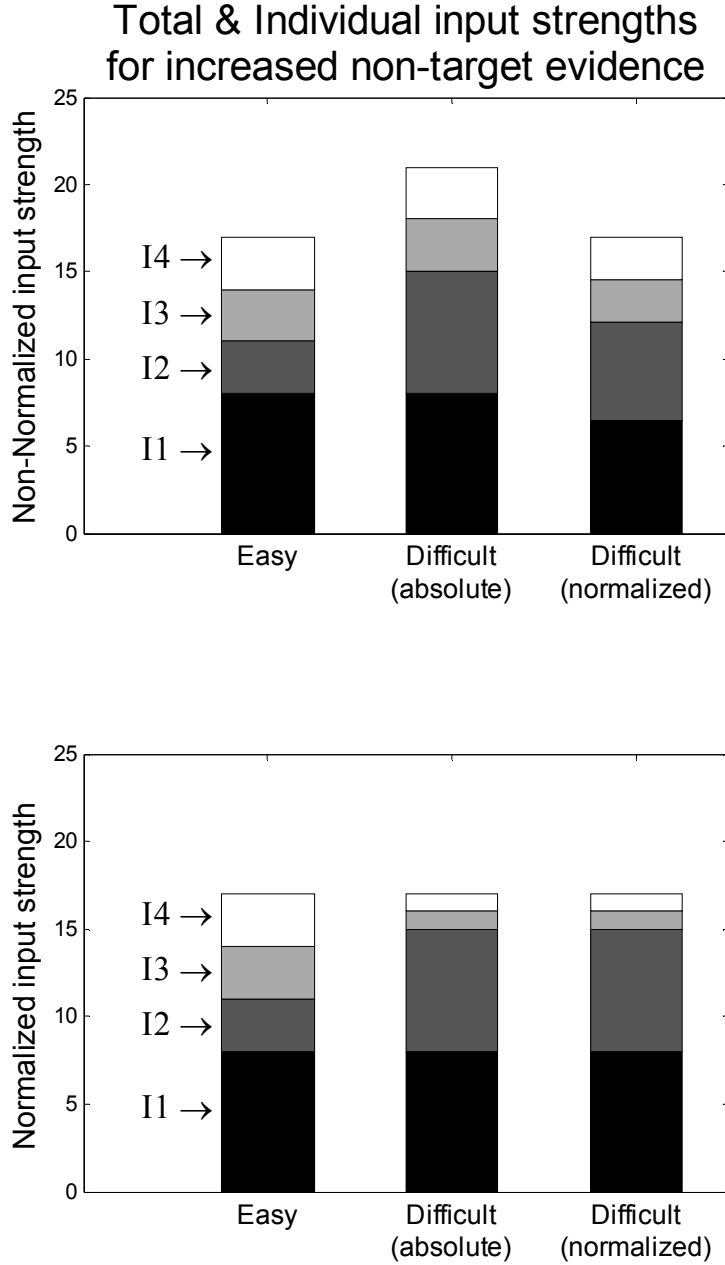


**Figure 4:** Illustration of stimuli for the two conditions in experiment 1. Numbers in gray patches represent brightness levels (0=black to 10=white); assume a dark background (unlike in this illustration). As can be seen, the sum of total brightness level was kept constant throughout the different conditions.

Assume, for example, that participants are presented with a stimulus that corresponds to four patches of fluctuating brightness, whose mean brightness are,  $b_1 > b_2 > b_3 > b_4$ . We can now carry out a manipulation in which the mean brightness of  $b_2$  increases towards  $b_1$  while  $b_3$  and  $b_4$  are kept constant (i.e., we make the task more difficult; see Figure 5 (top panel) for illustration). This is equivalent to replacing runner (S) with runner (M) in the *runner* metaphor. We can then measure whether this manipulation results in a speed-up of mean response times (RT), as predicted by independent models or in a slow-down, as predicted by competitive models. This procedure, however, does not discriminate between the various types of

5 Note that these simple correspondences between the physical intensity, its momentary representation as a perceptual value and the input that is accumulated towards the decision are, to a large extent, task dependent and may change with task demands. For example, if the task is to choose the brightest alternative, then physical intensity, perceptual value and input can be mapped to each other through monotonic, positively correlated mappings such that the brighter a stimulus is the higher its perceptual value is and the stronger the input to its corresponding accumulator. In fact, in this simple setting the perceptual value and the input are identical. If the task, however, is to choose the dimmest alternative then the perceptual value might have to be mapped to the input through a monotonic, *negatively* correlated mapping such that the higher the perceptual value is, the weaker the input to the corresponding accumulator. Alternatively the positively correlated mapping may be maintained and an eliminatory strategy used instead. Similarly, more complex stimuli such as orientations, letters and words do not convey information through physical intensity but rather through more complex features and therefore would necessitate more complex mappings between stimulus and perceptual representation. This observation further stresses the importance of choosing appropriate stimuli that are simple and easily mapped to task demands. We thank an anonymous reviewer for drawing our attention to this.

choice competition, such as, response (e.g., MMN or LCA) vs. input (e.g., normalized race models or FFI).



**Figure 5:** Illustration of the effects of normalization on a hypothetical, brightness discrimination task similar to the one depicted in Figure 4. The left bar represents the perceived brightness (the height of each distinct grey-shade corresponds to the brightness of one spot) of the stimuli for baseline input strengths as in the easy condition of Experiment 1a & b (for clarity inputs strengths have been altered from the ones used in the experiment). Middle and right bars represent the perceived brightness of the stimuli in the difficult condition for two alternative model types (normalized and absolute accordingly). The top panel illustrates these effects when input strength for one non-target alternative is increased while all others inputs remain the same (easy condition:  $I_1(\text{target})=8$ ,  $I_2=I_3=I_4=3$ ,

sum( $I_i$ )=17; difficult condition:  $I_1$  (target)=8,  $I_2$ =7,  $I_3=I_4$ =3, sum( $I_i$ )=21). The bottom panel illustrates these effects for an input manipulation where one non-target input is increased (as in the top panel) but the two remaining non-target inputs are reduced to compensate for the former increase and keep the total input normalized (easy condition:  $I_1$  (target)=8,  $I_2=I_3=I_4$ =3, sum( $I_i$ )=17; difficult condition:  $I_1$  (target)=8,  $I_2$ =7,  $I_3=I_4$ =1, sum( $I_i$ )=17).

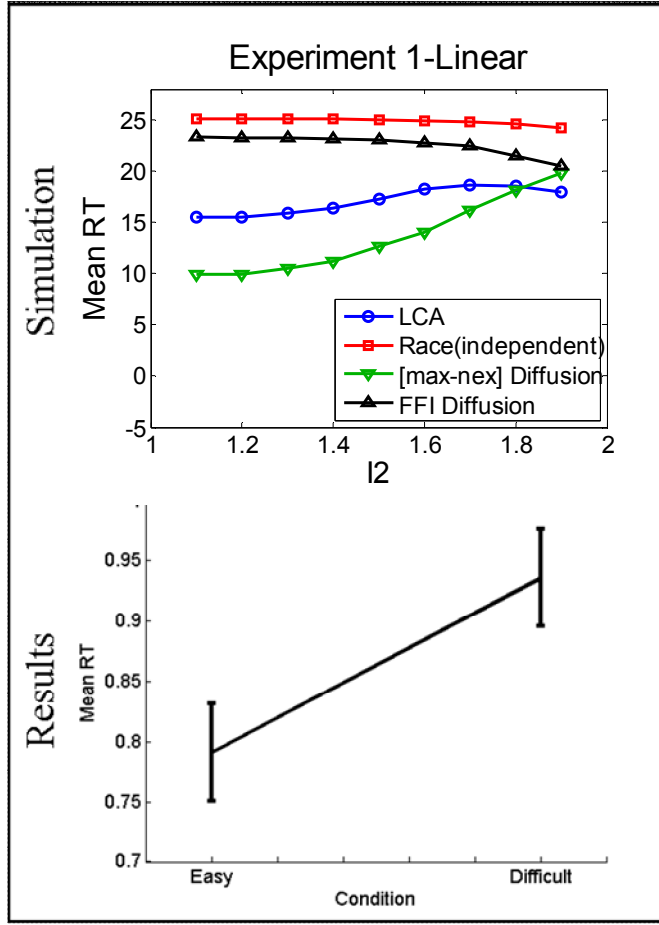
We can illustrate this lack of discrimination by considering the normalised race model -- a case of global-input competition (Figure 3, right panel). As shown in Figure 5 (top panel, right bar), a normalization of input strengths has the effect of reduced input for the target ( $I_1$ ) as a result of increased input for the strongest non-target ( $I_2$ ). This could then account for a slow-down of RT with increased non-target input. A similar slow-down is also predicted by the FFI model, where the momentary activation for any accumulator is computed as the momentary input of that respective alternative minus the average momentary input of the other alternatives. Since the increased non-target value is subtracted from the total input to the target decision-unit, the result is a general slow-down of RTs.

We can, however, distinguish between response-competition and some input competition models, by co-varying the brightness of the remaining two spots ( $b_3$  &  $b_4$ ). To do this, we can make the task more difficult by increasing  $b_2$  towards  $b_1$  while at the same time maintain the normalization by lowering the brightness of the other two spots (see Figure 4, Figure 5 (bottom panel) for illustration). With this additional manipulation, the normalized race model will also predict a speed up with increased input for strongest non-target since target input is not affected by this change (this is because the sum of the inputs is kept constant and there is no re-normalization; see Figure 5 (bottom panel)); we thus effectively equate the predictions of the normalized race model with those of the purely independent race model (shown together in the simulations). As we show below (Figure 6), this manipulation not only discriminates between response-competition, independent and normalized race models, but also discriminates the former from the FFI model. In addition, these predictions are also robust to non-linear, concave, psychophysical transformations of physical intensities into input strengths such as logarithmic and power law functions (see Appendix A for simulations demonstrating this).

Computer simulations were run to formally evaluate the effect of increasing task difficulty via the augmentation of the brightness value of the strongest non-target ( $I_2$ ) on mean-RT. Five models were used: a purely independent (*race*) model (red line; Figure 6), and four competitive models: a normalized race model (red line<sup>6</sup>), *MMN* (green line), *FFI diffusion* (black line) and *LCA* (blue line). As one can see in Figure 6 (top), as the input for the main non-target ( $I_2$  - x axis in Figure 6) increases the independent race and FFI diffusion models predict a speedup of RT, while the response competition models (*LCA* and *MMN*) predict a slowdown in RT with increasing task difficulty.

---

6 Note that, due to our specific choice of a normalized input structure, the independent race model and the normalized race model make equivalent predictions for this manipulation. Therefore both are represented by the same color in the figure.



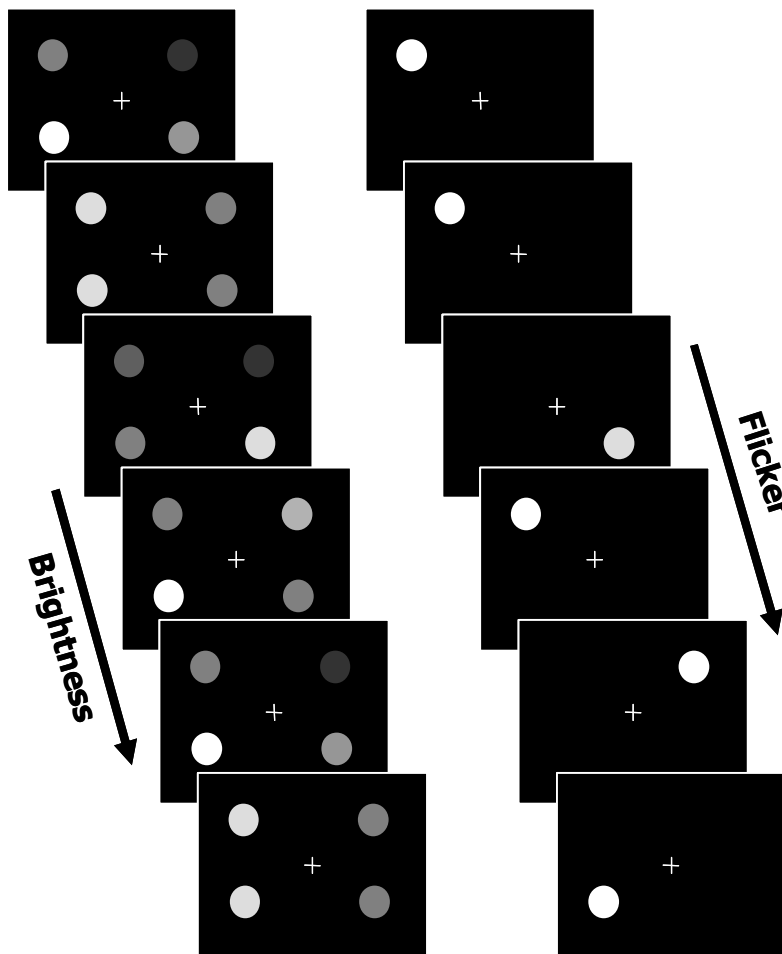
**Figure 6: Top** (Simulation): Mean-RT for three choice models, as a function of the input strength of the brightest non-target ( $I_2$ ); Decision criteria were set so that models predicted approximately the same accuracies for all the various input alternatives (Race: 50; LCA: 12; MMN: 5.5; FFI: 21). Input strengths in the simulations were varied in the following manner:  $I_1$  was kept constant ( $I_1=2$ );  $I_2$  increased in increments of 0.1 from 1.1 to 1.9;  $I_3=I_4$  were decreased in increments of 0.05 from 1.1 to 0.7 in accordance with increases in  $I_2$  such as to maintain overall normalization. The sum of all inputs was kept constant ( $\sum_{i=1}^4 I_i = 5.3$ ). All models were modeled according to the equations described in the introduction ( $\sigma = 1$ ;  $\lambda = \beta = 0.1$ ). **Bottom:** Experimental results; Mean RT for experiment 1a. Error bars correspond to within subject standard errors calculated according to Cousineau (2005) which discounts irrelevant between-subject variance.

The observed speedup effect for the race (red line), either with or without normalization, is due to statistical facilitation in the absence of competition, as discussed above. The speedup observed in the FFI diffusion (black line), however, is the result of both statistical facilitation and competition. To understand why the FFI diffusion behaves like the race model under our manipulation, one can note two things: i) the activation of the target accumulator ( $\Delta X_{\text{target}} = I_{\text{target}} - \frac{1}{n-1} \sum_{i \neq \text{target}} I_i$ ) is not affected by the manipulation since  $I_1$  (the target input) as well as the average of  $\{I_2, I_3, I_4\}$  are maintained constant by the manipulation; ii) the activation of the

strongest distractor accumulator increases (eq. 2) in our difficult condition, not only because the value of  $I_2$  goes up, which is in itself enough to produce statistical facilitation, but also because the mean of  $\{I_1, I_3, I_4\}$  (the inhibition felt by  $X_2$ ) goes down. Thus, we have an even stronger statistical facilitation effect than in the race model: as the strongest distractor finishes faster, it steals more of the slower runs from the target, speeding up, both total and the correct RT. Unlike the race and FFI diffusion, the two response-competition models, (MMN and LCA) show a slowdown of RT with difficulty (green and blue lines), which is the direct result of increased competition between the target and the largest non-target working against the statistical facilitation effect. Experiment 1 was designed to test these diverging predictions.

### **Experiment 1a – brightness task**

Participants were asked to choose, as rapidly and accurately as possible, which of four round gray patches of fluctuating brightness was the brightest overall (Caspi, Beutter & Eckstein, 2004; Ludwig, Gilchrist, McSorley & Baddeley, 2005). Since the brightness of each patch was independently re-sampled on each frame (with noise drawn from a normal distribution), either one of the four circular patches could be the brightest on a particular frame, requiring the participants to integrate the patch-brightness values across time (Figure 7, left). The experiment included two conditions (easy and difficult), which differed in the brightness value of the brightest non-target, effectively mimicking low and high  $I_2$  values in the simulation. The critical dependent variable is the mean-RT for the two (*easy/difficult*) conditions. While all models predict a drop in accuracy in the *difficult* condition compared to the *easy* condition, they differ on their predictions regarding mean-RT. Independent and input-competition models predict a speedup, while models with response competition predict a slow down.



**Figure 7:** Illustrations of experimental timelines for experiments 1a (left) and 1b (right). For the brightness stimuli (left), the brightness levels of each patch varied randomly on each frame (16.6ms). For the flicker stimuli each patch had a certain probability to be white otherwise it was black. Note that in the flicker condition there was also a high probability of all black frames between frames containing white patches (not displayed in figure for reasons of compactness).

## Method

### Participants

Eight Tel-Aviv University undergraduate students (6 female) participated in the experiment in exchange for course credit. Each participant was tested in two, sixty minute long sessions (no more than four days apart). All participants had normal or corrected to normal vision. The projects were approved by the department's ethical committee.

### Materials

All stimuli in this experiment were presented on a ViewSonic Graphics Series G90fB 19" CRT monitor. The monitor was gamma corrected using a TES-1332A photo meter. The stimuli were composed of four homogenous, round, gray patches on a black background (width 1.1cm) positioned at the four corners of an imaginary square relative to a fixation-cross (total width from left edge of left patch to right edge of right patch: 3.7cm). Each patch's gray level fluctuated randomly and independently of the other patches over the course of each trial. For the *easy* condition the gray levels of the target and non-targets were normally distributed around means of 0.4 and 0.2 (on a 0 (black) to 1 (white) scale) respectively. For the

*difficult* condition the gray levels of the target, principal non-target and secondary non-targets were normally distributed around means of 0.4, 0.3 and 0.15 respectively. On each frame the gray level for each individual patch was separately recalculated as the sum of its designated mean plus a Gaussian random variable  $x=N(0,0.1)$  that was cutoff below -0.1 and above 0.1 to prevent obvious flickering of the stimuli that might attract attention to it in a bottom-up fashion. Refresh rate was set at 60Hz (16.6ms per frame) and tests were run to evaluate the probability of dropped frames. No frames were dropped after a full hour of continuous presentation. The location of the target was randomly drawn on each trial.

### Procedure

*Easy* and *difficult* trials were randomized within each block. Responses were given on the 1, 3, 9 & 7 keys of the keyboard number keypad for the bottom left, bottom right, top right & top left responses respectively. Subjects were instructed to use the right index finger and thumb for the 3 & 9 keys respectively and the same fingers on the left hand for the 1 & 7 keys. The stimuli stayed on until the response was entered, after which a short 1sec ISI preceded the next trial. The task was divided into blocks of 60 trials. Each block consisted of an equal number of trials from each condition for a total of 1000 trials per participant. After each block there was a self timed intermission to allow the subject to rest. During each of these breaks, the average accuracy and RT for the last block were presented on the screen. The participants were instructed to try and maximize both accuracy and RT such that if they reached 100% accuracy they should try to respond faster and were given a 30 trial practice block. Subjects were also told to keep their eyes focused on the fixation cross throughout the trial though in the absence of an eye tracker there was no way to verify that they actually complied with this request. The experiment was held in a partially darkened room.

### Results

Participants were less accurate in the *difficult* condition ( $M=0.83s$ ,  $SD=0.07$ ) compared with the *easy* condition ( $M=0.96s$ ,  $SD=0.02$ ;  $z=2.2$ ,  $p<0.05$ ; Wilcoxon Matched Pairs Test). More importantly, however, the manipulation also had an effect on RTs. As shown in Figure 6 (bottom), participants were slower to respond in the *difficult* condition ( $M=1.21s$ ,  $SD=0.27$ ) than in the *easy* condition ( $M=1.11s$ ,  $SD=0.21$ ;  $z=1.99$ ,  $p<0.05$ ; Wilcoxon Matched Pairs Test). This pattern of RT is in accordance with the predictions of the response competition models (LCA or MMN). Independent models are inherently unable to account for such results due to the lack of interaction between channels. Independent architecture inevitably leads to statistical facilitation and can, thus, only predict a speedup in RTs under these circumstances.

The input competition models tested in this study (Normalized race & FFI diffusion) also failed to account for the data presented here. However, each one failed for a different reason. The input normalization race model (Figure 3) is unable to account for the results, precisely because the inputs were designed to make normalization completely invariant to our manipulation. On the other hand, the FFI diffusion is subject to an even stronger statistical facilitation than the independent race model, and it predicts that increasing difficulty in this way involves speeding up



the evidence accumulation of the strongest non-target, even more than for an independent race model.

### **Fitting the models to data**

We fit five models to the data from Experiment 1, one independent model (independent race), two input competition models (normalized race and Feed Forward Inhibition; FFI) and two response competition models (Max Minus Next; MMN, and Leaky Competing Accumulator; LCA). Since the predictions and fits of the independent race model and the normalized race model converge for this input manipulation, we present them together under the title of ‘race model’. For a detailed discussion on fitting methods see Appendix B.

We chose to fit the model to quantile RT data that was averaged over subjects (Forstmann, Brown, Dutilh, Neumann & Wagenmakers, 2010; Ratcliff, 1979; Ratcliff & Smith, 2004; Thomas & Ross, 1980; Wagenmakers, Ratcliff, Gomez & McKoon, 2008;). There were several reasons for this. Since the models made narrow, qualitatively different prediction, we wanted to emphasize the qualitative aspects of the data. Using averaged, instead of individual, quantile RTs helps bring out the qualitative differences by reducing between subject variability (Ratcliff & Smith, 2004). This makes particular sense since all the subjects displayed the same qualitative effects (Ratcliff & Tuerlinckx, 2002). Furthermore, in large studies with multiple subjects per group, it has been demonstrated that parameter values obtained from fitting the model to data averaged over subjects were in good agreement with averaged parameter values from individual subjects (Ratcliff, Thapar, & McKoon, 2001, 2003, 2004; Ratcliff, Thapar, Gomez, & McKoon, 2004; Thapar, Ratcliff & McKoon, 2003).

