Errors, Fast and Slow

Carlos Alós-Ferrer*[1] and Michele Garagnani[2]

[1]Lancaster University Management School, Department of Economics
[2]Department of Finance, University of Melbourne

August 28, 2024

Abstract

Human errors in cognitive, attentional, and decision-making tasks are sometimes faster than correct responses, and sometimes slower. Several existing models can fit response time distributions exhibiting these phenomena. However, it is hard to predict *ex ante* (i.e., before data collection) when errors will be fast or slow. Relying on 20 different datasets comprising 31 experiments from different domains, we empirically validate a simple nonparametric model which successfully predicts when errors will be faster or slower than correct responses. The predictions also include a generalized Stroop effect, as well as error rate differences. The model formalizes how the interaction of multiple processes determines behavior and makes predictions which depend on whether those processes are in alignment or conflict in a given trial, which can be determined before data collection (e.g., congruent vs. incongruent trials in conflict tasks). This yields new testable hypotheses which are overwhelmingly supported in the data. The model's predictions can also be seen as a test of whether process multiplicity is a reasonable assumption in a given task.
Keywords: Errors, Response Times, Multiple Processes, Speed of Errors

*Corresponding author: c.alosferrer@lancaster.ac.uk. The authors thank Sudeep Bhatia, Roger Ratcliff, and Clintin Davis-Stober for helpful comments, and the authors of the 31 experiments we reanalyze for sharing their data.

Table of Contents

1 Introduction **3**

2 Model and Predictions: The Binary Case **6**
2.1 The Basic Model I: Conflict, Alignment, and a Generalized Stroop Effect 7
2.2 The Basic Model II: Errors, Fast and Slow 9
2.3 Extended Model I: Conflict Detection and Non-Decision Time 10

3 Results (Binary Choice) **11**

4 Description of the Studies: Binary Choice **14**
4.1 Cognitive Control 14
4.2 Attentional Processes 18
4.3 Social Cognition 20
4.4 Memory 22
4.5 Decision Making 23

5 Beyond Binary Choice **25**
5.1 Extended Model II: The Multi-Alternative Case 25
5.2 Results (Non-Binary Choice) 27
5.3 Description of the Studies: Non-Binary Choice 28

6 Neither Conflict Nor Alignment: Neutral Trials **30**
6.1 Neutral Trials in the Considered Datasets 30
6.2 Error Rates in Neutral Trials 30
6.3 Response Times in Neutral Trials 32
6.4 Slow Errors in Neutral Trials 32

7 Extended Model III: Option-Dependent Process Response Times **33**

8 General Discussion **35**
8.1 Relationship to Single-Process Models 35
8.2 Other RT Asymmetries 36
8.3 Comparison to Other Dual-Process Models 37
8.4 Conclusions 37

APPENDICES **46**

A Proofs of Theoretical Results **46**
A.1 Prediction D1: Generalized Stroop Effect 46
A.2 Predictions D2 and D2': The Frequency of Errors 47
A.3 Predictions N1, N2, and N2': Neutral Trials 48
A.4 Prediction T2: Slow Errors in Case of Alignment 49
A.5 Prediction T1: Fast Errors in Case of Conflict 49

B Detailed Analysis of Predicted Effects for the Individual Datasets **51**
B.1 Cognitive Control 51
B.2 Attentional Processes 55
B.3 Social Cognition 57
B.4 Memory 59
B.5 Decision Making 62
B.6 Non-Binary Choice 63

C Analysis of Neutral Trials in the Individual Datasets **65**

1 Introduction

A long discussion has examined whether errors are on average faster or slower than correct responses. Empirically, both phenomena are often observed, depending on the specific task, experimental implementation, and type of trials considered (e.g., Laming, 1968; Luce, 1986; Ratcliff and Rouder, 1998, 2000; Ratcliff et al., 2004). Further, both asymmetries can often also be observed within the same task, depending on its implementation. For example, the relative speed of errors might depend on whether response speed or accuracy are emphasized (Swensson, 1972; Luce, 1986; Ratcliff and McKoon, 2008). Even for a fixed task and implementation, the relative speed of errors might also depend on the stimuli. For example, White et al. (2011) showed that errors tend to be slower than correct responses in congruent trials of the Flanker task, but the opposite is true in incongruent ones. Similarly, in a random-dot motion paradigm, Mulder et al. (2012) found slow errors (compared to correct responses) for trials where a previous cue was consistent with stimulus direction, and fast errors for trials where the cue was inconsistent.