To facilitate the visual evaluation of the quality of fits, fitting results are presented as Quantile Probability Functions (QPF; Ratcliff, 2001; Ratcliff & Smith, 2004). This presentation mode allows for the simultaneous observation of both RT distributions and accuracies and therefore is most suitable for our purpose. QPFs are created by plotting the 0.1, 0.3, 0.5, 0.7 & 0.9 RT quantiles on the Y axis and the response probabilities for each condition on the X axis. The result is 10 data points for each condition (5 for correct and 5 for errors). Equivalent RT quantiles of the simulated data are then connected across all response types and conditions by a line to represent the trends of changes in the form of the distribution across conditions (all 0.1 quantiles are connected by one line, all 0.3 quantiles by another line etc...). The empirical RT quantiles are represented by black Xs.

Since the models we examine here have all been previously tested in accounting for choice-RT data, all of them should be able to account for RT distributions and accuracies. We expect the differences between the models to be most evident in their ability to fit the dependency of mean RT on difficulty. Furthermore, to compensate for the asymmetric significance of conclusions based on successful fits versus those based on failed fits<sup>7</sup>, we will test whether the models that fail (qualitatively) to fit

---

7 Since a good fit to complex data is difficult to achieve a success is considered as strong support for the model. For the same reason, however, a poor fit can not be considered as strong support against a model. This is because, when dealing with the optimization of stochastic models to complex data, a poor fit can result from either a failure of the model, a failure of the optimization process or a failure of the modeler. To illustrate consider the following hypothetical example: Finding a needle in a random haystack is considered strong evidence for the a-priori hypothesis that needles exist in haystacks. Failing to find one, however, is not strong evidence against the existence of needles in haystacks because of the high probability of accidentally missing it (type 2 error).

complex data also fail to fit the main RT effect even under simplified, favorable conditions (i.e. fit only two data points: main RT effect and accuracy).

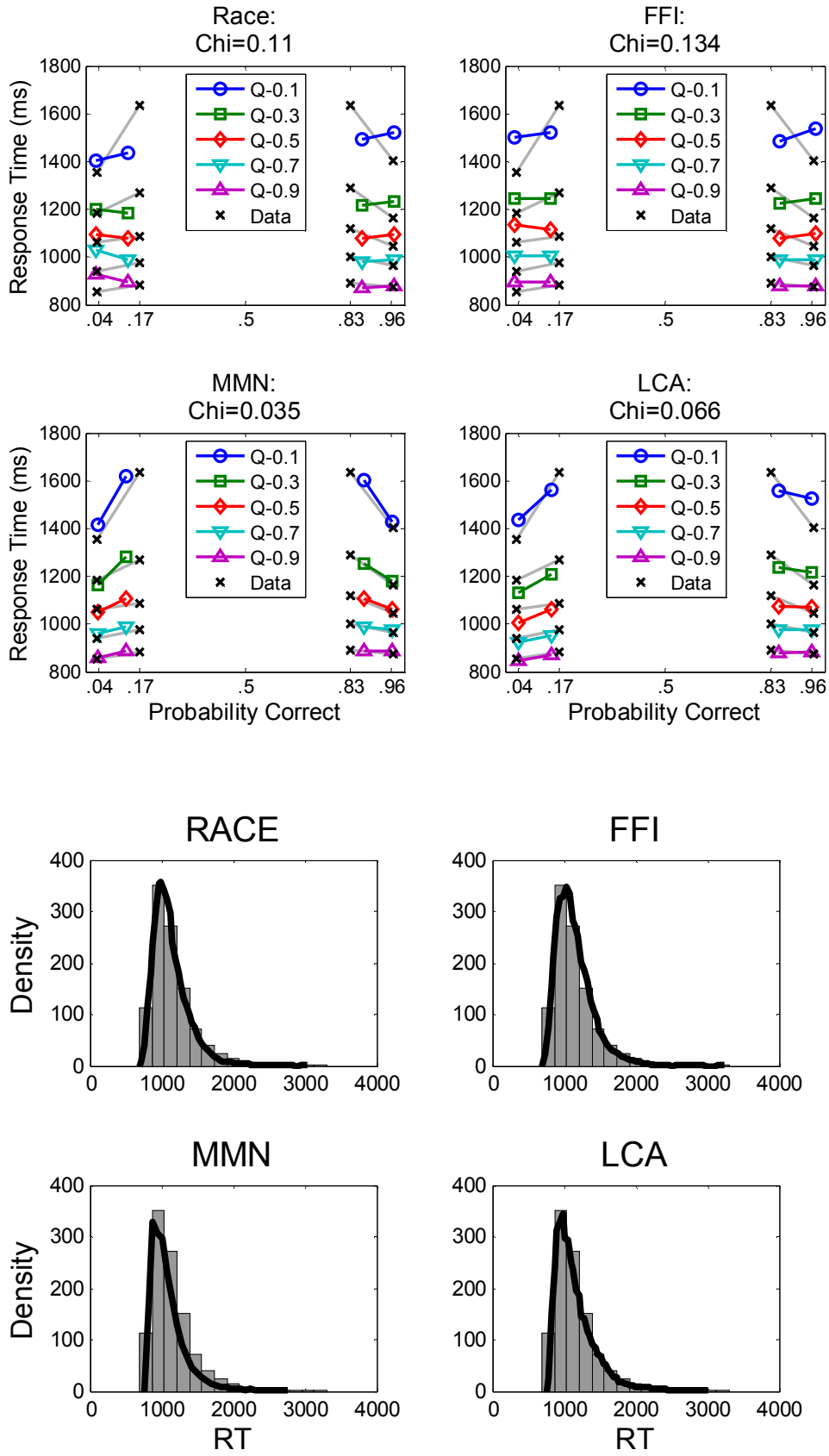
## Results

Best fitting parameters for the models are presented in Table 2. Figure 8 (top) shows QPFs of the model fits with the Xs representing the empirical data and the circles, squares, diamonds and triangles connected with lines representing the RT quantiles of simulated data for models with the best fitting parameters.

	Dv	Cr	$\sigma$	s	T <sub>nd</sub>	T <sub>s</sub>	Cv	$\lambda$	$\beta$	Chi
Race	0.042	19.047	0.364	5.215	417	12.05	7.023			0.103
FFI	0.001	6.73	0.52	0.865	585	18.336				0.135
MMN	0.013	3.313	0.422	0.915	726	14.64				0.035
LCA		0.844	0.189	0.154	749	28.593		0.391	0.391	0.066

**Table 2:** Best fitting parameters. Dv = drift rate; Cr = criterion;  $\sigma$  = internal noise; s = starting point variation; T<sub>nd</sub> = non-decision time; T<sub>s</sub> = step size; Cv = criterion variability;  $\beta$  = inhibition;  $\lambda$  = leak =  $\beta$ .

On the whole, all models were able to capture the general RT distribution in terms of spread and skewness (see Figure 8). In accordance with the qualitative predictions from the simulations presented in Figure 6, the race models and the FFI model were not able to capture the slow-down in the transition from the easy to the hard condition as was observed in the empirical data. An indication for this can be observed in the opposite slopes of gray (data) and colored (model) lines in Figure 8 (top panel). Note that, in terms of number of free parameters, these models had either just as many (FFI) or more parameters (race) than the response competition models. Although the MMN model achieved the best fit, both the LCA and MMN models were able to fit the data well despite small quantitative misses. For example, the LCA underestimated the amount of slow-down in the difficult condition while the MMN overestimated the accuracy in the difficult condition. While both input competition models we tested here (normalized race & FFI) failed to capture the qualitative pattern of the data, they did so not because they belong to the input competition class of models but rather because of the specific manner in which this competition is implemented (see discussion below).



**Figure 8:** QPFs (upper panel) and distribution fits (lower panel) for the race, FFI, MMN and LCA.

Several ways to improve model fits can be proposed, such as additional sources of between trial variability, different distributions of parameter values and psychophysical transformations of inputs. Except for the Race model that benefitted from variability in Cr, additional sources of variability such as variability in  $T_{nd}$  and variability in Cr did not improve the models chi scores significantly. For the race model, the specific choice of Cr distribution was important. The exponential distribution provided the best results with regards to the skewness of RT distribution. Different parameter distributions could also have similar effects on other models (but see Ratcliff, 2012 in press). However, such manipulations affect only the form of the distribution for all conditions simultaneously and cannot affect the pattern between conditions.

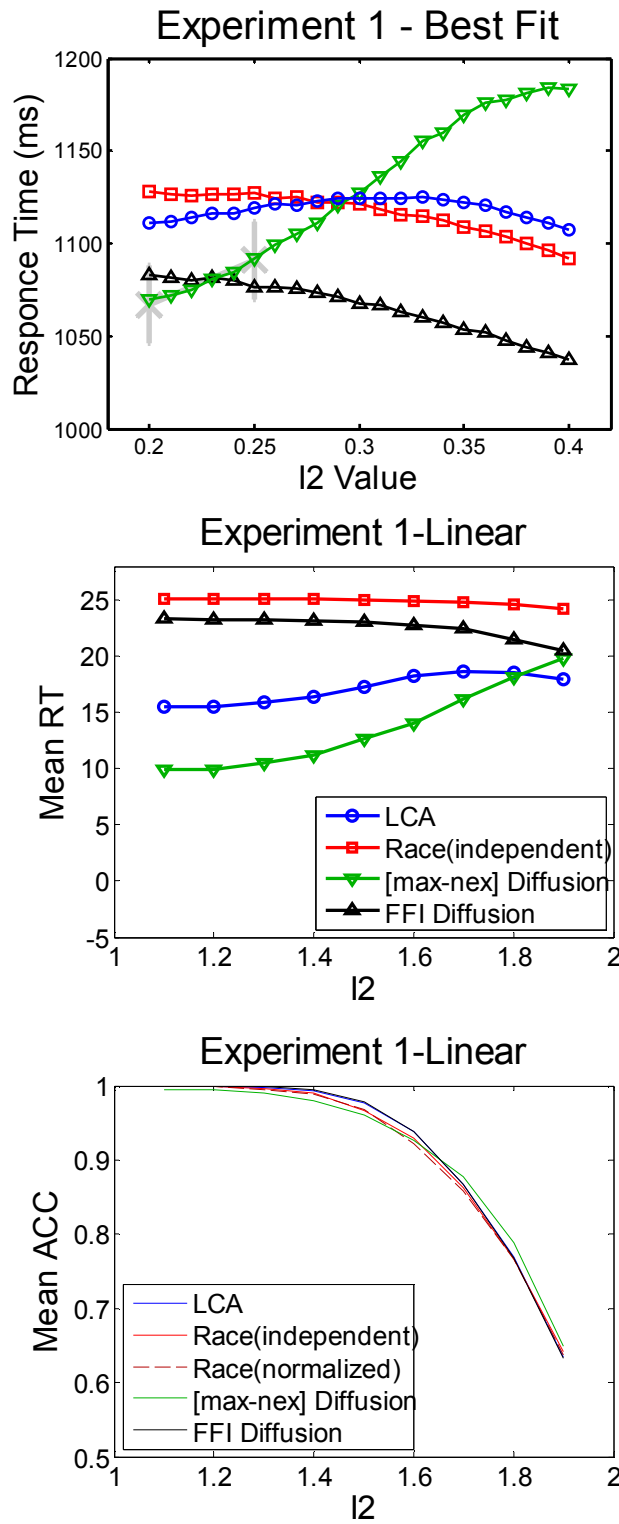
Another potentially beneficial manipulation, inspired by psychophysical research, is to transform the physical magnitudes into model-inputs via logarithmic or power law formulas before the accumulation stage. For example, with linear inputs the LCA underestimated the difference in accuracy between the easy and difficult conditions. This problem was mostly resolved when the inputs were transformed via a simple logarithmic formula<sup>8</sup>. Such a transformation (and, in fact, any monotonic, concave function) affects the larger inputs more than the smaller ones. Consequently, this reduces the relative difference between the target and strongest non-target ( $I_1$  &  $I_2$  respectively) thus reducing accuracy in the difficult condition<sup>9</sup>. However, as we demonstrate in appendix A, for our purpose, such non-linear psychophysical transformations can affect predictions *quantitatively* but do not alter the *qualitative* predictions of the models. For this reason and in the interest of clarity, we do not include all possible (model) x (transformation) combinations.

Each experiment in this study was designed with the purpose of bringing out the differences between model predictions. In the process of designing the experimental conditions, we used simulations to aid us in choosing the specific input manipulations that would produce the desired effects. To do so we had to first choose parameter values that would allow for a meaningful comparison between model predictions. Parameter values for a-priori simulations were chosen to produce a close match between the different models' accuracy predictions for all  $I_2$  values simultaneously. Figure 9 demonstrates how model predictions based on best fitting parameters (top panel) relate to predictions based on a-priori simulations based on accuracy matching (middle panel). The bottom panel in Figure 9 shows the accuracy predictions on which the choice of parameters for the a-priori simulations was based. While not perfectly identical *quantitatively* due to constant vertical shifts, the two graphs are nearly identical *qualitatively* with regard to the form or trend of the graphs for each model.

---

8 The fits of the LCA model in Figure A1 are of the model with log transformed inputs.

9 The MMN model has a similar miss in the fit presented in Figure A1. A logarithmic or power-law transformation could have also improved the fit of the MMN model. However, since the MMN model already achieved the best chi-square score with the least amount of free parameters there was no need for the added complexity.



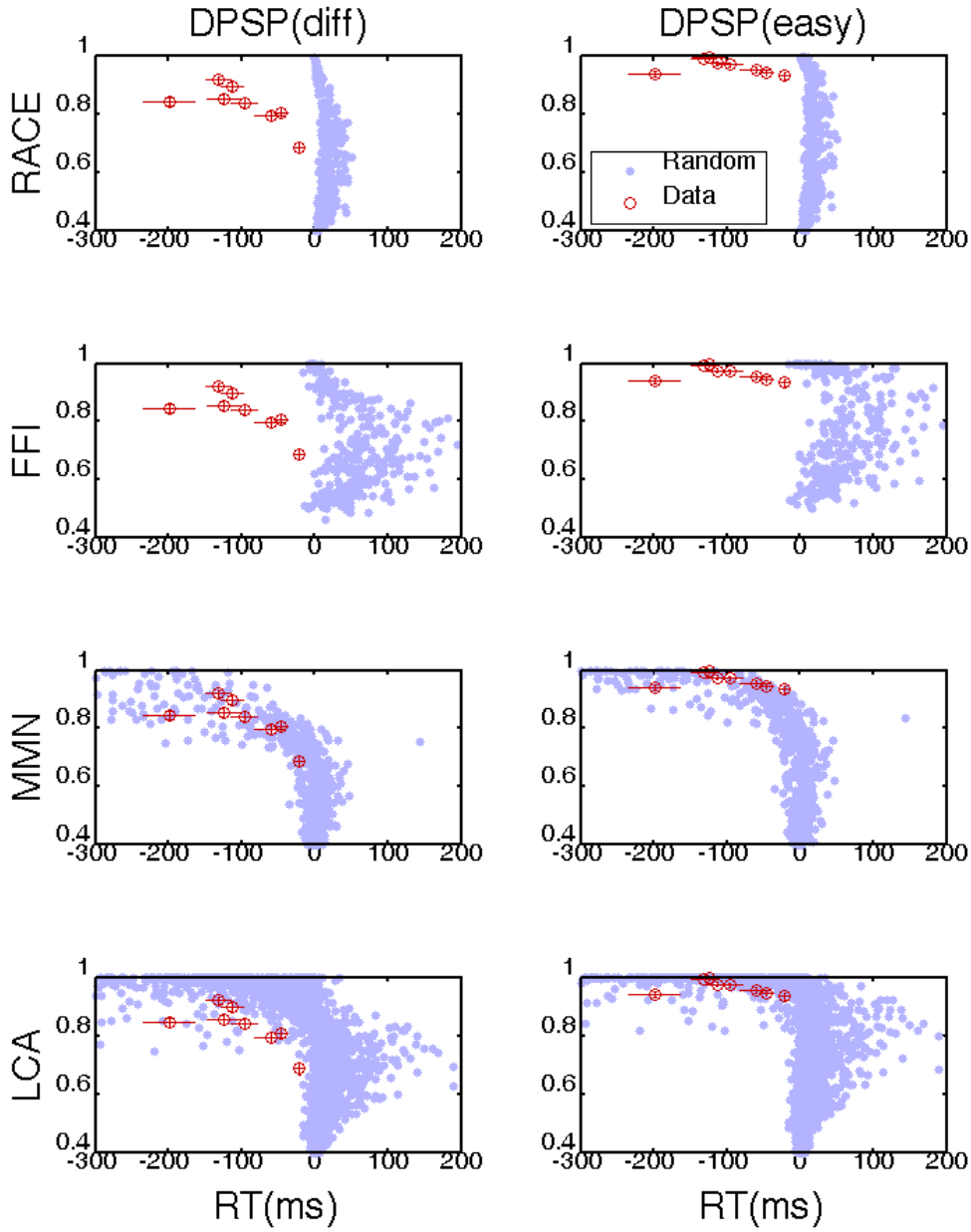
**Figure 9:** head to head comparison of post-hoc model predictions generated with best fitting parameters and a-priori predictions (middle panel) for models matched on accuracy (bottom panel).

### Contrasting Qualitative Predictions

So far, we derived distinguishable, qualitative, a-priori predictions and compared the predictions against empirical data generated with the relevant constraints in mind. By fitting the models to the data we both corroborate the validity of our, simulation based, a-priori predictions and put the models to the test to see how well they can accommodate the results. The last stage in this approach to model testing is to outline the boundaries of possible (and impossible) predictions for each model while comparing them to the data and its variability. This should enable us to see what other results can or cannot be accommodated by each of the models and, possibly, to derive additional predictions for future testing.

The main qualitative prediction of interest in this study concerns the difference in mean RT between various types of difficult and easy conditions where the target input is kept constant and the difficult condition is created by increased support to one of the non-target alternatives. To explore the boundaries of this effect we generate multiple predictions, for each of the models, by randomly sampling the parameter space from a uniform distribution around the best fitting set of parameters. This simplified method was chosen because a brute force, exhaustive grid search or a completely random sampling of the entire parameter space (which is not always well defined) would be too computationally intensive and would include many unreasonable parameter sets resulting in predictions that are pure noise. Presentation of model predictions in this way is reminiscent of two statistical methods used to evaluate relative complexities, Prior Predictive Distributions (PPD) and Response Surface Analysis (RSA). PPD is a Bayesian method that gauges complexity by comparing the Universal Interval (UI), which is the range of outcomes that could potentially be observed, irrespective of any model, to the Predicted Interval (PI) which contains all possible predictions for a specific model (Vanpaemel, 2009). RSA is a statistical method used to study relations among areas covered by responses generated from competing mathematical models (Bates & Watts, 1988; Pitt et al, 2002).

For each parameter set we calculated mean RTs and probability correct for both the easy and difficult conditions. Then for each model we generated two Delta Probability Scatter Plots (DPSP). In a DPSP each set of parameter values is represented by a point with  $\Delta = [\text{RT}(\text{easy}) - \text{RT}(\text{difficult})]$  as the x-value and probability correct as the y-value. The y-value can stand for probability correct in either the easy or difficult conditions hence the two DPSPs for each model (DPSPdiff & DPSPeasy). Figure 10 shows DPSPdiff and DPSPeasy pairs for each model.



**Figure 10:** DPSPs for the Race, FFI, MMN and LCA models. Left column displays combined model predictions for  $\Delta\text{RT} = \text{RT}(\text{easy}) - \text{RT}(\text{difficult})$  on the x axis and accuracy in the difficult condition on the y axis. Right column does the same for accuracy in the easy condition. Red circles with error bars represent individual subjects. Gray dots represent model predictions for randomly generated parameter sets. Black crosses represent model predictions for best parameters from different optimization runs that began from random starting points.

As can be seen in Figure 10 response competition models predict increasing slowdowns with higher accuracies. A more specific prediction is made by the LCA model, which predicts positive  $\Delta\text{RT}$ s with low accuracies and negative  $\Delta\text{RT}$ s with higher accuracies, a prediction that can be empirically tested. On the other hand, the independent race and FFI models are intrinsically incapable of slowing down under such input manipulations in direct contradiction to the data.

It is possible however that, due to the random parameter sampling method we used to generate the DPSPs in Figure 10, we may have missed some particular parameter combination that would have allowed the race and FFI models to reproduce the observed slowdown. Furthermore, it may be argued that due to the poorly-behaved nature of the error space and the complexity of the target data, the poor fits presented in Figure 8, could be due to a failure of the *modeler* (or the optimization algorithm), rather than a failure of the model. To circumvent these issues, we also fit the race and FFI to only two data points: the main effect RT difference between the easy and difficult conditions and accuracy in the difficult condition. If these models *can* somehow produce a slow down, then they should be able to fit two data points with 7 (race) and 6 (FFI) free parameters. Fitting results demonstrate that even under these highly favorable conditions, the independent race and the FFI models fail to reproduce the observed slowdown (see appendix C, Figure C1 ). This provides strong evidence that the failure of the race and FFI models is indeed due to attributes of the models and not to randomness or to the limitations of optimization algorithms.

### Experiment 1a: Discussion

This study demonstrated that both the independent and normalized race as well as the FFI model speedup in the difficult condition for *all* parameter values while response competition models are able to account well for the observed slowdown in the data. However, it might be possible to adjust some of the *peripheral* assumptions of the two input competition models to accommodate this pattern. For example,

changing the calculation of inhibition in the FFI from:  $\text{Inh}_i = \frac{1}{(n-1)} \sum_{j \neq i} I_j$  to:

$\text{Inh}_i = \max_{j \neq i} (I_j)$  or:  $\text{Inh}_i = \sqrt{\sum_{j \neq i} I_j^2}$  would cause the target input to be weaker in the

difficult condition consequently slowing the whole process down in accordance with the empirical findings. In the case of the normalized race, a more complex normalization function could possibly be found that would be able to produce negative  $\Delta$ s for this type of manipulation. For that to happen, the normalization would have to weigh the increase in  $I_2$  as more influential than the cumulative decreases in  $I_3$  &  $I_4$ . Such a weighting scheme, however, goes against the well established and empirically supported concept of diminishing sensitivities in most perceptual domains. Therefore, if applied it would probably have difficulties accounting for other behavioral effects. Although we did not test every possible independent model, our simulation results should generalize to all independent models. This assertion is based on the mathematical fact that the minimum of two independent random variables will always be lower then (or equal to) the minimum of each random variable separately. Therefore, faster independent racers will always finish faster than slower independent racers (see racer metaphor above).