The empirical response time patterns comparing errors and correct responses have motivated many theoretical developments in the psychological literature. Some models can accommodate one of the two asymmetries. For example, Poisson parallel-counter models (e.g, Townsend and Ashby, 1983; Smith and Van Zandt, 2000; Townsend and Liu, 2020) predict that errors should always be slower than correct responses. A few models can accommodate both asymmetries. For example, Blurton et al. (2020) show that a Poisson random walk model with random starting points can accommodate fast or slow errors (compared to correct choices). Another prominent example that can fit either pattern is the drift-diffusion model (DDM; Ratcliff, 1978; Ratcliff et al., 2016) with intertrial variability in drift rates and starting points (Ratcliff et al., 1999; Ratcliff, 2002; Ratcliff et al., 2016). With fixed drift rate and symmetric boundaries, the DDM predicts identical response time distributions for errors and correct responses, but intertrial drift-rate variability has been used to accommodate slow errors, while intertrial variability in starting points (or, equivalently, in the asymmetry of the boundaries) can accommodate the opposite pattern (Ratcliff and Rouder, 1998). Similarly, the linear ballistic accumulator (LBA; Brown and Heathcote, 2005, 2008) accommodates fast or slow errors through the interplay between starting point and drift rate variability. It is important to highlight that these models can accommodate fast or slow errors (compared to correct responses) in the sense that they can fit data exhibiting those characteristics. Crucially, however, these models provide no theoretical explanation of why the speed of errors should be asymmetric, and cannot predict when one of the two asymmetries will occur for a fixed, given task (e.g., without changing instructions) before fitting the data.

In this contribution, our objective is different. Models as the DDM and other examples accommodate asymmetries in the relative speed of errors and correct responses in the sense that specific parameter values or realizations can be found which produce one or the other pattern. When applied to existing data, this corresponds to a fitting approach, which seeks to account for empirical patterns by fitting a parametric model to the data. Instead of parametric fitting, our motivation is nonparametric prediction. That is, we focus on the *ex ante* prediction of asymmetries in response times and error rates. Specifically, we propose and validate a simple model which can predict whether errors are faster or slower than correct responses for specific types of trials, but does so before data is collected, and then show its empirical bite in a large variety of tasks. The approach is nonparametric in the sense that none of the predictions depends on

the values of underlying parameters. Thus, we provide a simple model which is able to predict *ex ante* when and why errors are systematically faster or systematically slower than correct choices within the same paradigm. The model is also able to capture other features of the data, as asymmetries in error rates or a generalized Stroop effect.

There are two further differences between our approach and previous work in the literature. The first is that we aim to predict general (parameter-independent), directional effects, as e.g. whether errors are faster or slower than correct responses, rather than studying quantitative relations between certain parameter values and response time distributions. In this sense, the model serves the conceptual purpose of showing the link between a specific theoretical structure (process multiplicity and the concept of process conflict, as explained below) and widespread empirical patterns which transcend any specific task or paradigm. However, this also means that the model is not comparable to parametric models which aim to provide quantitative fits of specific datasets. In particular, it is not possible to ask the question of whether our nonparametric model has a better fit than other, parametric models or not. This is simply not the purpose of our work.

Another important difference is that we focus on predictions showing that errors will be slower than correct responses for certain, pre-defined sets of trials, and faster for other trials. For example, this will typically translate in different predictions for congruent and incongruent trials in conflict tasks, where participants face potentially-conflicting stimuli (e.g., Stroop, Flanker, and Simon tasks). This is in sharp contrast with speed-accuracy effects. As remarked by Luce (1986, Section 6.4.3), errors are often found to be slower than correct responses in experimental treatments where accuracy is emphasized, and faster in other treatments where speed is emphasized or time pressure is implemented (see, e.g., Hawkins and Heathcote, 2021). Our predictions are very different, since they concern the relative speed of errors for different types of trials within the same experimental treatment.