In fitting the models to the data we tried to keep the models as constrained as possible in order to bring out their qualitative differences. This was achieved by keeping the input parameters locked directly to the physical stimuli. By not letting the input parameters vary freely either within a condition or between conditions we limit, to a considerable degree, the ability of the models to mimic each other. Despite the strong constraints on input parameters, response competition models perform surprisingly well with very few free parameters.



These results, demonstrate that such constrained, information based approach to model fitting is not beyond the capabilities of the models (see also Gao, Tortell & McClelland, 2011; Niwa & Ditterich, 2008; Palmer, Huk & Shadlen, 2005; Rorie, Gao, McClelland & Newsome, 2010; Tsetsos, Gao, McClelland & Usher, 2012). Furthermore, this approach holds important benefits with regards to the issue of model mimicry and can provide valuable insights into the mechanisms that underlie cognitive processes, such as, but not restricted to, decision making.

The remainder of the article focuses on experimental replications of the same type of manipulation: making the task more difficult by increasing the value of the non-target alternative while keeping the target constant. Except for Experiment 2 that does not involve a manipulation of input strengths, the basic manipulation remains the same throughout the remaining experiments and the predictions of the models also do not qualitatively change from those outlined in Figure 6 (top). Since Experiments 1b, 3a & 3b are meant to serve as replications and converging evidence for the ones obtained in Experiment 1a, we do not endeavor to fit the models to the data in these experiments. Except for some peculiarities of each experimental manipulation that will be discussed separately in the corresponding sections, model predictions are directly constrained by the architecture of the model and the design of the experimental manipulation. For illustration purposes and clarity, where necessary, model predictions are presented as simulation results for models matched on accuracy as in Figures 6 and 9. Experiment 2, however, involves a different type of manipulation, so data fits are carried out.

### **Experiment 1b – flicker task**

This experiment was designed as a replication of Experiment 1a with discreet flickering lights instead of continuously varying brightness patches. There are two basic motivations for doing so. First, a replication with a different type of stimuli provides converging evidence and demonstrated the robustness of the results. Second, we wished to eliminate alternative explanations to our findings from Experiment 1a that hinge on the concept that normalization can be applied to evidence coming from partial sets of stimuli. One such account is that subjects first rule out the two weak alternatives and only then proceeds to normalize the two remaining ones before accumulation. Under this assumption, a normalized (though not a completely independent) race model can capture the slowdown observed in Experiment 1a without difficulty. This is true since once the two weak alternatives are eliminated the two remaining alternatives are no longer intrinsically normalized. Therefore, in order to maintain normalization, an increase in evidence for the non-target alternative (as in the *difficult* condition) will lead to a decrease of evidence for the target alternative (see Figure 5 top). This will then result in an overall slowdown of mean RTs (as observed in the data).

Flickering light stimuli have several properties that make them especially attractive for testing decision processes (Vickers, Caudrey, & Wilson, 1971; Vickers, 1995) while at the same time excluding normalization as a plausible assumption. Perceptual normalization of stimuli intensities requires not only spatial proximity (Luck, et al., 1997; Moran & Desimone, 1985) but also temporal overlap. Stimuli that do not appear together on the screen are unlikely to be processed together and therefore are also unlikely to be normalized with respect to each other. In addition, flickering lights do not have different physical "intensities" per se since all discreet flickers share the same brightness and duration. From an information theory perspective, each flicker provides the same amount of information as any other

flicker. In fact, in a flickering lights paradigm information is carried mainly by the flicker's arrival times and not by its brightness or duration, which does not vary.

Still, one can think of flickers in terms of temporal "rates", which could theoretically be normalized. However, we used relatively slow flicker rates of 5.4fps (frames per second) to 1.2fps (i.e. 185 to 833ms mean intervals between consecutive flickers). Unlike brightness, which is instantaneously transformed into some amount of neural activation, rate magnitudes are encoded by averaging over time. To correctly evaluate the rate of discrete events arriving at stochastic time intervals, one needs at least two or three samples. In our case, for the fastest stimuli this would result in an absolute minimum rate evaluation time of  $185 \times 2 = 370\text{ms}$ . Therefore, in order to evaluate the flicker rate of each alternative, one must first integrate it over some considerable period of time. Importantly, if two rates are to be normalized with respect to each other then the minimum time step for normalization would have to be determined by the slower of the two alternatives (in our case  $333 \times 2 = 666\text{ms}$  minimum). This makes the mechanism of normalization difficult to apply at both the local and global input competition levels since both are exclusively dependent on momentary evidence. Taken together, these arguments support the notion that using flickering lights as perceptual inputs further reduces the plausibility and justifiability of assumptions relating to evidence normalization.

## Method

### Participants

Six Tel-Aviv University students, all female, participated in the study as part of their Introduction to Psychology course requirements. Each participant was tested in three to four sessions no more than four days apart. Each session was about 45 minutes long. All participants had normal or corrected to normal vision.

### Materials

All stimuli in this experiment were presented on a Samsung SyncMaster 943b LCD monitor. The stimuli were composed of four homogenous, round, white patches on a black background. The alternatives were positioned at the four corners of an imaginary square relative to a central plus sign fixation point (same as Experiment 1a). Each patch flickered randomly and independently of the other patches over the course of each trial. A flicker is defined as an abrupt transition from a gray level of 0 (black) to 1 (white) for one monitor refresh cycle (one frame) and back to 0. For the *easy* condition, the mean interval between consecutive 'on' frames was set to  $\text{Int}(1)=11.11$ ,  $\text{Int}(2-4)=33.33$  frames for the target and three non-targets respectively. For the *difficult* condition, mean intervals were set to  $\text{Int}(1)=11.11$ ,  $\text{Int}(2)=20$  and  $\text{Int}(3-4)=50$  frames for the target and three non-targets respectively. Mean flicker intervals were varied uniformly in a  $(-2, +2)$  range.

For the *easy* condition average rates were  $R(1)=5.4\text{fps}$  and  $R(2-4)=1.82\text{fps}$  and for the *difficult* condition average rates were  $R(1)=5.4\text{fps}$ ,  $R(2)=3\text{fps}$  and  $R(3-4)=1.2\text{fps}$ . The first 'on' frame of each alternative was determined randomly on each trial (uniform distribution between 1 and 10) to prevent repetitive circular (periodic) patterns and to discourage heuristics based on first onset. Refresh rate was set at 60Hz (16.6ms per frame) and tests were run to evaluate the probability of dropped frames. No frames were dropped after a full hour of continuous presentation. The location of the target was randomly drawn on each trial.

### Procedure

All trials were randomly assigned to one of two possible conditions: *easy* or *difficult*. Each trial began with a large fixation cross at the center of the screen which

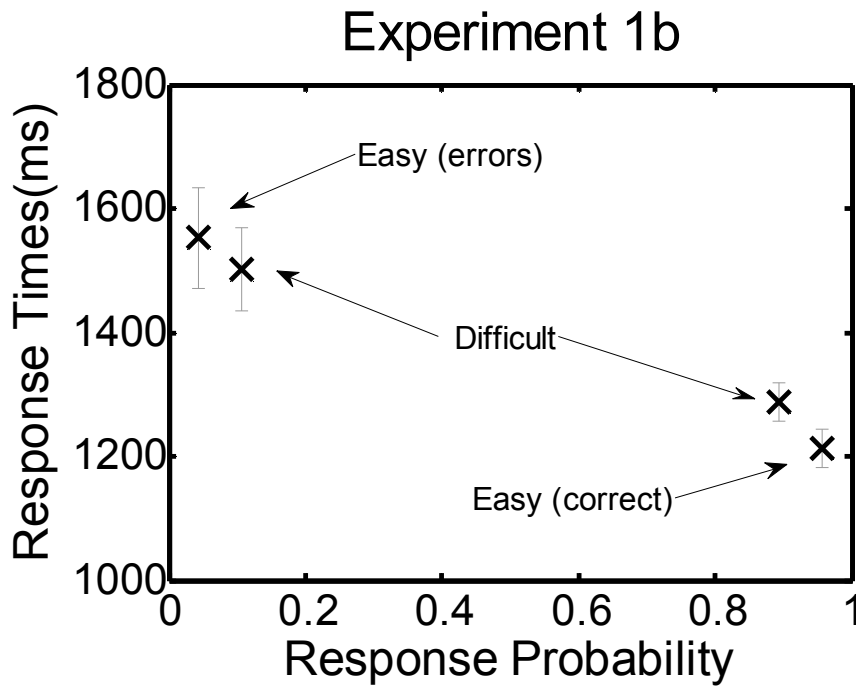
stayed on for one second then turned into a smaller one just as the target and distractors stimuli appear. Stimuli stayed on until the subject made his response. After the participants' response came a short 1sec ISI followed by the next trial. The fixation cross remained on the screen throughout the experiment but it briefly increased in size during each of the ISIs and came back to its original size at its end in order to draw attention back to the center of the screen before the beginning of the next trial.

Answers were given via the right number pad of a standard keyboard. The 1, 3, 7 and 9 keys represented the lower left, lower right, upper left and upper right patches respectively. Subjects were asked to place two fingers from the right hand on the 3 and 9 keys and two fingers from the left hand on the 1 and 7 keys. Participants were presented with blocks of 20 trials and each participant performed on average 1300 trials. Each block consisted of an equal number of trials from each condition. After each block there was a self timed intermission to allow the subject to rest his eyes and prepare for the next block. During each of these breaks the subjects were presented on the screen with their average accuracy and RT for the last block. Breaks were ended by pressing any key on the keyboard.

Subjects were instructed to respond to the alternative that, on average, flickers the most or, in other words, that has the fastest over all flickering rate. Subjects were also instructed to try and maximize both accuracy and response time such that if they reached 100% accuracy they should try to respond faster and were given a 30 trial practice block. Error feedback was given by an auditory tone. Participants were also instructed to keep their eyes focused on the fixation cross throughout the trial though in the absence of an eye tracker there was no way to verify that they actually complied with this request. The experiment was held in a partially darkened room.

## Results and Discussion

Participants performed less accurately in the *difficult* condition ( $M=0.87$ ,  $SD=0.05$ ) compared with the *easy* condition ( $M=0.95s$ ,  $SD=0.02$ ;  $z=2.2$ ,  $p<0.05$ ; Wilcoxon Matched Pairs Test). As predicted by competitive models an analysis of overall RTs revealed that participants responded more slowly in the *difficult* condition ( $M=1.31s$ ,  $SD=0.17$ ) than in the *easy* condition ( $M=1.22s$ ,  $SD=0.16$ ;  $z=2.2$ ,  $p<0.05$ ; Wilcoxon Matched Pairs Test). A separate analysis of correct and error responses revealed that correct RTs in the easy condition were faster ( $M=1.21s$ ,  $SD=0.16$ ) than in the difficult condition ( $M=1.29s$ ,  $SD=0.16$ ;  $Z=2.2$ ,  $p<0.05$ ). No significant effect was found between error RTs in the easy condition ( $M=1.55s$ ,  $SD=0.27$ ) and error RTs in the difficult condition ( $M=1.5s$ ,  $SD=0.27$ ;  $z=0.3$ ,  $p=0.75$ ). However, errors were very few due to the high accuracy. The results for overall and correct RT are consonant with experiment 1a and provide a replication of the same effects with a different type of stimulus.



**Figure 11:** Latency probability plots of results from Experiment 1b. Xs represent mean correct (right) and error (left) RTs for the two experimental conditions: *easy* and *difficult*. Error bars stand for two within subject standard errors. Lateral position on the x axis indicates the probability a specific type of response (correct>0.5; error<0.5) within that condition.

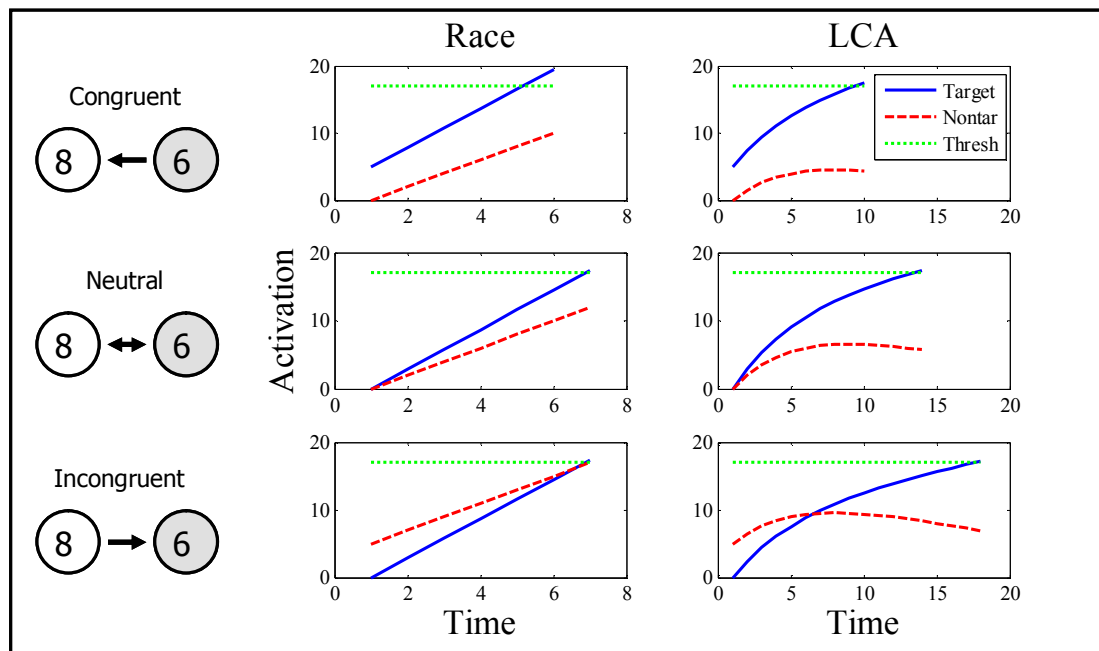
This replication demonstrates the robustness of the effect and provides converging evidence for our conclusions. Nevertheless, it is still possible that subjects do adopt a two stage strategy whereby the two slowest alternatives are eliminated first and then only the two remaining alternatives continue to compete for the determination of the response. However, as discussed before, the use of discreet, temporally separated flicker stimuli with relatively slow presentation rates renders the application of online input normalization impractical. As a result, even if a two stage strategy is utilized, the data is still difficult to account for within the framework of a normalized race model and impossible with a completely independent race model. In Experiment 3b, we will further examine this issue by replicating these results with a two alternative choice task where the alternative 'two stage' account does not apply.

Our second study was intended to further distinguish the predictions of independent and competitive models while also laying the ground for a paradigm that could begin to distinguish between input competition and response competition models in two alternative choice tasks and to provide certain constraints for input competition models. Here, priors, rather than input strengths, were manipulated by controlling observers' prior knowledge regarding the likelihood for the target's spatial location. As discussed above, response competition is a function of the *total* accumulated input and, as such, is affected not only by the momentary information but also by the priors in favor of each alternative (assumed to affect the starting point of the accumulation process; Bogacz et al., 2006; Gao, Tortell & McClelland, 2011; Gold & Shadlen, 2001, 2002; but see discussion below for alternative

interpretations). On the other hand, input competition is sensitive only to the momentary input which excludes the starting points. The next study thus enables us to further focus our attention on a narrower distinction between the different levels of competition: independent, input and response.

## Study 2: A Manipulation of Priors

Consider a two alternative choice task with the same type of stimuli as in experiment one, only now we introduce variation of the prior belief for each alternative via a 75% valid pre-cue (Posner, Snyder & Davidson, 1980), which can be *congruent*, *incongruent* or *neutral* (Figure 12, left). Such manipulations of prior knowledge have been studied in the context of decision making (Edwards, 1965; Summerfield & Koechlin, 2008) and found to have distinct effects on the decision process. Congruent cues tend to improve accuracy and reduce response times while incongruent cues sometimes do the opposite and sometimes selectively reduce accuracy without significantly affecting RTs (Forstmann et al., 2010). Prior information in favor of one alternative is operationalized here as an increase in the starting point of that alternative's respective accumulator (Bogacz et al., 2006; Gold & Shadlen, 2001, 2002). As was illustrated in the *scientists' competition* metaphor, a head start to one competitor slows down RTs for models with response competition (see also Figure 12, right). Contrary to this, models with input competition and independent models predict the opposite effect. For the independent race model, a head-start for the incorrect accumulator speeds up its finishing times (see Figure 12, middle). Consequently, this results in a speedup of correct RT's due to statistical facilitation. The same takes place in the FFI diffusion model since the head-start does not alter the value of the non-target alternative which could slow down the target accumulator via subtraction. Note that in the FFI model, unlike the MMN, the starting point of one accumulator does not affect the finishing time for the other accumulator.



**Figure 12: Middle and right:** Noiseless, illustration of the effects of prior knowledge (congruent/neutral/incongruent pre-cues) on the accumulation process. Prior knowledge was

implemented in the simulations as +5 to target accumulator/no change/+5 to non-target accumulator accordingly Red and blue lines trace activation trajectories (blue(solid): Target; red(dashed: Non-target) for the race (middle) and LCA (right) models as a function of time and the type of pre-cue. Green (dotted) lines indicate decision thresholds. In the *incongruent* condition the race model clearly does not slow down. The race model also does not seem to speedup in comparison to the *neutral* condition as our predictions dictate. However, this is just an artefact of the lack of noise in a single trial accumulation. Given some noise in the accumulation process, fast incorrect responses would cause mean correct RT to accelerate (statistical facilitation) giving us the predicted speedup effect. **Left:** Schematic illustration of stimuli (numbers indicate brightness values on a 1-10 scale) with the three possible pre-cues.

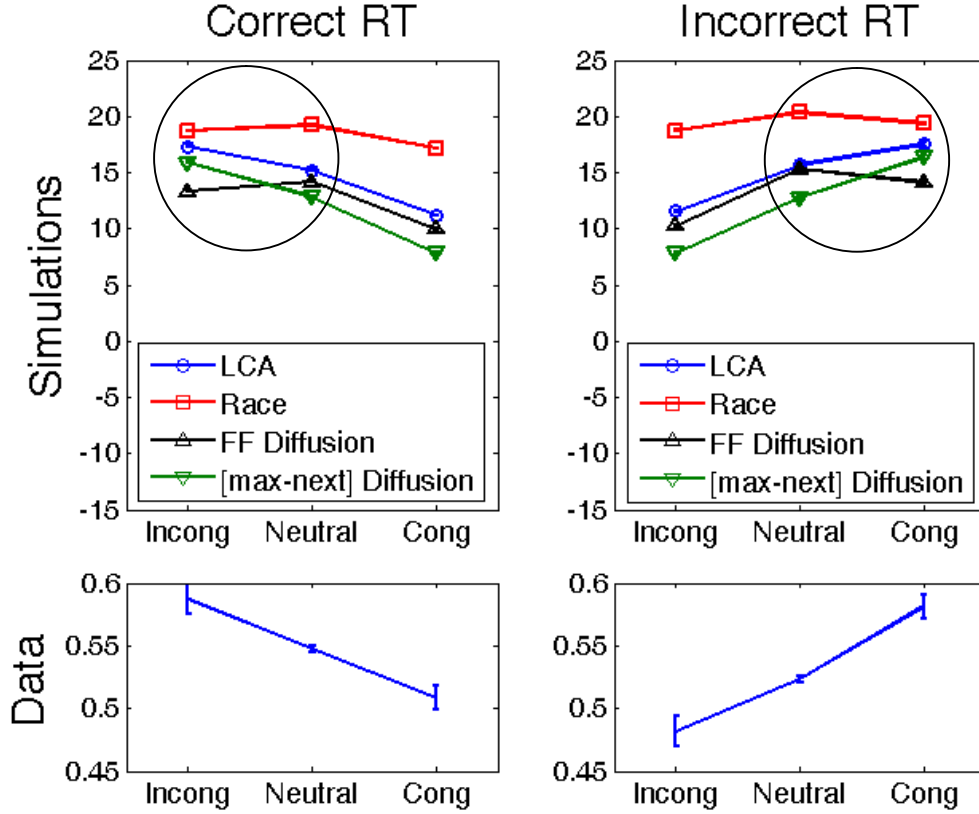
The effect of prior information on the starting conditions of the decision process can be operationalized in several ways. While this issue is still under debate and may be sensitive to the specific choice of experimental manipulation and stimuli, we chose to represent prior knowledge as an increase to the starting point of the cued alternative without a simultaneous change to the starting point of the un-cued alternative (un-correlated starting points assumption). This choice is supported by observed RT distributions and data fits (see Experiment 2 discussion below). A second reason for using this assumption is its parsimony and natural mapping to neural activity.<sup>10</sup> Some alternative accounts of prior knowledge, such as changes in drift rate and anti-correlated starting points, can invalidate part (or all) of the qualitative predictions outlined above. For example, changes in drift rate would allow input competition (but not independent) models to slow down following an incongruent cue while anti-correlated starting points would also allow independent models to slowdown. Fortunately, the drift rate account makes clearly distinguishable predictions regarding RT distributions and will therefore be evaluated directly from the data. Other accounts, such as anti-correlated changes in starting points, however, are difficult to tell apart from un-correlated changes in starting points just by looking at the data. Therefore, such accounts will be explored quantitatively via data fits and qualitatively by mapping prediction spaces to derive constraints for the various models.

Simulations were run to evaluate the effect of a manipulation of priors on the mean RT of correct and incorrect responses. The same five models were used: a purely independent (*race*) model (red line), and four competitive models: a normalized race model (red line<sup>11</sup>), MMN (green line), *FFI diffusion* (black line) and *LCA* (blue line; see Figure 13). As can be seen in Figure 13 (top), for both the correct and incorrect responses the race and the FFI diffusion models predict faster response times for all conditions relative to the neutral condition. For the correct responses, the LCA, as well as the MMN predict a speedup in the congruent (easy) condition relative to *neutral* and a slowdown for the *incongruent* (hard) condition (Figure 13, top left). For these two models exactly the opposite pattern is observed for the incorrect responses (Figure 13, top right). Thus the critical comparisons for this

<sup>10</sup> Accordingly, the cue acts as a brief input that increases the activation of the cued-alternative unit.

<sup>11</sup> Note that, due to the specific choice of constant inputs, the independent race model and the normalized race model make equivalent predictions for this manipulation. Therefore both are represented by the same line color (red) in the figure.

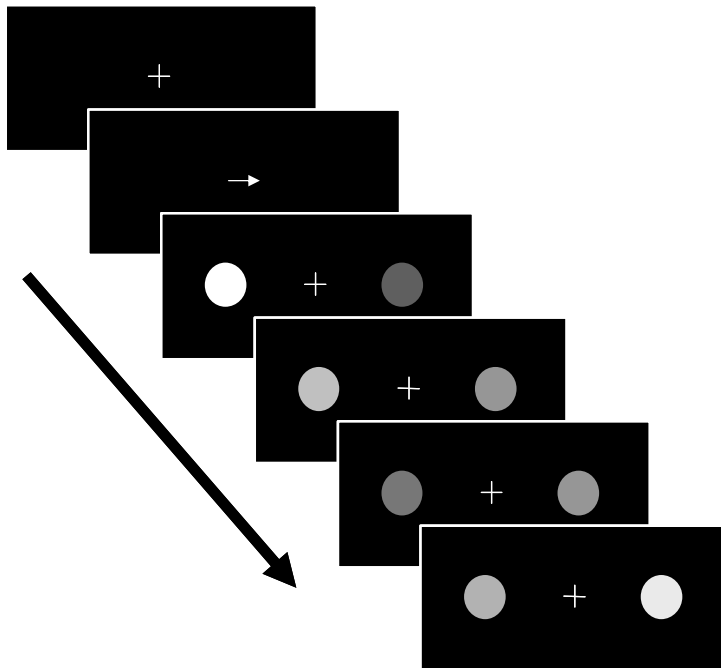
simulation are *neutral* vs. *incongruent* for the correct responses and *neutral* vs. *congruent* for the incorrect responses, where input-competition and response-competition models make contradicting predictions (see Figure 13, top). Experiment 2 was designed to test these predictions.



**Figure 13:** Mean-RT for correct responses (left) and incorrect responses (right), in the 3 conditions (*congruent/neutral/incongruent*) of Ex.2. **Top:** Simulations for the various models with starting points differing according to condition (see Figure 12 for illustration). Decision criteria were set so that all models predicted approximately the same accuracies for the neutral condition (Race: 40; LCA: 14; MMN: 9; FFI: 13). Input strengths remained constant throughout the simulations ( $I_1 = 2$ ,  $I_2 = 1.5$ ). For the *neutral* trials the two accumulators started from the same starting point ( $SP_i = 1$ ). Before the initiation of simulated *congruent/incongruent* trials, the starting point of the target/non-target accumulator, respectively, was increased by 5 ( $SP_i = 6$ ). All models were modeled according to the equations described in the introduction ( $\mu = 2$ ;  $\xi = \beta = 0.1$ ). **Bottom:** RT data [sec.] from Ex.2. Error bars correspond to within-subject variance (Cousineau, 2005).

## Experiment 2

In this experiment the participants were asked to decide, as fast and as accurately as they can, which of two fluctuating gray patches is the brightest. This time, however, each trial was preceded either by an arrow indicating where the brightest patch is most likely to appear (75% validity) or by a two headed arrow that provided no predictive information (see Figure 12, left; Figure 14). Importantly, brightness levels remained constant throughout the experiment and the only difference between the conditions was in the priors.



**Figure 14:** Illustration of the time course of Experiment 2. The cue appeared 1500ms before stimulus onset and remained on for 1000ms. Then, the fixation cross returned for another 500ms after which the brightness patches appeared and remained on until response.

## Method

### Participants

Ten TAU students (7 female) participated in the study in exchange for course credits. Participants were tested in two, 45 minute long sessions (no more than four days apart). One participant (female) was excluded from the final analysis due to chance level performance. All participants had normal or corrected to normal vision.

### Materials

The stimuli were composed of two homogenous, round, gray patches on a black background (width: 1.2cm) that were positioned to the right and to the left of the fixation point (total width from right edge of right patch to left edge of left patch: 5cm). On all trials gray levels of the target and non-target were normally distributed around means of 0.8 and 0.6 (on a 0 (black) to 1 (white) scale) respectively with a SD of 0.2. All trials were preceded by a white arrow that replaced the fixation cross for 500ms before stimulus onset. The arrow pointed either towards the right, left or in both directions. Except for the above mentioned details and the fact that there were only two patches instead of four, experiment 2 was identical to experiment 1.

### Procedure

All trials were randomly assigned to one of three possible conditions: congruent, *incongruent* and *neutral*. Participants were presented with blocks of 50 trials for a total of 1600 trials per participant. Each block consisted of 60%, 20% and 20% of *congruent*, *incongruent* and *neutral* trials respectively, making the arrow 75% valid (60 out of 80 non-neutral trials). Participants were informed that the arrow preceding each trial was a good, though not perfect, predictor of the targets' location. All other procedures were identical to experiment 1.



## Results

Participants were less accurate when the arrow pointed in the wrong direction (*incongruent* condition;  $M=0.84$ ,  $SD=0.07$ ) than when it pointed in both directions (*neutral* condition;  $M=0.91s$ ,  $SD=0.02$ ,  $z=2.55$ ,  $p<0.05$ ; Wilcoxon Matched Pairs Test) and were more accurate when the arrow did point in the right direction (*congruent* condition;  $M=0.95$ ,  $SD=0.02$ ,  $z=2.55$ ,  $p<0.05$ ; Wilcoxon Matched Pairs Test). More importantly, participants responded more slowly in the *incongruent* condition ( $M=0.57s$ ,  $SD=0.09s$ ) than in the *neutral* condition ( $M=0.55s$ ,  $SD=0.08s$ ;  $z=2.66$ ,  $p<0.01$ ) and faster in the *congruent* condition ( $M=0.52s$ ,  $SD=0.08s$ ;  $z=2.07$ ,  $p<0.05$ ). Furthermore, a separate analysis of correct and error responses revealed that participants were slower to correctly respond in the *incongruent* condition ( $M=0.58s$ ,  $SD=0.09s$ ) than in the *neutral* condition ( $M=0.55s$ ,  $SD=0.08s$ ,  $z=2.66$ ,  $p<0.01$ ; Wilcoxon Matched Pairs Test) and faster in the *congruent* condition ( $M=0.52s$ ,  $SD=0.08s$ ;  $z=2.19$ ,  $p<0.05$ ). Furthermore, for error trials, participants were slower to respond in the *congruent* condition ( $M=0.58s$ ,  $SD=0.09s$ ) than in the *neutral* condition ( $M=0.53s$ ,  $SD=0.1s$ ,  $z=2.66$ ,  $p<0.01$ ; Wilcoxon Matched Pairs Test) while also marginally faster in the *incongruent* condition ( $M=0.5s$ ,  $SD=0.1$ ;  $z=1.84$ ,  $p=0.07$ ).

We also tested for differences between Leading Edge (LE) and tail effects of the congruent and incongruent conditions. The leading edge effect for a specific condition is defined as the difference between the RT of the 0.1 quantile in the neutral condition and the 0.1 quantile in the test condition. The tail effect is defined in a similar manner for the 0.9 quantile. No significant differences were found between either the LE ( $M=0.061s$ ,  $SD=0.084$ ) and tail ( $M=0.051$ ,  $SD=0.028$ ;  $z=0.65$ ,  $p=0.5$ ; Wilcoxon Matched Pairs Test) of correct congruent trials, LE ( $M=0.101s$ ,  $SD=0.110$ ) and tail ( $M=0.1s$ ,  $SD=0.056$ ;  $z=0.77$ ,  $p=0.44$ ; Wilcoxon Matched Pairs Test) of correct incongruent trials, LE ( $M=-0.061s$ ,  $SD=0.059$ ) and tail ( $M=-0.055s$ ,  $SD=0.096$ ;  $z=0.18$ ,  $p=0.86$ ; Wilcoxon Matched Pairs Test) of incorrect congruent trials and LE ( $M=-0.105$ ,  $SD=0.109$ ) and tail ( $M=-0.113s$ ,  $SD=0.171$ ;  $z=0.18$ ,  $p=0.86$ ; Wilcoxon Matched Pairs Test) of incorrect incongruent trials.

## Fitting the models to the data

To test the capabilities of the models to account for additional aspects of the data we fit the models to the data from Experiment 2. Fitting methods were similar to Experiment 1a. The effect of the cue in the congruent, neutral and incongruent conditions was modeled as an increase in starting point for the cued accumulator. The increase parameter ( $s_0$ ) was assumed to remain constant between conditions. It was always set to zero in the neutral condition and was equal in the congruent and incongruent conditions with the difference between them manifesting only in the type of cued alternative (i.e. the target or non-target accumulator accordingly). Inputs were held constant across conditions with their mean and variance identical to the ones used in the experiment. Thus there were no free parameters that varied between conditions. Finally, to test the assumption of anti-correlated bias in starting points we also fit the Race and FFI models with starting points calculated as  $X_0 + s_0/2$  &  $X_0 - s_0/2$  for the cued and un-cued alternatives, respectively (see appendix C for illustrations)

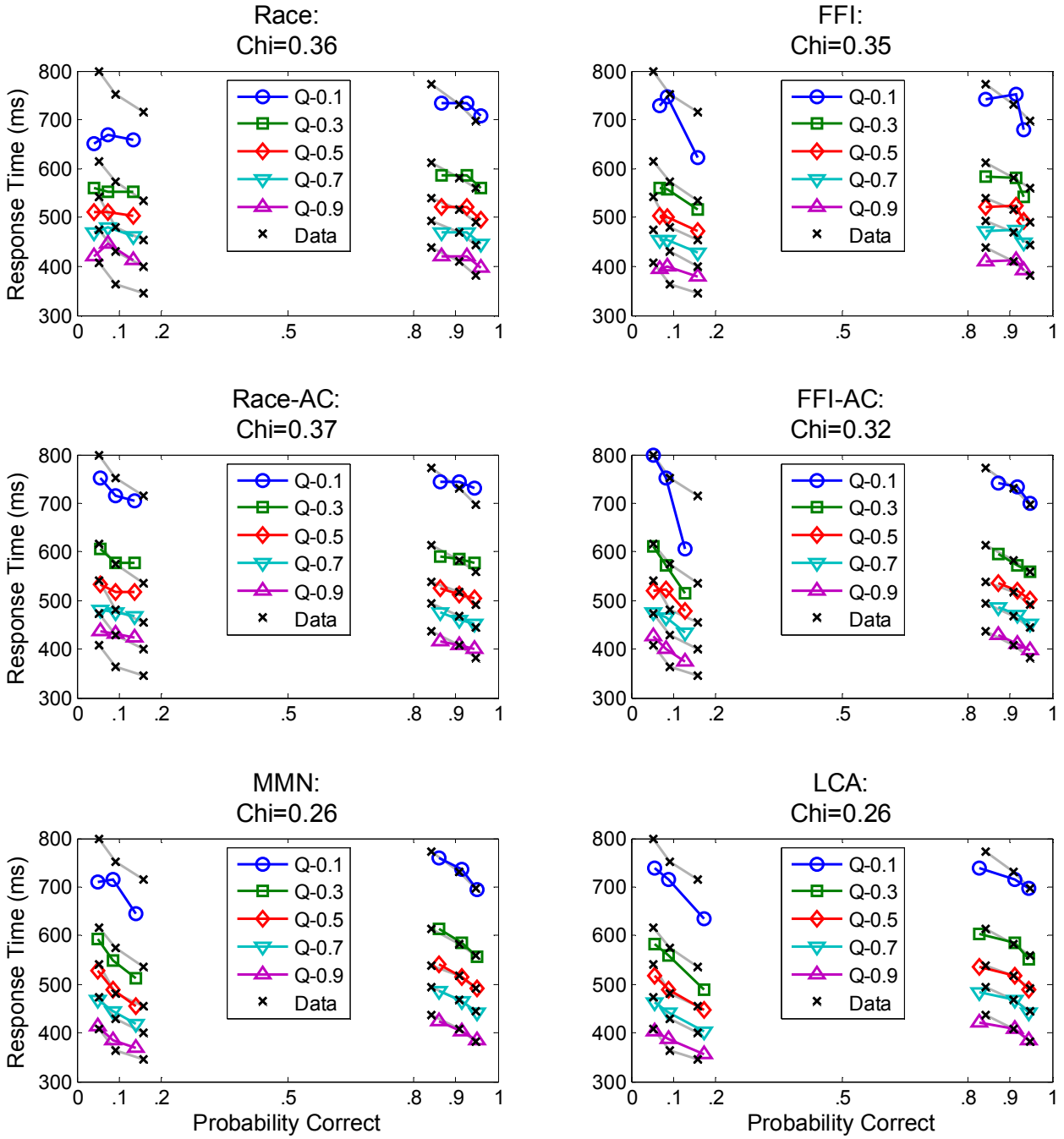
## Results

Fitting results are presented in Figure 15 and best fitting parameters are presented in Table 3 together with Chi-square scores. All models accounted well for the correct RT distributions. However, all models somewhat underestimated the spread of the distribution of error RTs. While this underestimation is small for the MMN and LCA models, it is more pronounced for the FFI model and largest for the Race model. More importantly, both independent and input competition models (independent race, normalized race and FFI) with uncorrelated starting points failed to produce slower than neutral correct RTs for incongruent trials and slower than neutral error RTs for congruent trials.

Changing the prior to starting point assumption from uncorrelated to anti-correlated, enabled the independent Race and FFI to predict slower correct RTs in the incongruent condition. However, for the race model the RT-slowdown was still smaller than the one observed in the data and also came at the expense of underestimating the RT speedup in the congruent condition (shallower slopes), without any improvement in the Chi-Square goodness of fit. For the FFI model the Chi-square scores improved slightly (see Figure 15), but were not as good as those of response competition models (although the number of parameters is larger than MMN and equal to LCA).

	Dv	Cr	$\sigma$	s	$T_{nd}$	$T_s$	Cv	$\lambda$	$\beta$	$s_0$	$T_{nd}Var$	Chi
Race	0.02	35.58	0.511	10.072	109	8.221	11.759			2.221		0.3553
FFI	0.06	7.139	0.919	4.746	409	4.06	3.424			1.704	176	0.3453
MMN	0.02	10.49	0.828	8.36	375	3.332				2.05	145	0.261
LCA		6.635	0.79	4.625	318	2.762		0.149	0.173	1.727	153	0.26

**Table 3:** Best fitting parameters for Experiment 2. Dv = drift rate; Cr = criterion;  $\sigma$  = internal noise; s = starting point variation;  $T_{nd}$  = non-decision time;  $T_s$  = step size; Cv = criterion variability;  $\beta$  = inhibition;  $\lambda$  = leak;  $s_0$  = change to starting point of cued alternative;  $T_{nd}Var$  = non-decision time variance.

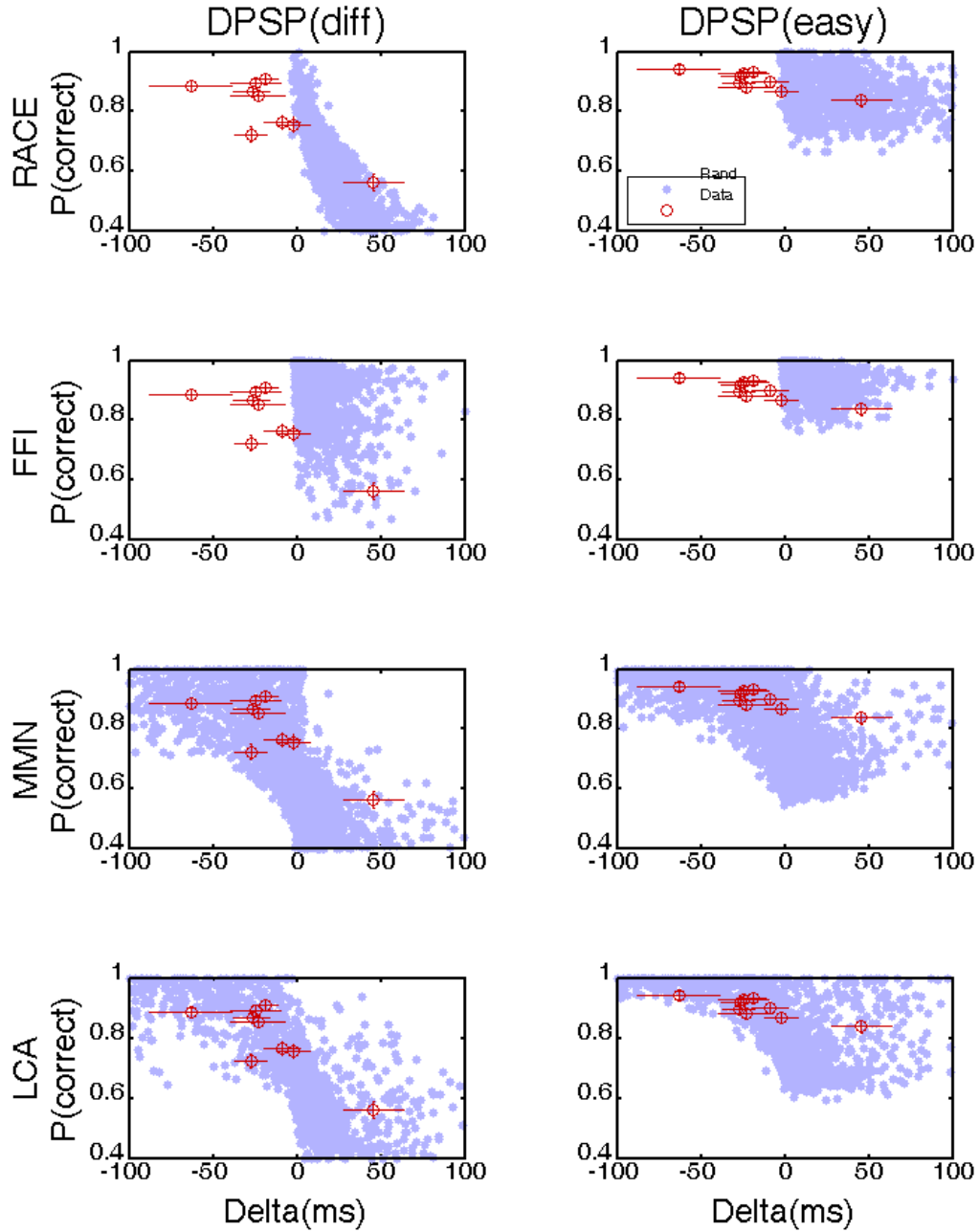


**Figure 15:** QPFs for race, FFI , race with anti-correlated starting points, FFI with anti-correlated starting points, MMN and LCA models. Xs linked by gray lines represent the data. Color lines represent model predictions for best fitting parameters.

### Contrasting Qualitative Predictions

We start with the qualitative predictions for the four models under the uncorrelated starting point assumption. The distinctive qualitative prediction involves the difference in *overall* RTs between the neutral and incongruent conditions. To test the models' ability to account for this main effect (as observed in the data) while also exploring the predictions of the models for a wide range of parameter combinations

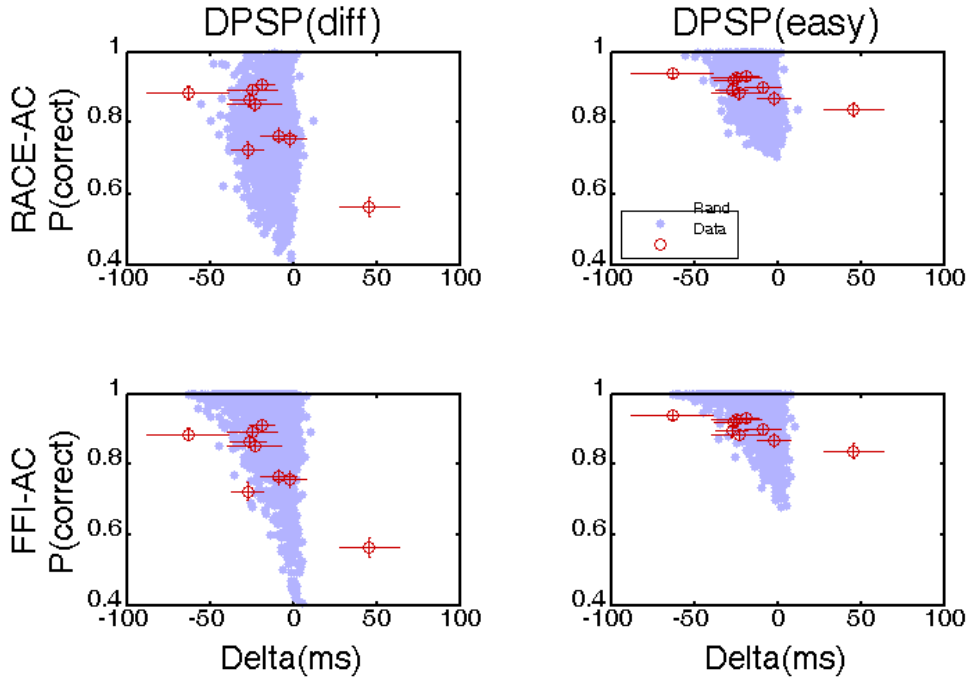
we generated random parameter combinations for each model and produced DPSPs. Figure 16 shows the range of possible predictions for each model with regards to the relation between accuracy and  $\Delta = RT(\text{neutral}) - RT(\text{incongruent})$ .