In this work, we show that the directional predictions of the model hold in a wide variety of different contexts. For this purpose, we collected 20 different datasets comprising 31 experiments from the domains of cognitive control, attention, social cognition, memory, and decision making. Those include laboratory, field, and on-line experiments, spanning different levels of cognitive complexity, and involving a heterogeneous array of cognitive processes. None of the datasets was collected with the aim to test our predictions. The tasks covered in the data include conflict tasks as, e.g., Flanker, Stroop, and Simon tasks, but also many other paradigms which are not usually viewed as conflict tasks, e.g., attentional paradigms using Gabor patches, kinematograms, and clouds of moving dots, imitation and perspective-taking tasks, word recognition tasks, false memory paradigms, and decision-making tasks ranging from value-based decisions and probability judgments to the Cognitive Reflection Test. We also go beyond binary choices and include several datasets where more than two alternatives are available to decision makers. The predictions of our model, including the relative speed of errors and correct responses, are overwhelmingly supported in the data.

The model we rely on is a formalized, stylized dual process model assuming that behavior is codetermined by two cognitively different processes (see, e.g., Evans, 2008; Weber and Johnson, 2009, for reviews). Specifically, one process is assumed to be more deliberative, and its modal response should correspond to normatively correct responses. The other process is assumed to be more impulsive or intuitive in nature, and in particular to be faster and closer to a stimulus-response mapping. Both processes are stochastic, in the sense that they do not always select the same choice. Crucially, the more intuitive

process is assumed to react to different cues than the deliberative process. For example, in a Stroop task, the deliberative process would be identifying the color in which a word is printed, while the intuitive process would simply be reading the word. Thus, depending on stimuli, in some trials the typical, modal answer of the intuitive process will be the same as that of the deliberative process (e.g., the word "RED" printed in red), and we say that the processes are *aligned*. In other trials, modal answers will be different (e.g., the word "BLUE" printed in red), and we say that the processes *conflict*. That is, we speak of alignment when the modal responses of the processes coincide, and of conflict if they differ. For some paradigms, alignment and conflict trials correspond to what the paradigm-specific literature calls congruent and incongruent trials, respectively. For other paradigms, as we shall see, the classification is more subtle.

As remarked by Diederich and Trueblood (2018), dual process models are often expressed as verbal theories rather than formalized accounts. Our model, which is based on Achtziger and Alós-Ferrer (2014) and Alós-Ferrer (2018), is a formal model which allows for precise, tractable mathematical derivations and testable predictions. The model, however, is nonparametric in the sense that predictions obtain independently of any fine-tuning or fitting of particular model components. Specifically, the model predicts that errors will be slower or faster than correct responses depending on whether a set of trials captures alignment or conflict, respectively. For many specific paradigms (e.g., Stroop, Flanker, etc.), the candidate processes are clear *a priori*. Thus, which trials correspond to alignment or conflict is also known before data collection, and hence the model delivers clear predictions. Those predictions can also be used as a test of whether in the specific paradigm considered behavior can be explained as the result of the interaction of different processes or not, hence providing a formal test for dual-process effects. In other words, the model is fully falsifiable (since the predictions do not depend on fitting model's parameters), but what is falsified is the joint hypothesis that the model applies and two processes, one more deliberative than the other, codetermine behavior in the considered task.

The intuition for the model's predictions is as follows. For conflict trials, the intuitive process often selects erroneous responses, and hence the average response time of errors is brought down (fast errors). For alignment trials, the intuitive process is a cognitive shortcut which often selects the correct response and does so faster than the deliberative process. By virtue of being closer to a stimulus-response mapping, the intuitive process is also more internally consistent. Hence, conditional on a erroneous response, it is more likely than the decision is made by the deliberative process and is hence relatively slower (slow errors). For example, in the domain of decision making, many decision problems are designed to elicit intuitive but incorrect answers (Cognitive Reflection Test, Base Rate Neglect, Syllogisms, etc.), which are often faster than correct ones (Raoelison et al., 2020). This corresponds to fast errors under conflict. In a different domain, when cueing attention is congruent with a subsequent stimulus (Denison et al., 2018; Hu and Rahnev, 2019; Heathcote et al., 2019), errors are infrequent, but we will show that they are often slower than correct choices. This corresponds to slow errors under alignment.