**Figure 16a:** DPSPs for the Race, FFI, MMN and LCA models. Left column displays combined model predictions for  $\Delta = RT(\text{neutral}) - RT(\text{incongruent})$  on the x axis and accuracy in the incongruent condition on the y axis. Right column does the same for accuracy in the neutral condition. Red circles with error bars represent individual subjects. Gray dots represent model predictions for randomly generated parameter sets. Black crosses represent model predictions for best parameters from different optimization runs that began from random starting points similar to the gray dots.

As one can see, independent and input competition models predict positive deltas for *all* parameter combinations and thus fail to capture the (negative) main effect observed for nine out of the ten participants. As before, to verify that this effect, as well as the poor fits presented in Figure 15, is not due to randomness or limitations of optimization procedures, we also fit the race and FFI models to only two data points: main effect of RT difference between neutral and incongruent conditions and accuracy in the incongruent condition. Results show that even under these favorable conditions the race and FFI models still can not account for the observed slowdown in RT (see Appendix C). On the other hand, response competition models predict a combination of negative effects for high accuracies and positive effects for low accuracies, thus accounting for all ten participants including the outlier that performed at chance level in the incongruent condition. Although response competition models show greater flexibility than other types of models they still cannot predict *any* pattern of results. For example, positive deltas with high accuracy are well within the predictions of input competition and independent models while clearly beyond the capabilities of response competition models. A combination of positive effects for high accuracies and negative effects for low accuracies can not be captured either by response competition, input competition or independent models.

We turn now to the qualitative predictions for the race and FFI-models with anti-correlated starting points. As shown in Figure 16b, with this additional assumption, both models are able to predict RT slowdowns in the incongruent condition. These predictions, however, are still distinct from the ones made by response competition models with uncorrelated starting points. The assumption of anti-correlation seems to restrict the models' predictions exclusively to slowdowns. This is in contrast to response competition models that predict slowdowns for high accuracies but speedups for low accuracies. In our study there was only one subject that had low accuracy (close to chance in the incongruent condition but around 80% in the neutral condition) and this subject conformed to the prediction of response competition models and was outside the reach of the independent race and FFI models with anti-correlated starting points.



**Figure 16b:** DPSPs for the Race and FFI models with Anti-Correlated (AC) starting points. All other details are the same as Figure 16a.

## Experiment 2: Discussion

The results of Experiment 2 demonstrate that subjects slow down in response to an incongruent cue. Under the assumptions that priors selectively affect the starting point of the cued alternative this effect is incompatible with both independent and input competition models. The race model (either independent or normalized) and the FFI diffusion model predict opposite patterns to the ones observed in the data since they are only affected by the head start of the incorrect accumulators without any corresponding slow down in the correct-RT accumulator and are thus subject to statistical facilitation. On the other hand, the results are consistent with the predictions of response competition models represented by the LCA and the MMN. These response competition models are able to account for the observed patterns precisely because the competition they utilize is introduced only at the level of the integrated input. This type of competition corresponds to the total amount of accumulated information and includes the priors. Therefore, response competition leads to a slowing down of correct responses to incorrectly cued displays and also of incorrect responses to correctly cued displays. Since accuracy was fairly high, this sums up to an overall RT slow down in the incongruent condition. Consequently, by taking longer to respond the response competition models allow for more information to be integrated thus preventing accuracy from suffering too much.

Under the alternative assumption that priors generate anti-correlated shifts in the starting points of the two accumulators (Forstmann et al., 2010), qualitative distinctions between the independent, the input and response-competition model are less clear. However, our quantitative evaluation of the race model fits to data, indicates that this assumption still does not improve the quality of fit for this model, which remains lower than for the competitive models. For the FFI-model, the fits with anti-correlated starting points are better able to account for the effect of prior manipulation, resulting in a slight improvement in the quality of fit over the non-

correlated version. One distinctive prediction that can be derived from the DPSP figures (16a & 16b) relates to the interaction of accuracy with the main RT effect (the RT difference between neutral and incongruent conditions). Independent and input competition models predict either only speedups (with un-correlated starting points) or only slowdowns (with anti-correlated starting points), while response competition models predict slowdowns for high accuracies and speedups for low accuracies. In this regard, the results of Experiment 2 provide some evidence in favor of response competition models. This is because one of our subjects that had low accuracies also displayed a speedup in RT, supporting the hypothesis predicting a reversal in the main RT effect as a function of accuracy. Further testing with various accuracy instructions, however, is needed to fully establish this result.

A second alternative assumption about how priors affect the parameters of the choice process is that prior knowledge affects the gain (drift rate) rather than the starting point of the accumulation process (Diederich & Busemeyer, 2006; Ratcliff, Van Zandt, & McKoon, 1999; Ravenzwaaij, Mulder, Tuerlinckx & Wagenmakers, in press). So far, fits to empirical data tend to support the hypothesis that starting points rather than drift rates are affected by such manipulations of prior knowledge (Mulder et al., in press; Simen, Contreras, Buck, Hu, Holmes & Cohen, 2009; but see Ratcliff et al., 1999; Ravenzwaaij et al., in press). The two hypotheses can be further distinguished by their qualitatively different predictions regarding RT distributions. The effect of a manipulation of starting points on RT distributions is characterized by equal shifts in both the leading edge and tail and has an opposite affect on error RT than on correct RT. The effect of a manipulation of drift rates, on the other hand is characterized by small shifts in the leading edge and large shifts in the tail of the RT distribution and affects correct and error RT in the same manner (Mulder et al., in press). The results of Experiment 2 also shed light on this issue. Since no significant differences were found between leading edge and tail effects, and since the manipulation had an opposite effect on error vs. correct RT, we can consider the data as supporting the account of biased starting points over biased drift rates (see also Figure 15 for graphical illustration).

The results we presented so far provide some support for response competition choice models over independent and input-competition models. In our next study we examine in more detail the difference between independent, response competition and one type of input competition (the normalized race model; Figure 3 (right)) by using both a more direct manipulation of the starting point, which is free of the correspondence assumptions between priors and starting points (Experiment 3a) and a replication of the difficulty manipulation from Experiment 1 for two alternative choice (Experiment 3b). These manipulations do not distinguish between response competition and some input competition models, such as the FFI diffusion, and we defer further comparisons between these models to the General Discussion.

### Study 3: Probing for Normalization

The state of each accumulator at the beginning of a trial can be manipulated, without resorting to assumptions about the representation of priors, by independently controlling the initial values of each alternative. Let us consider, for example, the *incongruent* condition from experiment 2, where the starting point of the incorrect accumulator was higher than that of the correct accumulator (see Figure 12). Instead of using priors to produce these starting activations, we can use a temporal

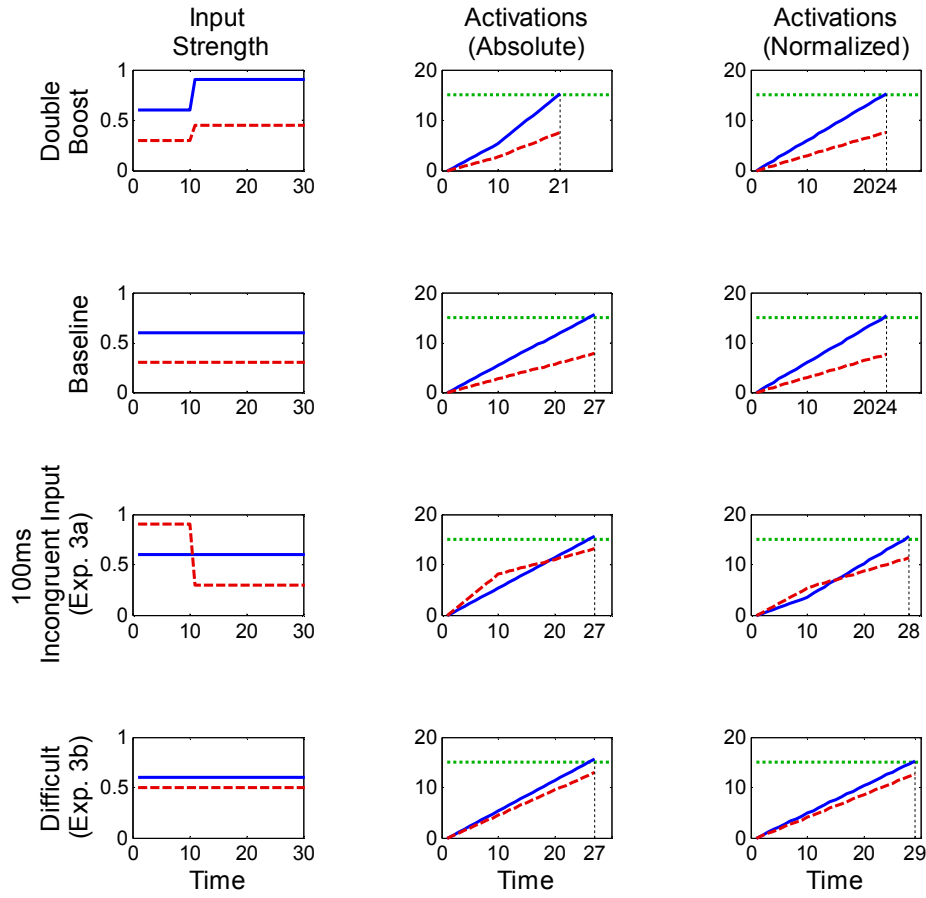
manipulation of brightness that briefly favors the non-target alternative before returning to its baseline values (Figure 17 third row, left; see also Brown & Heathcote, 2005, for an experiment using a similar manipulation). Thus, in Experiment 3a, the non-target stimulus will have a higher mean brightness than the target stimulus ( $I_2' = I_1 + (I_1 - I_2)$ ) for the first  $n$  frames ( $2n \ll \text{mean RT}$ ) of the trial (we name this the *100ms incongruent input* condition since  $n=100\text{ms}$  in our experiment). After  $2n$  frames the accumulated input for the target will overcome that of the non-target rendering it the correct response and returning the balance of evidence to its correct state (see Figure 17, third row, middle, for an illustration). However, from the point of view of an independent model, after the first  $n$  frames we are effectively at a state in the accumulation process that is equivalent to the beginning of an incongruent trial in experiment 2 (at the time of stimulus onset and just after the presentation of the pre-cue arrow). At this point the integrated value will slightly favor the wrong alternative, as in the incongruent prior condition in experiment 2. Importantly, when compared to a baseline condition with exactly the same input for the target alternative but with a *constant* lower non-target input (lacking the first few frames of high input; Figure 17, second row), this manipulation will produce a speedup in RT under an independent race framework (Figure 17, middle column<sup>12</sup>) and a slowdown for the other competitive models such as the normalized race model (Figure 17, right column; other competitive models not shown in Figure 17, but see Figure 12 (right) for similar illustration of LCA).

Another way to distinguish between independent and competitive models we implement in Experiment 3b, is to compare the RT in the baseline condition to a more difficult condition where the target brightness is the same as in the baseline condition but the non-target brightness level is higher (see bottom row of Figure 17). This is in fact a replication of experiment 1a for a two alternative task, and will produce the same qualitative predictions. Here independent models will predict faster RTs in the difficult condition while competitive models (including the normalized race model) will predict slower RTs (see "Qualitative Predictions" section for the rationale behind this). While the procedures described above can distinguish between independent and competitive models, they do not distinguish between normalized input competition models, such as the race with normalized input (Figure 17, right column), and other types of competitive models. In such models any increase in non-target input also negatively influences the activation of the target accumulator itself.

---

12 although there is no visible speedups in the figure due to the noiseless nature of the simulations, speedups due to statistical facilitation occur every time the manipulation causes the finishing times of the non-target accumulator to approach those of the target accumulator





**Figure 17:** Noiseless illustrations of input strengths and simulations of the accumulation process (activations) in both a standard race model and a normalized race model, as a function of time for the four different conditions of experiments 3a & 3b (rows represent conditions). **Left column:** temporal evolution of input strengths. **Middle column:** accumulation of input up to decision criteria, as would be expected from a non-normalized, independent race model. Speedups in RT compared to the baseline condition are expected in the two bottom boxes of the middle column. Since statistical facilitation, which is responsible for the faster overall RTs, is a stochastic phenomenon, it can't be directly seen in these 'noiseless' illustrations. However, some statistical facilitation should be expected anytime the amount of accumulated input in the non-target (red) accumulator at response is higher than in the baseline condition. **Right column:** accumulation of input up to decision criteria as would be expected from an input normalization race model. Blue lines represent target stimuli, Red (dashed) lines are non-targets and green (dotted) lines are the decision criteria.

It is possible, however, to distinguish between normalized models and other competitive models, by adding another condition to the design that would directly probe the process of normalization. Note that applying normalization to inputs by maintaining their sum constant is equivalent to using their ratio rather than their absolute values. As a result, a process accumulates normalized inputs would be invariant to a multiplication of both inputs by a constant (see Figure 17, top row: *double boost* condition). Consequently, if inputs are manipulated so that they are multiplied by a constant some time after the trial has begun we would expect no

change in RT under a normalized model<sup>13</sup> (Figure 17, first row, right panel). On the other hand, a non-normalized model, whether independent or competitive, would predict a speedup of RTs due to the increase in absolute input strength (Figure 17, first row, middle panel).

This manipulation, in conjunction with a manipulation of difficulty via the increase of non-target input, produces two different sets of predicted outcomes. If no effect is observed in the *double boost* condition then it is possible that inputs are normalized and thus the difficulty manipulation has no discriminating power. That is, as long as inputs are normalized, the prediction will always be the same: a slowdown in the *100ms incongruent input* (Experiment 3a) or *difficult* (Experiment 3b) condition, no matter which type of competition takes place. If, however, faster than baseline response times are observed in the *double boost* condition then independent and competitive models diverge with regard to their predicted outcome for the *100ms incongruent input (difficult)* condition. Here, independent models foretell a speedup relative to baseline and competitive models, the opposite (see Table 1). Thus, in the absence of an assumption of normalization, any observed slowdown in the *100ms incongruent input (difficult)* condition must be the result of competitive mechanisms and cannot be accounted for with an entirely independent, non-normalized model. As before, these predictions are also robust to, and even accentuated by, non-linear, concave, psychophysical transformations of physical input strengths such as logarithmic and power law functions (see Appendix A for simulations demonstrating this).

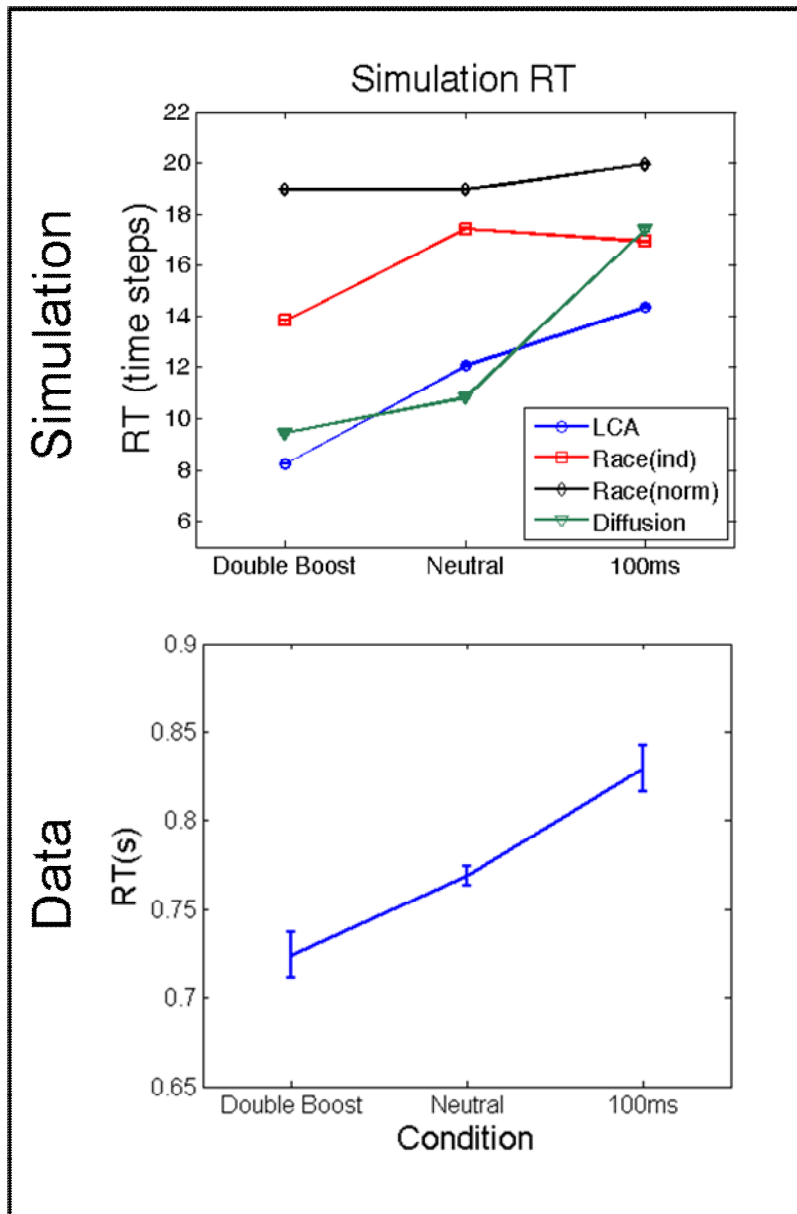
<b>Double Boost</b>	<b>Incongruent Initial input/ Difficult</b>	<b>Models consistent with result</b>
No Effect	Speed Up	-----
	Slow Down	Independent with normalization Competitive with normalization
Speed Up	Speed Up	Independent with no normalization
	Slow Down	Competitive with no normalization

**Table 4:** Possible experimental outcome combinations and corresponding model predictions for study 3. Left and middle columns present possible experimental effects when comparing the baseline condition to the *double boost* and *100ms incongruent input* conditions respectively. The right column shows which competition types predict each possible outcome combination.

Simulations were run to formally evaluate both the effect of increasing task difficulty via a transient augmentation of the input to the non-target ( $I_2$ ) and the multiplication of all inputs by a constant, on the mean-RT of several models (the effect of the *difficult* condition of Experiment 3b are virtually identical to the *100ms incongruent input* and therefore are not shown). Four models were used: a purely independent (*race*) model (red line), a normalized race model (black line), a diffusion model (green line) and an LCA model (blue line; see Figure 18). As one can see in Figure 18 (top), simulation results reveal that the *double boost* condition resulted in

<sup>13</sup> Note that input normalization can be introduced into all types of models and not just independent (race) models (for an application of input normalization to response-competition models (LCA) see Usher & McClelland, 2001).

the fastest mean RTs for all but the normalized race model (black vs. other lines). RTs in the *100ms incongruent input* were slowest for all but the independent race (red vs. other lines). Critically, while the independent race (red line) predicts a speedup in the *double boost* condition it also predicts a small speedup in the *100ms incongruent input* condition due to statistical facilitation. The normalized race (black line), on the other hand, does predict a slow down in the *100ms incongruent input* condition but fails to predict a speedup in the *double boost* condition. It is important to note here that *any* model in which inputs are normalized will predict a null effect for the *double boost* manipulation. Both diffusion and LCA predict a slowdown in RT for the difficult *100ms incongruent input* condition and a speedup for the *double boost* condition.



**Figure 18:** Top: simulation results depicting average RTs as a function of experimental condition (Double Boost/ Neutral/ 100ms incongruent information) for four models: Independent race, Normalized race, Diffusion and LCA. Decision criteria were set so that models predicted approximately the same accuracies for all the various input alternatives (Race: 35; Normalized Race: 38; LCA: 11; Diffusion: 5) Input strengths in the simulations were varied in the following manner: (1)

in the double boost condition, 7 time steps after the onset of the stimuli, both  $I_1$  and  $I_2$  were multiplied by a factor of 1.5 and remained so until a response was given (2) In the 100ms incongruent input condition,  $I_1$  was kept constant ( $I_1=2$ ) while  $I_2$  was set to  $I_2 + 2*(I_1 - I_2)$  for the first 7 time steps after which it returned to its standard value of  $I_2=1.5$ . All models were simulated according to the equations described in the introduction ( $\sigma = 1$ ;  $\lambda = \beta = 0.1$ ). **Bottom:** Response time results as a function of the different conditions in Experiment 3a.

### Experiment 3a: temporal manipulation of inputs.

As in Experiment 2, the participants were asked to decide which of two fluctuating gray patches is the brightest. This time, however, participants were presented with three types of trials that differed in the way mean stimulus brightness evolved during the time course of the trial. The first condition was the *neutral* condition, where both the target and non-target stimuli remained constant until a response was entered. The second was the *double boost* condition which began exactly the same as the *neutral* condition but after 100ms the mean brightness values of both the target and non-target stimuli grew by a factor of 1.5 and remained so until response. The third and last condition was the *100ms incongruent input* in which the target stimulus behaved just like in the *neutral* condition but the non-target's mean brightness level was initially set to a higher value than the target's mean brightness for a duration of 100ms after which mean brightness returned to its baseline level and remained at that level until the subject responded.

## Method

### Participants

Ten TAU students (7 female) participated in the study in exchange for course credits. Participants were tested in two, 45 minute long sessions (no more than four days apart). All participants had normal or corrected to normal vision.

### Materials

For *neutral* trials gray levels of the target and non-target were normally distributed around means of 0.4 and 0.3 (on a 0 (black) to 1 (white) scale) respectively with a SD of 0.1. For the *100ms incongruent input* condition non-target mean gray level was 0.5 for the first 100ms and 0.3 thereafter. Here, target mean gray levels remained unchanged throughout the trial. Last, the *double boost* condition began the same as a *neutral* trial but 100ms into the trial both target and non-target mean gray levels increased from 0.4 and 0.3 to 0.6 and 0.45 respectively. Except for these details, experiment 3 was identical to experiment 2.

### Procedure

All trials were randomly assigned to one of three possible conditions: *neutral*, *100ms incongruent input* and *double boost*. Participants were presented with blocks of 60 trials for a total of 1200 trial per participant. Each block consisted of 40%, 30% and 30% of *neutral*, *100ms incongruent input* and *double boost* trials respectively. All other procedures were identical to experiment 2.

## Results

In comparison to the *neutral* condition ( $M=0.94$ ,  $SD=0.03$ ) participants were less accurate in their responses in both the *double boost* condition ( $M=0.91$ ,  $SD=0.04$ ,  $z=2.1$ ,  $p<0.05$ ; Wilcoxon Matched Pairs Test) and the *100ms incongruent input* condition ( $M=0.86$ ,  $SD=0.06$ ,  $z=2.5$ ,  $p<0.05$ ; Wilcoxon Matched Pairs Test). Importantly, participants responded significantly faster in the *double boost* condition ( $M=0.71s$ ,  $SD=0.14$ ) when compared to the *neutral* condition ( $M=0.76s$ ,  $SD=0.16$ ,  $z=2.5$ ,  $p<0.05$ ; Wilcoxon Matched Pairs Test) thus supporting the hypothesis that inputs were not normalized prior to accumulation. Next, we turn to how the participants performed in the *100ms incongruent input* condition. As can be seen in Figure 18 (bottom), compared to the *neutral* condition, participants responded more slowly to *100ms incongruent input* trials ( $M=0.81s$ ,  $SD=0.18$ ,  $z=2.5$ ,  $P<0.05$ ; Wilcoxon Matched Pairs Test). In tandem, these results validate the interpretation that the slowdown in RT observed in the *100ms incongruent input* condition is the result of some sort of competition that does not originate from a standard, pre-integration normalization of input strengths but rather involves some sort of either input or response competition acting on non-normalized inputs.

Although the results of this experiment do not rule out all input competition models (for example, the FFI model is equivalent to the MMN model for this manipulation and both account well for the results; see Figure 18 top panel 'Diffusion' model), it does create strong constraints for choice models in general. Thus, we can see that not only independent models are unable to account for these findings but also some input competition models such as race with normalized inputs.

One may argue, however, that there is a certain amount of ambiguity in the *100ms incongruent input* condition due to the switch in the designation of the target stimulus 100ms after stimulus onset, which may induce some confusion in the participant regarding task demands (but see Experiment 3b for a design that avoids this). This could provide an alternative account of the observed slowdown. Note, however, that in order for the participant to be confused by the 'switch' he must first perceive or be aware of it at some level. On the one hand, if the subject does not perceive the 'switch' then this type of confusion can be directly translated into additional conflict during the accumulation process. This additional conflict is then naturally accounted for in competitive models as increased competition. If, on the other hand, the participant *does* perceive the 'switch' or is somehow aware of it then this information can (and optimally should) be used advantageously. Note that, only one of the alternatives, the non-target, changes in a 'switch' trial (*100ms incongruent input*) while the other alternative, the target, remains constant. Therefore, the alternative that switches is in fact perfectly correlated with the non-target alternative and provides a 100% valid cue for the incorrect response. Considering the fact that participants are provided with trial by trial feedback (beep on error), then if the 'switch' is indeed detectable subjects should learn to use this information to improve their accuracy and possibly, though not necessarily, reduce their RTs (see e.g., Bauer, Cheadle, Parton, Mueller & Usher, 2009; Figure 2A). This, however, is not compatible with our results which show that participants are both slower and less accurate in the *100ms incongruent input* condition.

Experiment 3b was designed to replicate the results from Experiment 3a with a manipulation that did not involve potentially confusing, non-stationary stimuli as in the *100ms incongruent input* condition. As discussed above, one possible interpretation of the results in Experiment 3a is that the slowdown observed in the

*100ms incongruent input* condition was the result of subjects being confused by the temporal inconsistency of the stimuli. In addition, the experiment was designed to replicate the results of experiment 1a and 1b with two instead of four response alternatives. The purpose of doing so was to eliminate alternative accounts of these results. One possible interpretation of the results of Experiment 1a & 1b is that subjects first rule the two weakest alternatives and only then proceed to accumulate the values of the two remaining alternatives. If this is correct then it should be possible to account for the data with a normalized (though not an independent) race model. This is true since once the two weak alternatives are eliminated the two remaining alternatives are no longer normalized and therefore the increase in the input for the non-target alternative (as in the *difficult* condition) will result in a decrease of the input for the target alternative in order to maintain normalization. This will then lead to an overall slowdown of mean RTs as observed in the data.

### **Experiment 3b: non-temporal manipulation of inputs**

The design of the experiment 3b closely resembles that of Experiment 3a, with the bottom row, replacing the third row in Figure 17. Thus, like in Experiment 3a, three types of trials are randomly presented: i) *neutral* (easy), ii) *difficult*, and iii) *double boost*. The *difficult* condition is created by taking the *neutral* condition and increasing the brightness level of the non-target alternative, while the *neutral* and *double boost* conditions are identical to the ones from Experiment 3a. Due to the profound similarities, model predictions for this experiment are also practically identical to the ones from Experiment 3a (see Figure 18 top panel).

## **Method**

### **Participants**

Nine TAU students participated in the study in exchange for course credits. Participants were tested in two, 45 minute long sessions (no more than four days apart). All participants had normal or corrected to normal vision.

### **Materials**

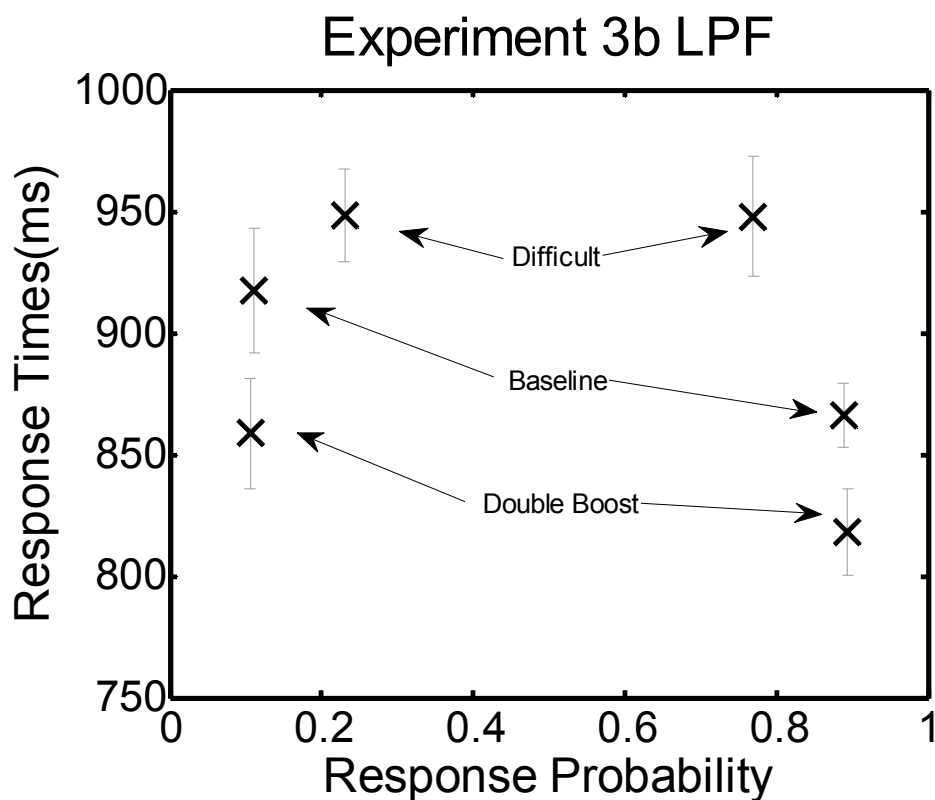
For *neutral* trials mean gray levels of the target and non-target were 0.4 and 0.3 (on a 0 (black) to 1 (white) scale) respectively. Gray levels were perturbed by a random noise variable that was drawn from a normal distribution with a SD of 0.2 that was truncated above 0.1 and below -0.1 to avoid obvious flickering of the stimuli. The *double boost* condition began the same as a *neutral* trial but 100ms into the trial both target and non-target mean gray levels increased from 0.4 and 0.3 to 0.6 and 0.45 respectively (a factor of 1.5). In the *difficult* condition, mean gray levels were 0.4 and 0.34 for the target and non-target stimuli respectively.

### **Procedure**

All trials were randomly assigned to one of three possible conditions: *difficult*, *neutral* and *double boost*. Participants were presented with blocks of 60 trials for a total of 1200 trial per participant. Each block consisted of 40%, 30% and 30% of *neutral*, *difficult* and *double boost* trials respectively. All other procedures were identical to experiment 3a.

## **Results and Discussion**

In comparison to the accuracy of the *neutral* condition ( $M=0.88$ ,  $SD=0.02$ ) participants were less accurate in their responses in the *difficult* condition ( $M=0.77$ ,  $SD=0.03$ ,  $z=2.7$ ,  $p<0.01$ ; Wilcoxon Matched Pairs Test). However, subjects were just as accurate in their responses in the *double boost* condition ( $M=0.89$ ,  $SD=0.02$ ,  $z=0.3$ ,  $p=0.77$ ; Wilcoxon Matched Pairs Test) as in the *neutral* condition. Therefore, we did not replicate the decrease in accuracy observed in experiment 3a. Importantly, participants responded significantly faster in the *double boost* condition ( $M=0.82s$ ,  $SD=0.26$ ) when compared to the *neutral* condition ( $M=0.87s$ ,  $SD=0.26$ ,  $z=2.7$ ,  $p<0.01$ ; Wilcoxon Matched Pairs Test) thus replicating the finding from experiment 3a and supporting the hypothesis that inputs were not normalized prior to accumulation. In addition, compared to the *neutral* condition, participants responded more slowly in *difficult* trials ( $M=0.95s$ ,  $SD=0.35$ ,  $z=2.7$ ,  $P<0.01$ ; Wilcoxon Matched Pairs Test). This result replicated the finding from Experiment 1 supporting the hypothesis that subjects slow down when non-target input is increased. Error RTs for the *double boost* condition ( $M=0.86s$ ,  $SD=0.29$ ) were only marginally faster than *baseline* ( $0.92s$ ,  $SD=0.3$ ,  $z=1.8$ ,  $p=0.07$ ; Wilcoxon Matched Pairs Test) while RTs were not significantly slower in the *difficult* condition ( $M=0.95s$ ,  $SD=0.34$ ,  $z=0.77$ ,  $p=0.44$ ; Wilcoxon Matched Pairs Test) compared to *baseline*. Mean overall RTs for correct and error responses are presented in Figure 19.



**Figure 19:** Results of Experiment 3b. Xs represent mean correct (left) and error (right) RTs for the three experimental conditions: *double boost*, *neutral* and *difficult*. Error bars stand for two within subject standard errors. Lateral position on the x axis indicates the probability a specific type of response (correct>0.5; error<0.5) within that condition.

When taken together, the results of experiment 3b provide converging evidence supporting the hypothesis that the observed slowdowns in both experiments 1a & 1b as well as experiment 3b are a result of some sort of competition that can be captured with neither independent models or normalized models. These findings also rule out the possibility that the results of Experiment 1 were solely due to a dual stage process where subjects first rule out the two weakest alternatives and then continue with a normalized race between the remaining two alternatives. Finally, the results of experiment 3b also rule out the alternative account of experiment 3a where the observed slowdown in the *100ms incongruent input* condition is explained by the added confusion in task demands due to the temporal inconsistency of the target alternative.

## **General Discussion**

Sequential sampling models have been highly successful in accounting for complex choice RT data in a variety of tasks. However, the great resilience of these models has also marked an equally great challenge – how to tell them apart. The task of distinguishing sequential sampling models remains complex and has so far produced mixed, inconclusive results. We addressed this problem by comparing qualitative model predictions and quantitative model fits to results from a set of experiments specifically designed to discriminate between several classes of models by manipulating the values of the non-target alternatives.

To this end we combine theoretical, experimental and computational approaches. First a functional mapping of possible competition loci was proposed: stimulus vs. input vs. response competition. We argue that some properties of the 1D stimuli used to produce the data to which models are compared can lead to phenomena such as stimulus competition and low level, local input competition. The danger in competitive 1D stimuli is that they introduce strong correlations between physical attributes of the choice alternatives. Similarly, stimuli that induce local input competition can introduce strong correlations between the neural representations of the stimuli. Interactions at the local input level can be considered automatic in the sense that they occur whether or not a decision is required. Thus, we treat them as belonging to a perceptual encoding, rather than a decision, stage. Importantly, if not controlled or eliminated, these types of pre-decision competition could be confounded with higher level decision competition and thus might disguise important functional differences between perceptual choice models. In addition, such stimuli also limit the discriminability of the models by restricting the freedom the experimenter has to independently manipulate the values of the various alternatives. Without such independent manipulation it is difficult to separately quantify the amount of input available to each channel in the decision process and one can only approximate them by ad hoc fits to data at the added cost of increasing the number of free parameters. More free parameters inevitably lead to higher model complexity making it harder to discriminate between models.



The perceptual stage where input competition takes place is complex and stimulus specific. Fortunately though, one can study the more general decision mechanism without a detailed perceptual model since perceptual interactions are highly dependent on physical stimulus properties and can be minimized by carefully controlling certain attributes, such as spatial or temporal separation and stimulus complexity. Stimuli that are simple and well separated in space (for ex. brightness at distinct non-overlapping locations) avoid problems, such as physical correlations (stimulus competition) and correlations in early processing (input competition) since the value of each alternative is represented at a separate, non-overlapping, retinotopic location in the primary visual cortex. Consequently, by using brightness stimuli for which the neural response is (up to a monotonic, psychophysical transformation) proportional to the physical magnitudes, we were able to constrain model inputs to a degree sufficient for the production of qualitatively diverging predictions.

We further specify our theoretical taxonomy of competition loci by proposing a distinction according to which 'input' and 'response' are two distinct competition loci within the decision mechanism itself that operate on the basis of different available information. Input competition is proportional to the momentary value of each alternative and does not include the total, accumulated values. Response competition, on the other hand, is a function of the total accumulated values, rather than the momentary value of each response alternative.

Building on these distinctions, we developed three experimental paradigms that were aimed at contrasting independent models with models employing different types of decision competition. To illustrate the differences between these model classes we used several computational models, which encompass the various types of decision competition. Our experimental manipulations were designed in such a way as to produce diverging predictions for the different models classes. The first study carried out a manipulation of the value of the strongest non-target alternative while maintaining the total sum of all values constant. In the second study we manipulated the priors for each alternative. In the third, we manipulated the amount (Experiment 3b) and temporal distribution (Experiment 3a) of the value of the non-target alternative while simultaneously probing for normalization. All these manipulations are based on the principle of increasing non-target value while maintaining target value constant. In the absence of interactions between processing channels, independent models can predict slower overall RTs only if the input for at least one of the alternatives is reduced. Otherwise, independent models predict faster RTs due either to stronger target activation (faster target accumulator finishing times) or to stronger non-target activation (faster non-target finishing times leading to statistical facilitation). Competitive models, on the other hand, vary with regard to their predictions for such manipulations and can show complex behavior. In all three studies we found that participants respond more slowly when the non-target value is stronger even when the experimental design (randomized within a block) makes it unlikely that participants change the decision criteria to accommodate different levels of difficulty for each trial (Ratcliff & Smith, 2004; Ratcliff & McKoon, 2008, VanZandt et al., 2000).

To further test these predictions, models were fit to data from Experiment 1a and Experiment 2. All of the models tested here have been successfully fit in the past to many variations of standard RT data such as RT distributions and accuracies. Indeed, all the models captured these aspects of the data well. We therefore assumed that the

models possess the necessary flexibility to also do so for the rest of the experiments in this paper. The real challenge for the different categories of models, however, was to fit the main effect of RT difference between conditions. In accordance with a-priori simulations and theoretical model predictions, when model inputs were constrained to the physical stimulus values, the independent race model was unable to fit the slowdown in overall RT between easy/baseline and difficult/incongruent conditions in Experiment 1a and 2 and in fact predicted the opposite pattern (but see discussion below regarding starting point assumptions in Experiment 2). The remaining experiments (1b, 3a & 3b) are variations and replications of the same basic underlying manipulation (i.e. increased non-target value). Therefore, independent models would make qualitatively equivalent predictions (i.e. speedup instead of slowdown) for these experiments as well as for any experiment with similar manipulations

Some input competition models also find our main effect results hard to account for, due to a variety of reasons both idiosyncratic and general. More specifically, the two input competition models we tested here (normalized race and FFI) failed, to various degrees, to account for the observed effects. The normalized race could not account for the overall RT slowdown main effect observed in studies 1 & 2 and for the RT speedup in the *double boost* condition in study 3a & 3b. The FFI model had difficulties only with the RT slowdown effects observed in studies 1 & 2. Some of these failures are idiosyncratic to a specific model assumption. For example, the normalized race could not account for results of study 1 & 3 because of the normalization of input strengths. Although normalization is a type of input competition, this problem is not intrinsic to a particular class of models since in fact *any* model, either independent or competitive, employing such an input normalization would produce the same results. Similarly, the failure of the FFI model in Study 1 is a result of the specific assumption that inhibition on each accumulator is calculated as the *mean* of the inputs to the other alternatives. Future investigations are needed to examine in more detail differences in predictions between input and response-competition models, as well as differences between different input competition models.

It is interesting to note that some input-competition models not tested here, such as a multiple choice version of Vickers' accumulator (Vickers, 1970), could be consistent with the data from Studies 1 and 3. Vickers' accumulator model has a competitive mechanism that, for each momentary information sample, increases only one accumulator – the one with the strongest momentary support. Therefore, in Experiment 1a, this n-choice accumulator model predicts that the accumulator pertaining to the strongest non-target 'steals' winning samples from the target accumulator, and thus slows down RT (simulations not reported). Such a model, however, would find it difficult to account for the data obtained in Experiment 1b where (except for rare occasions) at any given moment the display provides input for only one alternative and therefore the non-target accumulator can no longer steal winning samples from the target accumulator. This, however, depends on the definition of the *duration* of a sample. In Experiment 1b the fastest average flicker arrival rate was 10 per second. Thus, for longer sample intervals, this probability increases gradually allowing the non-target accumulator to steal more and more wins from the target, leading to increasing RT slowdowns. On the other hand, longer

samples mean fewer samples per decision, making choice RT more discreet and therefore more dependent on variability in non-decision processes to account for RT.

Unlike the accumulator model, models that assume an independent race acting on normalized input (for example the normalized race model; see Figures: 3 right, 18 top), do not account for data from either of our experiments. While, the assumption of input normalization is quite frequent in perceptual choice models (Brown & Heathcote, 2008; Leite & Ratcliff, 2010; Usher & McClelland, 2001), we argue that such normalization procedures are appropriate only when stimulus properties mandate it. For example, input normalization could be valid for 1D stimuli, reflecting the constraint that such stimuli impose on the balance of evidence in the display. In practice, input normalization is often also used with other types of stimuli for several possible reasons. One reason is that normalization is a mathematical, simplifying assumption that reduces one model parameter for fitting purposes; see also discussion in Donkin, Brown and Heathcote (2009), who provide data that argues against this procedure. Another reason for using input normalization is that it can help establish an equivalence between structurally different models, such as the LCA and the drift diffusion (see Bogacz et al., 2006, p. 712).

Despite the mathematical convenience, one has to be alert to the problematic nature of assuming input-normalization without a mechanistic interpretation. This is especially the case when trying to distinguish rather than equate models. We suggest that if a model with input normalization does provide a good account for experimental data, which was obtained with stimuli that are not on a 1D continuum, it is important then to carefully interpret *the process* that this normalization corresponds to. Such a mechanistic interpretation can be provided within the framework of input competition (see models described in Ditterich (2010), which provide explicit formulations of processes that mediate input normalization).<sup>14</sup>

Experiment 2 provides insights into possible ways of discriminating between input and response competition models. Affecting the initial conditions of the accumulation process can, under certain assumptions, have differential effects on the two classes of models. The interactions between processing channels in input competition models are a function of the momentary inputs and therefore are not *intrinsically* sensitive to manipulations of initial conditions. On the other hand, interactions between processing channels in response competition models are a function of the *total* accumulated information which includes both momentary inputs and initial conditions, making them intrinsically sensitive to manipulations of the latter. The distinctiveness of these predictions is, however, muddled by the unresolved issue of how exactly prior knowledge is implemented in the decision process.

Prior knowledge is often conceptualized as affecting either drift rates, starting points or both. If priors affect drift rates then input competition, though not independent, models could account for the RT slowdown effect observed in Experiment 2. The lack of significant differences between leading edge and tail effects in Experiment 2, and the opposite effects observed for error and correct RT (see also Mulder et al., in press), however, provide evidence against the drift rate

---

14 A normalization constraint in the linear ballistic model (Brown & Heathcote, 2008) could be interpreted to correspond, either to stimulus-inhibition (1D stimuli) or to shunting inhibition (Andrew Heathcote, personal communication, 8-Jan, 2011).

interpretation. In addition, it has been shown that, biasing only starting points is optimal and that drift rates are not strategically adapted to produce response bias (Simen et al., 2009; but see Ravenzwaaij et al., in press).<sup>15</sup>

In addition, the effect of priors on starting points could be either correlated or uncorrelated. The uncorrelated starting points assumption postulates that prior knowledge in favor of one alternative affects only the starting point of its corresponding accumulator. Alternatively, the anti-correlated starting points assumption postulates that an increase in the starting point of one accumulator is accompanied by a decrease in the starting point of the other accumulators. This is important because under the uncorrelated assumption, both independent and input competition models fail to qualitatively account for the observed main RT-slowdown effect of Experiment 2, while under the anti-correlated assumption both succeed (though not as well as response competition models).

Computational explorations of model predictions for Experiment 2, however, provide a possible avenue to discriminate between these alternative hypotheses and therefore also between input and response competition. As can be seen in Figures 16a & 16b, the uncorrelated assumption allows both independent and input competition models to predict *only* RT speedups while the anti-correlated assumption allows for *only* RT slowdowns. Response competition models with uncorrelated starting points, on the other hand, predict RT slowdowns for high accuracies and RT speedups for low accuracies. While exploring these predictions empirically would require a more detailed, dedicated study where difficulty and accuracy instructions are also manipulated to produce a wide range of accuracies, the results of Experiment 2 provide some preliminary evidence for the uncorrelated response competition alternative. While most subjects in Experiment 2 exhibited RT-slowdowns with high accuracy, one individual subject exhibited the opposite pattern, RT speedup with low accuracy, a pattern which is in accordance with the predictions of uncorrelated response competition models. In addition, the MMN and LCA models with uncorrelated starting points provided better fits to the data than the FFI and race models with anti-correlated starting points.

Finally, response competition models accounted well for *all* of the results reported in this paper. When difficulty is increased by increasing the value or the prior favoring a weak (non-target) alternative, decision-time of response competition models such as the LCA (Bogacz, et al., 2007; Usher & McClelland, 2001) and the MMN diffusion (McMillen & Holmes, 2006; Ratcliff & Rouder, 1998), slows down due to stronger inhibition of the target by the, now more strongly activated, non-target accumulator (see Hanks, Ditterich & Shadlen, 2006, for additional converging evidence supporting this finding and Ridderinkhof, Wildenberg, Wijnen, & Burle, 2004, for evidence supporting response competition in attentional tasks). This precise pattern was obtained in all our experiments, revealing evidence for a compensatory mechanism which naturally accounts for RTs, while also allowing the model to improve accuracy by integrating more information without any top-down executive control (i.e. threshold changes). Note that our finding that people slow down to

---

15 Alternatively, it is also possible that interpreting priors as uncorrelated vs. correlated starting points could be dependant on the specific task demands and the stimuli used. For example, 1D stimuli might be more conducive to anti-correlation in starting points while independent, spatially separated stimuli might result in uncorrelated starting points.

compensate for higher stimulus difficulty may seem trivial at first glance. Contrary to this, however, a number of studies in the domain of eye-movement and saccade control show a reverse pattern: an increase in errors *without* a compensatory slow-down in RT (Caspi et al., 2004, Ludwig, et al., 2005). This is consistent with the idea that the employment of competition is contingent on the specificity of the system. For example, independent models could underlie mechanisms like saccade control, where speed may be prioritized over accuracy. Future research is needed to examine if response competition is a general property of the decision system and whether it depends on task demands (for example accuracy vs. speed emphasis).

Competition, both of the input and response type is often associated with optimality. For example, the variant of the diffusion model which involves a type of feed-forward competition (Mazurek, et al., 2003), has been argued to be optimal (Bogacz et al. 2006; Gold & Shadlen, 2001, 2002; Wald, 1947), in the sense that it provides the fastest RTs for a given accuracy level or, vice versa, the highest accuracy for a given RT. In our simulations (Figures 6, 9, 13), the decision criteria for all models were set so that they produce similar accuracies for all difficulty levels. Since optimality (in the sense above) implies the fastest RTs for a given accuracy, the model that predicts the fastest RTs in our simulations can be considered as the most optimal one. As can be clearly seen in Figures 6, 9 and 13, the MMN diffusion model produces the fastest RTs over most ranges. This conclusion needs to be qualified, however, as we only examined a small set of choice models, and since the proof of the diffusion model as optimal in the sense above only applies to specific situations, such as those that involve binary decisions with a single level of stimulus difficulty. However, more general optimal decision mechanisms have been proposed recently, which address multiple choice (Bogacz, 2008; Ditterich, 2010) and that still rely on some type of competitive interactions. We further qualify this by noting that a comprehensive discussion of optimality should also take into account the resources available to the organism (Anderson, J. R., 1990).

One of the reasons that competition leads to such efficient decision is that it allows the decision mechanism to integrate more information for difficult decision trials (slower RTs) and integrate less information for easy ones (faster RTs). Competitive mechanisms achieve this tradeoff without any *on-line* monitoring of difficulty or adjustments to thresholds. Such monitoring mechanisms which allow for rapid on-line (and within trial) adjustments of thresholds could allow race models to exhibit many of the characteristics of competitive models. Note, however, that the online evaluation of stimulus difficulty requires some process of evidence comparison (e.g., Vickers' balance-of evidence; Vickers, 1979), which, when feeding back into the decision process, would make such adjusted-race models not independent. Within the LCA model, for example, efficiency is enhanced by applying lateral inhibition to an otherwise independent race architecture (Bogacz et al. 2007; see also, Van Ravenzwaaij, Van der Maas & Wagenmakers, 2011).

To conclude, we demonstrated that progress towards distinguishing between different classes of models can be achieved by following several theoretical and methodological steps including: i) a taxonomy of model space that generates distinct mutually exclusive categories along a single dimension (e.g. independence vs. competition), ii) deliberate choice of manipulations that drive strong inference experimentation (Jewett, 2005; Platt, 1964) and iii) a coalition of statistical measures consisting of data fits, inferential statistics and prediction space mapping. The success

of such distinctions may often depend on the resolution of other, adjacent processes. For example, distinguishing between input and response competition may ultimately depend on first accounting for the mechanism through which prior information affects the initial conditions of the decision process. We propose that the best account for our results is that perceptual decision mechanisms are intrinsically competitive and that independent models are unable to account for these data. The results also provide some constraints on input competition models and suggest that competition is more likely to take place at the accumulator or response level. Future research is needed to extend the set of models examined to additional instances of input and response-competition models. For example, one needs to examine other input-competition models, in which the inhibition is not the average input of the other alternatives, but rather a more complex function, as well as other response competition models (e.g. the attractor model Wang, 2008; Albantakis & Deco, 2009)<sup>16</sup>.

Some steps in this direction were done in recent work. For example, decisions based on non-stationary (time-varying) evidence can distinguish between a FFI and LCA models (Tsetsos et al., 2011; Tsetsos et al, 2012; see also Zhou Wong-Lin, & Holmes, 2009). Additional studies could examine multi-alternative choice with variable set size (randomized within blocks), which is likely to probe the presence of inhibition-rescaling with set size, in the FFI-model. Finally, this methodology could be extended to examine other properties of the decision mechanism, such as leaky vs. perfect integration, absolute vs. relative stopping rules and noisy vs. ballistic integration.

---

16 Another interesting class of response-competition corresponds to top-down models, in which the accumulated evidence in each accumulator draws attention (by some gain increment) to the incoming input from the corresponding choice alternative. We thank an anonymous reviewer for this suggestion.



## References:

- Albantakis, L., & Deco, G. (2009). The encoding of alternatives in multiple-choice decision-making. *Proceedings of the National Academy of Science U.S.A.*, 106, 10308–10313.
- Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review*, 78(3), 171-206.
- Anderson, J. R. (1990). The adaptive character of thought. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bauer, F., Cheadle, S. W., Parton, A., Muller, H. J., & Usher, M. (2009). Gamma flicker triggers attentional selection without awareness. *Proceedings of the National Academy of Science*, 106(5), 1666–1671.
- Brown, S.D., & Heathcote, A. (2005) A ballistic model of choice response time. *Psychological Review*, 112, 117-128.
- Brown, S.D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57, 153-178.
- Bogacz, R. (2009). “Optimal decision making theories” in Dreher, L. (Eds.), *Handbook of Reward and Decision Making* (pp. 375–397). Oxford: Academic Press.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J.D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, 113, 700-765.
- Bogacz, R., Usher, M., Zhang, J., & McClelland, J. L. (2007). Extending a biologically inspired model of choice: multi-alternatives, non-linearity and value-based multidimensional choice. *Philosophical Transactions B*, 362, 1655.
- Bogacz, R., & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, 19, 442-477.
- Burr, D.C., Ross, J., & Morrone, M.C. (1985). Local regulation of luminance gain. *Vision Research*, 25, 717–727.
- Burr, D. C., Santoro, L., (2001). Temporal integration of optic flow, measured by contrast and coherence thresholds, *Vision Research*, Vol 41(15), 1891-1899.
- Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of the*



*National Academy of Sciences of the United States of America*, 101, 13086–13090.

Cook, P.B., & McReynolds, J.S. (1998). Lateral inhibition in the inner retina is important for spatial tuning of ganglion cells. *Nature. Neuroscience*, 1, 714–719.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology* 1(1), 42-45.

Desimone R., & Ungerleider L.G. (1989). Neural mechanisms of visual processing in monkeys. In Boller, F., & Grafman, J. (Eds.), *Handbook of Neuropsychology* (pp. 267–299). Amsterdam: Elsevier Science Publishers.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without attention effect. *Science*, 311(17), 1005–1007.

Diederich, A. & Busemeyer, J. R. 2006 Modeling the effects of payoffs on response bias in a perceptual discrimination task: threshold bound, drift rate change, or two stage processing hypothesis. *Percept. Psychophys.* 97, 51–72.

Ditterich, J. (2010). A comparison between mechanisms of multi- alternative perceptual decision making: Ability to explain human behavior, predictions for neurophysiology, and relationship with decision theory. *Frontiers in Neuroscience*, 4(184).

Donkin, C., Brown, S., & Heathcote, A. (2009). The over-constraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review*, 16, 1129–1135.

Donkin, C, Brown S, Heathcote, A, Wagenmakers, E (2011) Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin and Review* 18:61– 69.

Donner, T.H., Siegel, M., Fries, P., & Engel, A.K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making, *Current Biology* 19(18), 1581–1585.

Dyrholm, M., Kyllingsbæk, S., Espeseth, T., Bundesen, C., (2011). Generalizing parametric models by introducing trial-by-trial parameter variability: The case of TVA, *Journal of Mathematical Psychology*, 55(6), 416-429.

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., 2012; Independent race of colour and word can predict the Stroop effect. *Australian Journal of Psychology*, in press.
- Eidels, A., Townsend, J.T., & Algom, D. (2009). Comparing perception of Stroop stimuli in focused versus divided attention: Evidence for dramatic processing differences. *Cognition*, 114(2), 129-150.
- Fechner, G.T. (1877). *In Sachen der Psychophysik*. Leipzig, 1877.
- Forstmann, B. U., Brown, S., Dutilh, G., Neumann, J., & Wagenmakers, E.-J. (2010). The neural substrate of prior information in perceptual decision making: A model-based analysis. *Frontiers in Human Neuroscience*, 40:4.
- Gao, J., Tortell, R., & McClelland, J. L. (2011). Dynamic integration of reward and stimulus information in perceptual decision making. *PLoS ONE* 6(3)
- Glimcher, P.W. (2003). The Neurobiology of Visual Saccadic Decision Making. *Annual Review of Neuroscience*. 26: 133-179.
- Gold, J.I., & Shadlen, M.N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, 404, 390 - 394.
- Gold, J.I., & Shadlen, MN (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cog Sci* 5, 10-16.
- Gold, J.I., & Shadlen, M.N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, 30, 535-574.
- Grossberg, S., Mingolla, E., & Ross W.D. (1997). Visual brain and visual perception: how does the cortex do perceptual grouping? *Trends in Neuroscience*, 20(3), 106-111.
- Hanes, D.P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, 274 ,427-430.
- Hanks T.D., Ditterich J., Shadlen M.N., (2006). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nature Neuroscience*, 9, 682–89.
- Heathcote, A. (2005). A Ballistic Model of Choice Response Time. *Psychological Review*, 112(1), 117–128.
- Heekeren, H.R., Marrett, S., Bandettini, P.A., & Ungerleider, L.G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, 431, 859-862.
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15, 534–539.

- Howard, M.W., & Kahana, M.J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 269–299.
- Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition— sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 1311–1334.
- Jewett, D.L. (2005). "What's wrong with single hypotheses? Why it is time for Strong-Inference-PLUS". *Scientist* 19 (21): 10.
- Kastner, S., & Ungerleider, L.G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, 23, 315–341
- Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *Journal of Neuroscience*, 28(12), 3017.
- Kyllingsbæk, S , Marcussen, B & Bundesen, C., (2011), Testing a poisson counter model for visual identification of briefly presented, mutually confusable single stimuli in pure accuracy tasks, *Journal of Experimental Psychology: Human Perception and Performance* .
- Laming D. R. J., (1968). Information Theory of Choice-Reaction Times. London: Academic.
- La berge, D. (1962). A recruitment theory of simple behaviour. *Psychometrika*, 27(4), 375-396.
- Lee, D.K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature Neuroscience*, 2, 375-381.
- Leite, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception and Psychophysics*, 72, 246-273.
- Link, S.W., & Heath, R.A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40(1), 77-105.
- Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, 91, 295– 327.
- Logan, G. D., Cowan, W. B., & Davis, K. A. (1984). On the ability to inhibit simple and choice reaction time responses: A model and a method. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 276–291.
- Luce, R. D. (1959). Individual choice behavior: a theoretical analysis. New York: Wiley.

- Luce R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford, UK: Oxford Univ. Press.
- Ludwig, C.J.H., Gilchrist, I.D., McSorley, E., & Baddeley, R.J. (2005). The temporal impulse response underlying saccadic decisions. *Journal of Neuroscience*, 25, 9907–9912.
- Luck, S.J., Chelazzi, L., Hillyard, S.A., & Desimone, R. (1997). Neural Mechanisms of Spatial Selective Attention in Areas V1, V2, and V4 of Macaque Visual Cortex. *Journal of Neurophysiology* 77, 24-42.
- Mazurek, M.E., Roitman, J.D., Ditterich, J., & Shadlen, M.N. (2003). A role for neural integration in perceptual decision making. *Cerebral Cortex*, 13, 1257.
- McMillen, T., & Holmes, P. (2006). The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, 50(1), 30–57.
- Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in extrastriate cortex. *Science*, 229(4715), 782-784.
- Mordkoff, J. T. Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 520-38.
- Morton, J. (1964). A model for continuous language behavior. *Language and Speech*, 7, 40-70.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165–178.
- Morton, J. (1970). "A functional model for memory" in Norman D.A. (Eds.), *Models of Human Memory* (pp. 203-254). New-York & London: Academic Press.
- Mulder, M.J., Wagenmakers, E.J., Ratcliff, R., Wouter Boekel, W. & Forstmann, B.U. (in press). Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience*.
- Niwa, M., & Ditterich, J. (2008). Perceptual decisions between multiple directions of visual motion. *Journal of Neuroscience*, 28(17), 4435-4445.
- Palmer J, Huk AC, Shadlen MN. 2005. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, 5:376–404
- Philiastides, M.G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex* 16, 509 –518.

- Philastides, M.G., Ratcliff, R., & Sajda, P. (2006). Neural Representation of Task Difficulty and Decision Making during Perceptual Categorization: A Timing Diagram. *The Journal of Neuroscience*, 26(35), 8965–8975.
- Pike, R. (1973). Response latency models for signal detection. *Psychological Review*, 80, 53–68.
- Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, 6, 421–425.
- Pitt, M. A., Myung, I. J., Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, 109, 472–491.
- Platt, J. R. (1964, October 16). Strong inference. *Science*, 146, 347–353.
- Posner, M.I., Snyder, C.R.R., & Davidson, B.J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109(2), 160–174.
- Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, 117(4), 1113–1143.
- Raab, D.H. (1962b). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, 24, 574–590.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86, 446–461.
- Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review*, 95, 238–255.
- Ratcliff, R. (2001). Diffusion and random walk models. In *International encyclopedia of the social and behavioral sciences* (Vol. 6, pp. 3668–3673). Oxford, England: Elsevier.
- Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, 16, 323–341.
- Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception and Psychophysics*, 65, 523–535.
- Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, 50, 408–424.

- Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging*, 19, 278–289.
- Ratcliff, R., Hasegawa, Y., Hasegawa, R., Smith, P. L., & Segraves, M. (2007). A dual diffusion model for single cell recording data from the superior colliculus in brightness discrimination task. *Journal of Neurophysiology*, 97, 1756–1797.
- Ratcliff, R., & McKoon, G. (2008). The diffusion model: Theory and data for two-choice decision tasks. *Neural Computation*, 20, 873–922.
- Ratcliff, R., & Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347–356.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin and Review*, 9, 438–481.
- Ratcliff, R., & Rouder, J. N. (2004). Comparing categorization models. *Journal of Experimental Psychology: General*, 133, 63–82.
- Ratcliff, R., & Smith, P.L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychological review*, 111(2), 333–367.
- Ratcliff, R. & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relation between perceptual encoding and decision making. *Journal of Experimental Psychology: General*, 139, 70–94.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, 106, 261–300.
- Ratcliff, R. (in press). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review*.
- Ridderinkhof, K. R., van den Wildenberg, W. P. M., Wijnen, J., & Burle, B. (2004). Response inhibition in conflict tasks is revealed in delta plots. In M. Posner (Ed.), *Cognitive neuroscience of attention* (pp. 369–377). New York: Guilford Press.
- Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107, 358–367.
- Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience—A silver jubilee* (pp. 611–653). Cambridge, MA: MIT Press.

- Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, 108(2), 370-392.
- Rorie A.E., Gao J., McClelland J.L., Newsome W.T. (2010). Integration of Sensory and Reward Information during Perceptual Decision-Making in Lateral Intraparietal Cortex (LIP) of the Macaque Monkey. *PLoS ONE*, 5(2), e9308.
- Rorie, A.E., & Newsome, W.T. (2005). A general mechanism for decision-making in the human brain? *Trends in Cognitive Sciences*, 9(2), 41-43.
- Shadlen, M.N., & Newsome, W.T. (1996). Motion perception: Seeing and deciding. *Proceedings of the National Academy of Science U.S.A.*, 93, 628–633.
- Shadlen, M.N., & Newsome, W.T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation and information coding. *Journal of Neuroscience*, 18, 3870.
- Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., & Cohen, J. D. (2009). Reward rate optimization in two-alternative decision making: Empirical tests of theoretical predictions. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1865–1987.
- Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, 116, 283-317.
- Smith, P.L., Vickers D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*. 32, 135–168.
- Smith, P.L. & Van Zandt, T.(2002). Time-dependent Poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology*, 53, 293-315.
- Stevens, S.S. (1957). On the psychophysical law. *Psychological Review* 64(3), 153-181.
- Sugrue L.P., Corrado G.S., Newsome W.T. (2005). Choosing the greater of two goods: neural currencies for valuation and decision making. *Nat. Rev. Neurosci.* 6, 363–75
- Stone, M. (1960). Models for reaction time. *Psychometrika*, 25(3), 251-260.
- Summerfield, C., Koechlin, E. (2008) A neural representation of prior information during perceptual inference. *Neuron*, 59, 336–347
- Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging*, 18, 415–429.

- Thomas, E. A. C., & Ross, B. H. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology*, 21, 136–152.
- Townsend, J.T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, 25, 168-199.
- 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. (1990a). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee (and should) be distinguished. *Psychological Science*, 1, 46-54.
- Townsend, J. T., & Ashby, F. G. (1983). Stochastic modeling of elementary psychological processes. Cambridge: Cambridge University Press.
- 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.
- Tsetsos, K., Usher, M., & Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological Review*, 117(4), 1275-1291.
- Tsetsos, K., Usher, M., & McClelland J.L. (2011). Testing multi-alternative decision models with non-stationary evidence. *Frontiers in Neuroscience*, 5:63
- Tsetsos, K. Gao, J. Usher, M & McClelland J. L. (2012). Using time-varying motion stimuli to explore decision dynamics. *Frontiers in Decision Neuroscience*.
- Uchida N, Kepecs A, Mainen ZF. 2006. Seeing at a glance, smelling in a whiff: rapid forms of perceptual decision making. *Nature Review of Neuroscience*. 7:485–91
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550–592.
- Usher, M., McClelland, J. L., (2004). Loss Aversion and Inhibition in Dynamical Models of Multialternative Choice. *Psychological Review*, Vol 111(3), 757-769.
- Usher, M., Olami, Z., & McClelland, J.L. (2002). Hick's Law in a Stochastic Race Model with Speed–Accuracy Tradeoff. *Journal of Mathematical Psychology*, 46, 704–715.



- Van Ravenzwaaij, D., van der Maas, H. L. J., & Wagenmakers, E.-J. (2011). Optimal Decision Making in Neural Inhibition Models. *Psychological Review*. Advance online publication. doi: 10.1037/a0026275.
- Van Ravenzwaaij, D., Mulder, M., Tuerlinckx, F., & Wagenmakers, E. J. (2012). Do the dynamics of prior information depend on task context? An analysis of optimal performance and an empirical test. Manuscript submitted for publication.
- Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review* 7(3), 424-465.
- Van Zandt, T., Colonius, H., & Proctor, R.W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review* 7(2), 208-256.
- Van Zandt T., & Ratcliff, R. (1995). Statistical mimicking of reaction time data: Single-process models, parameter variability, and mixtures. *Psychonomic Bulletin & Review*, 2(1), 20-54.
- Vanpaemel, W. (2009b). Measuring model complexity with the prior predictive. In Y. Bengio, D. Shuurmans, J. Lafferty, C. K. I. Williams, & A. Culota (Eds.), *Advances in neural information processing systems* (pp. 1919-1927). Cambridge, MA: MIT Press.
- Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, 13, 37–58.
- Vickers, D. (1979). *Decision Processes in visual perception*. Academic Press.
- Vickers, D. (1995). The frequency accrual speed test (fast): A new measure of 'mental speed'. *Personality and Individual Differences*, 19, 863-879.
- Vickers, D., Caudrey, D., & Wilson, R. (1971). Discrimination between frequency of occurrence of two alternative events. *Acta Psychologica*, 35, 151-172.
- Wald, A. (1947). *Sequential analysis*. New York: Wiley.
- Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, 58, 140-159.
- Wang, X.J. (2002). Probabilistic decision making by slow reverberations in cortical circuits. *Neuron*, 36, 2336-2362.
- Wang, X.J. (2008). Decision making in recurrent neuronal circuits, *Neuron* 60, 215–234.

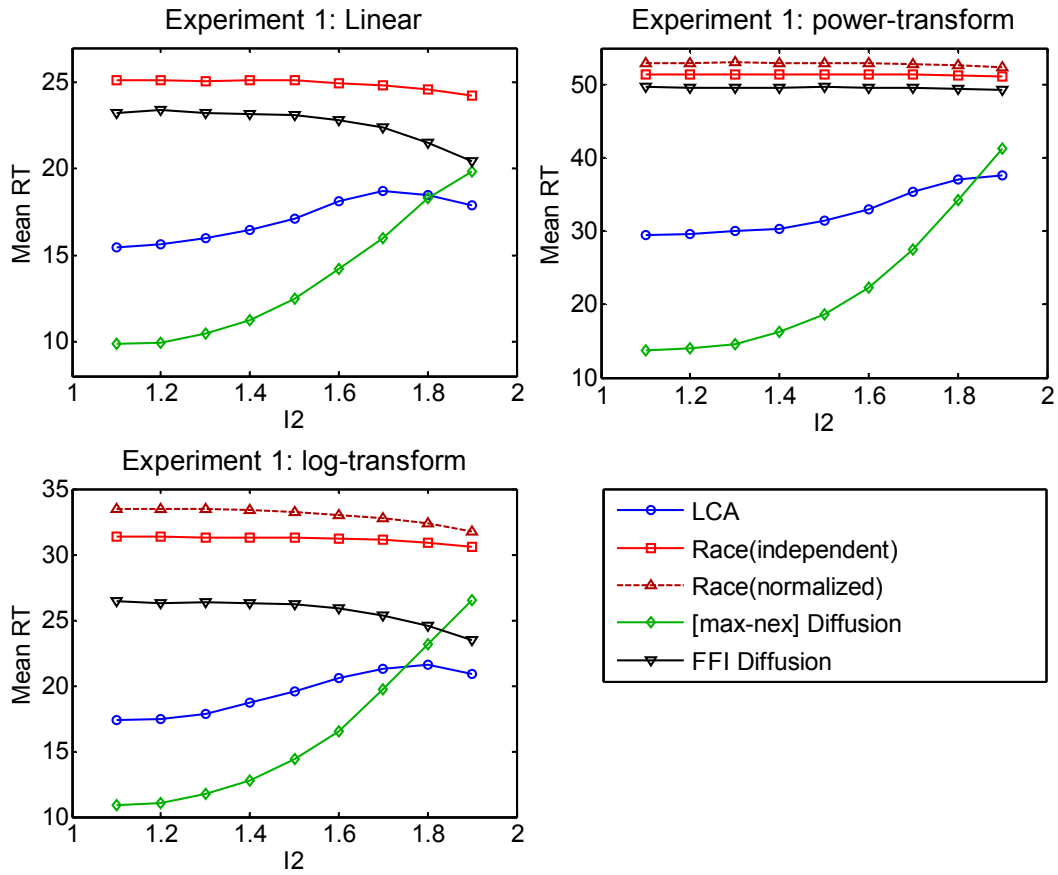
- Wong, K. F., & Wang, X. J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4), 1314.
- Zhou, X., Wong-Lin, K.-F., & Holmes, P. (2009). Time-varying perturbations can distinguish among integrate-to-threshold models for perceptual decision-making in reaction time tasks. *Neural computation*, 21(8), 2336\_2362.

## Appendix A

In this Appendix we address the issue of the robustness of our results when faced with non-linear transformations of inputs prior to their introduction into the model. For simplicity, simulation results presented in the paper correspond to a linear relation between physical stimuli strengths and model inputs. However, as mentioned in the discussion, one could argue that this correspondence is not realistic since neural encoding is often characterized by a saturation of firing rates at higher activation levels. These types of saturating functions have often been described by logarithmic or power law transformations (Fechner, 1877; Stevens, 1957). Brightness is also believed to be a saturating perceptual dimension with an exponent of about 0.5 (Stevens, 1957). Thus, for this argument we will assume that if our stimuli do go through some sort of transformation it will be a monotonically increasing function whose derivative is monotonically decreasing to zero such as a logarithmic or power law ( $\exp=0.5$ ) transformation.

How then, would the introduction of such a function affect model predictions and the interpretation of our data? We present simulations comparing linear vs. logarithmic vs. power law versions of the models used in our study. This non-linearity issue, however, is only relevant to the first and last studies since in our second study only prior knowledge was manipulated and stimuli strengths did not change throughout the experiment. Therefore the issue of psychophysical transformations is, in the case of study 2, reduced to a simple matter of scaling and does not require special notice.

Figure A1 illustrates the effects of non-linear transformations on model predictions for Experiment 1. The five different models discussed in our paper are simulated for the linear, logarithmic and power law transformations. Decision criteria for all models were set such as to produce nearly identical accuracies for all difficulty levels (see Table 2). As can be seen, model predictions remain qualitatively the same despite some variance in the magnitude of the effects.



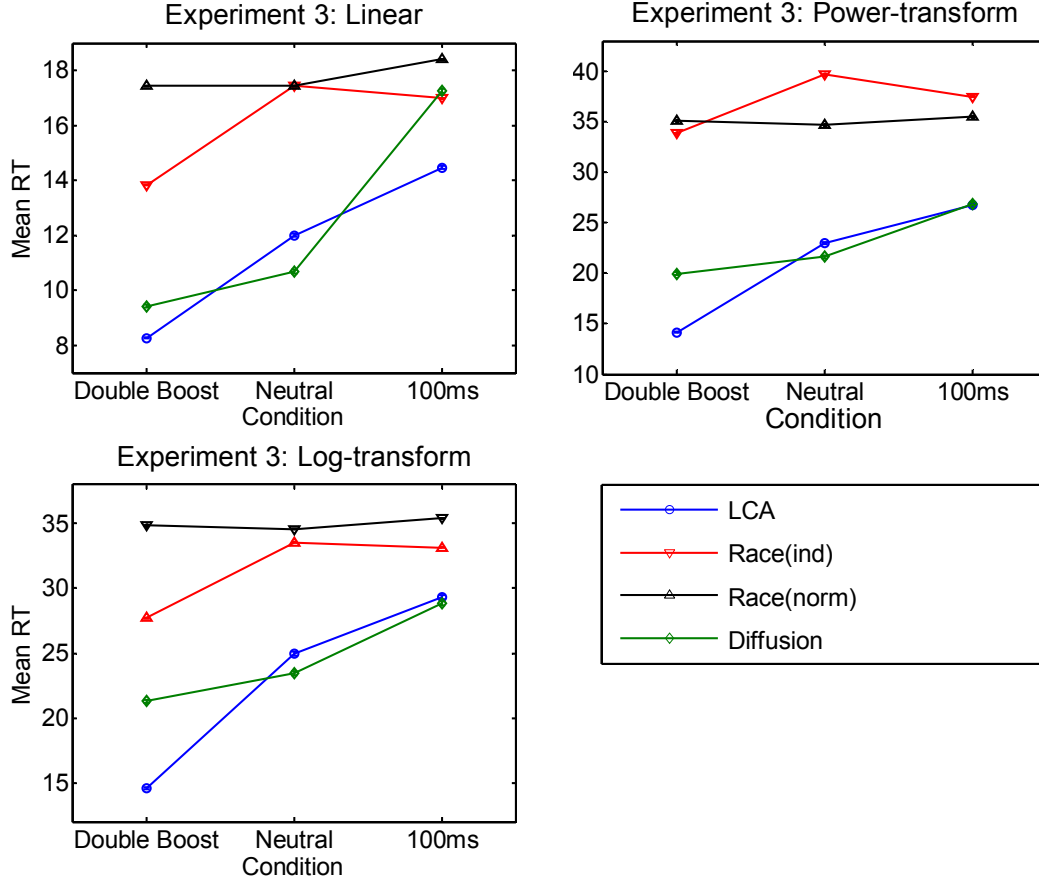
**Figure A1:** Study 1 simulation results, with linear (top), logarithmic (middle) and power (bottom) transformations applied to the inputs prior to accumulation. Mean inputs to the four choice alternatives were manipulated such that  $I_1=2$  remained constant throughout the simulation while  $I_2$  increased from 1.1 to 1.9 in steps of 0.1 and  $I_3=I_4$  decreased from 1.1 to 0.7 in steps of 0.05. For both the logarithmic and power transformation all input values were increased by 1 and noise was not allowed to drive input values below 1. This was done to avoid irregularities in the transformation for values between 0 and 1.

	Experiment 1			Experiment 3		
	Linear	Log	Power	Linear	Log	Power
Race (independent)	50	33	90	35	34	65
Race (normalized)	60	16	35	38	35	60
FFI Diffusion	21	10	85	--	--	--
[max-next] Diffusion	5.5	2.4	2.5	5	4.3	3.6
LCA	12	6	9.5	11	7	10

**Table 2:** Decision criteria used for the different models in appendix A. Each model was assigned three different decision criteria for each experiment, one for the linear case, one for the logarithmic transformation and one for the power transformation, resulting in a total of six different decision criteria values for each model.

Figure A2 illustrates the effects of non-linear transformations on model predictions for Experiment 2. As before, the five different models discussed in our paper are simulated for the linear, logarithmic and power law ( $\exp=0.5$ )

transformations. Decision criteria for all models were set such as to produce nearly identical accuracies for the neutral condition (see Table 2).



**Figure A2:** Study 3 simulation results, with linear (top), logarithmic (middle) and power (bottom) transformations applied to the inputs prior to accumulation. Mean inputs to the four choice alternatives were manipulated such that  $I_1=2$  remained constant throughout the simulation while  $I_2$  increased from 1.1 to 1.9 in steps of 0.1 and  $I_3=I_4$  decreased from 1.1 to 0.7 in steps of 0.05. For both the logarithmic and power transformation all input values were increased by 1 and noise was not allowed to drive input values below 1. This was done to avoid irregularities in the transformation for values between 0 and 1.

As can be seen, all model predictions remain qualitatively the same except for the normalized race model which predicts a slight slowdown in the *double boost* condition compared to the *neutral* condition when subject to non-linear transformations. This result is due to the way the non linear transformations affect the ratio between the target and non-target inputs. This ratio was intentionally kept constant in the original, linear, simulation, hence predicting a null effect for the *double boost* manipulation. However, for monotonic, concave functions the ratio shifts in favor of the non-target alternative the stronger the inputs become due to diminishing sensitivity at higher intensities. As a result the input to the target alternative is reduced resulting in the observed slowdown

## Appendix B – Data Fitting Methods

To evaluate the models' ability to account for response time distributions and accuracy of the data we used simulated data from the models and fit it to empirical data. To optimize model parameters and to determine how well the models fit the data we minimized the chi-square statistic (Ratcliff & Smith, 2004; Ratcliff & Tuerlinckx, 2002; REF) through a SIMPLEX algorithm (Nedler & Mead, 1965) as implemented by the 'fminsearch' function in MATLAB.

As a target for the optimization process we calculated the 0.1, 0.3, 0.5, 0.7 & 0.9 RT quantiles and accuracy for each experimental condition (easy/difficult) and for each type of response (correct/error) resulting in 4 quantile sets. For each such set, simulated RT data was then grouped into six bins confined by the empirical RT quantiles. This procedure resulted in 12 bins per condition (6 correct & 6 error). For each condition we calculated the proportion of trials (out of the total number of trials in that condition) that fall within the bounds of each bin. These proportions are equivalent to the probability for a response to end up in a particular bin. The proportions were then multiplied by the probability for a correct response and the probability for an error (1-p(correct)) to create the correct and error proportions respectively. Naturally, for each condition these proportions summed to one.

The chi-square statistic was calculated as the sum, over all bins in all conditions, of the differences between the proportions predicted by the model ( $M_i$ ) and the empirical (target) proportions ( $E_i$ ), divided by the target proportion ( $E_i$ ). Since the RT quantiles that divide the bins are calculated from the empirical data, the target proportions ( $E_i$ ) are always 0.1, 0.2, 0.2, 0.2, 0.2 & 0.1 multiplied by the probability for that type of response (correct/error). For example, if the accuracy in a particular condition was 90% then the target proportions would be 0.09, 0.18, 0.18, 0.18, 0.18 & 0.09 for the correct responses and 0.01, 0.02, 0.02, 0.02, 0.02 & 0.01 for the error responses.

$$Chi = \frac{(E_i - M_i)^2}{E_i}$$

Optimizing the models to predict binned response proportions allows us to test the models on their ability to simultaneously account for RT distributions and accuracies.

### Model parameters:

All the models we used in this study belong to the sequential sampling family and therefore had several common characteristics and assumptions. In the context of simulations and data fitting these shared components translate into model parameters that span across the various model architectures. The most common trait of sequential sampling models is the accumulation of stochastic evidence over time to a response criterion (Cr). The classic diffusion model (Ratcliff, 1979) for two alternative forced choice (2AFC) tasks implements one accumulator that accumulates evidence towards two response criteria, an upper bound and a lower bound. However, since the task in this study is a four alternative forced choice, we implemented two variants of the classic diffusion model that can be easily extended to  $N > 2$  alternatives. These variants, the FFI and the MMN, are, in fact, equivalent to the diffusion model for  $N=2$  alternatives but diverge for  $N > 2$ . The MMN and FFI are more like race models in the sense that the accumulation process is driven by  $N$  accumulators which race towards a common response criteria. Thus *all* our models share the assumption of  $N=4$  accumulators ( $X_1$ ,  $X_2$ ,  $X_3$  &  $X_4$ ) that accumulate evidence towards a common response criteria (Cr). We did not allow individual response criteria to vary

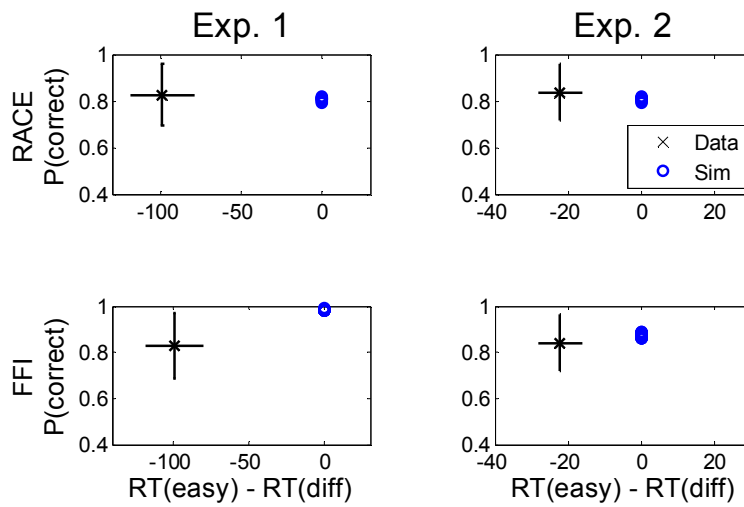
independently for each alternative since the stimulus locations were randomly intermixed and therefore participants had no way of knowing a-priori which location is associated with which alternative.

Another common assumption is that the  $N=4$  accumulators are driven by  $N=4$  separate evidence sources or inputs ( $I_1, I_2, I_3$  &  $I_4$ ). Traditionally, input parameters are left free to vary either separately for each difficulty condition or commonly for manipulations that do not alter the strength of the evidence such as speed vs. accuracy instructions. However, unlike traditional data fitting, one of our main goals was to constrain the models directly to the physical stimuli without the added freedom of fitting the input parameters to the data. During the experiment we had complete control over the exact brightness level presented to the subject on each individual frame and therefore also over the mean and standard deviation of the stimulus brightness. Thus, throughout all the simulations and data fitting in this paper inputs to the models were constrained to the exact magnitudes of the physical stimuli presented to the subjects. More specifically, for the purpose of fitting the results of experiment 1, mean input strengths (or drift rates) for the easy condition were modeled as  $I'_1=0.4$ ,  $I'_2=0.2$ ,  $I'_3=0.2$  &  $I'_4=0.2$  and for the difficult condition as  $I'_1=0.4$ ,  $I'_2=0.3$ ,  $I'_3=0.15$  &  $I'_4=0.15$ . Throughout the course of a simulated trial (i.e. for each time step) the true inputs  $I_i$  are calculated by repeatedly perturbing the mean input rates  $I'_i$  by a Gaussian noise random variable  $\sigma_i = N(0, 0.1)$  that is truncated at  $+0.1$  and  $-0.1$  such that  $I_i = I'_i + \sigma_i$ . Thus,  $\sigma_i$  represents the minimal within trial noise in the process, which is intrinsic to the stimulus and thus external to the processing of the perceptual/decision/response (or any other non-decision component) stage. In addition to the fixed mean stimulus noise variable  $\sigma_i$ , we also added another Gaussian noise random variable  $\epsilon_i = (0, \mu)$  where  $\mu$  is a free parameter. In the models,  $\epsilon_i$  stands for noise that is intrinsic to the process and therefore, represents the stochastic nature of the human cognitive and perceptual processing itself and not the stochastic nature of the stimulus.

In order for sequential sampling models to be able to correctly describe the form of RT distributions it is common practice to add some sources of between trial variability. One source of between trial variability, starting point variability ( $s_i$ ), was applied to all models in the same manner. In all our simulations the accumulation process for each alternative began at an arbitrary baseline activation level of 0.4 to which we added a uniformly distributed random variable  $\gamma_i = U(0, s_i)$ . All models, except for the LCA, also required an additional drift rate variability parameter ( $dv_i$ ) to account well for the form of response time distributions. This was implemented in the models as a Gaussian random variable  $\delta_{ij} = N(0, dv_{ij})$  that is drawn once for each trial  $j$  and added to all drift rates  $I_i$  for the duration of simulation trial  $j$ . The race model predicted distributions that were slightly too symmetrical so following Ratcliff & Smith (2004) we augmented it with a parameter for between trial variability in response criteria  $\lambda$ . Best fit for the race model was achieved when response criteria were drawn from an exponential distribution  $EXP(\lambda)$  with mean  $\lambda$ . The LCA model also required two additional parameters for leak ( $\xi$ ) and inhibition ( $\beta$ ). Finally, we needed to transform all simulated RTs (expressed as discrete time steps) to real world RTs (expressed in milliseconds). To this end, all model outputs ( $RT'$ ) were transformed using a non-decision time parameter ( $T_{nd}$ ) and a time step parameter ( $T_s$ ) such that  $RT = T_s * RT' + T_{nd}$ . The non-decision time parameter ( $T_{nd}$ ) corresponds to the time it takes to process all non-decision components like perceptual and response processes.  $T_s$  is a scaling parameter that transforms decision time from model time steps to milliseconds.

## Appendix C – Simplified Fits for main RT effect and Accuracy

In addition to the regular fits to quantile proportion data, the FFI and race models were also fit to simple mean RT and accuracy data. All fitting methods were identical except for the error function. The error for these fits was calculated as  $((RT\_delta(model) - RT\_delta(data))^2 / RT\_delta(data) + ((ACC(model) - ACC(data))^2 / ACC(data))$  where  $RT\_delta$  is the difference in RT between the *easy* and *difficult* conditions (equivalent to *neutral* and *incongruent* in Experiment2). Results are displayed in Figure C1. As can be seen, neither model can predict a negative effect as is observed in the data.



**Figure C1:** Simplified fits of the race (top row) and FFI (bottom row) models for Experiment 1 (left column) and Experiment 2 (right column). Blue circles denote ten runs of the model with the final parameters. Black X's denote the main RT effect averaged over subjects with error bars denoting two standard errors (horizontal) and  $(ACC * (1 - AC) / N)^{0.5}$  (vertical).  $N=9$ =number of subjects.

## Appendix D – Parameter ranges for DPSP plots

Experiment 1										
Model	parameter	inh	leak	thv	step	delay	SPV	Noise	Cr	dv
Race	min			5	5	150	0	0	10	0
	max			15	20	600	8	2	31	0.1
	best-fit			7.023	12.05	417	5.215	0.364	19.047	0.042
FFI	min				10	200	0	0	6	0
	max				24	600	2	1	14	0.1
	best-fit				18.336	585	0.865	0.52	6.73	0.001
MMN	min				10	200	0	0	2	0
	max				20	1000	4	2	6	0.1
	best-fit				14.64	726	0.915	0.422	3.313	0.013
*100										
LCA	min	0	0		20	700	0	0	0.5	
	max	0.5	0.5		30	1000	0.3	0.5	2.5	
	best-fit	0.391	0.391		28.593	749	0.154	0.189	0.844	

Experiment 2												
Model		DelVar	inh	leak	Cv	step	delay	SPV	Noise	Cr	dv	Boost
Race	min				5	5	100	3	0	15	0	1
	max				15	15	325	15	1	40	0.1	12
	best-fit				11.759	8.221	109	10.072	0.511	35.58	0.02	2.221
FFI	min	50			3	0	2	0	0	5	0	1
	max	250			5	6	4.25	5	1.5	9	0.09	3.5
	best-fit	176			3.424	4.06	409	4.746	0.919	7.139	0.06	1.704
MMN	min	50				1	200	4	0	5	0	0
	max	250				25	400	10	2	13	0.1	5
	best-fit	145				3.332	375	8.36	0.828	10.49	0.02	2.05
LCA	min	100	0.15	0.14		1	200	0	0.3	5.5		1.5
	max	200	0.25	0.16		12	400	5	1.5	7		3
	best-fit	153	0.173	0.149		2.762	318	4.625	0.79	6.635		1.727