The model also delivers other directional predictions. In particular, it predicts larger error rates and slower correct choices in conflict compared to alignment trials. The latter is a generalized version of the Stroop effect, which is commonly found in different implementations of this task (Liefooghe et al., 2019) and, as we will show, can also be found in many other tasks. For example, in memory paradigms (Brainerd and Lee, 2019; Charoy and Samuel, 2020) people seem to rely on semantic congruency as a short-cut (a heuristic which can be viewed as an intuitive process) to judge whether a presented word

was among a previously-memorized list or not. This implicit association biases people to wrongly claim word recognition when a new word is semantically similar to others that were actually in the list, resulting in more and faster errors in such (conflict) trials.

The model is kept as simple as possible while generating testable predictions as described above. The framework can be kept nonparametric because the key assumptions of the model are ordinal in nature, e.g. that the intuitive process is on average faster and more internally consistent than the deliberative process. The model, however, could be given specific, parametric microfoundations. For example, for the binary case it could be formalized as a combination of two simple DDMs, each with a fixed drift rate. While the deliberative DDM accumulates toward the correct response, the intuitive DDM accumulates toward the intuitive response, which coincides with the correct one or not depending on whether the trial is in alignment or in conflict, respectively. If the drift rate of the intuitive process is larger in absolute value than that of the deliberative process (hence making it swifter, as assumed in dual-process theories), all assumptions of our model are fulfilled. The predictions derived here will hence hold independently of the specific values of this new model's parameters. The predictions also obtain independently of whether other microfoundations are assumed, and extend to the multialternative case, which would not be covered by this "dual DDM" approach.

The paper is structured as follows. We first introduce the basic model, the assumptions, and its predictions, for the binary case. We also extend the model to account for non-decision times and process selection probabilities depending on conflict vs. alignment. We then briefly show that the predictions are overwhelmingly supported in all the binary-choice experiments that we reanalyze. Then, to clarify the generality of our approach, we give an overview of the experimental designs in the datasets. We start with applications in the domain of cognitive control, followed by tasks involving attentional processes, social cognition, memory, and finally decision making.

After completing the exposition of the binary choice case, we present an extension of the model which makes the analogous predictions for multialternative choice tasks. We then show that those predictions are supported in the corresponding experiments, and then give a more detailed overview of the involved tasks. We then consider predictions and results for a third type of trials often encountered in experimental paradigms, namely neutral trials where the intuitive process should be inactive. Finally, we remark that predictions on the relative speed of errors are obtained even though the model assumes no such differences for the response times conditional on a single, given process. However, we show that the model can be extended to the case where individual processes also display asymmetries as observed empirically for neutral trials.

2 Model and Predictions: The Binary Case

Consider a decision maker facing a task with two possible answers (binary case), a and b. We assume that two different decision processes codetermine behavior, a more deliberative one, D, and a more intuitive/impulsive one, I. Specifically, we are thinking of situations where the researcher has clear hypotheses on which decision processes affect the decision and what their nature is, in terms of (relative) automaticity. Since we will apply the model to a large number of different situations, however, we intentionally keep the setting as abstract as possible at this point.

We assume that all involved decision processes are noisy, and in particular each of the processes (D and I) can select each of the alternatives (a or b) with strictly positive probability. Denote by $P(a|D)$, $P(b|D)$, $P(a|I)$, and $P(b|I)$ the probabilities that each

alternative would be selected by each process, if that process would actually determine behavior. Which of the two processes will actually determine the answer is, however, a stochastic event (possibly reflecting central executive processing). Let $\Delta > 0$ be the probability that the actual response is selected according to the more automatic process I, and $1-\Delta$ the probability that it is selected according to the more deliberative process D. Thus, the actual probability of observing a choice of a is

$$P(a) = \Delta P(a|I) + (1-\Delta)P(a|D).$$

Response times t are also assumed to be stochastic, that is, there will generally be variation across different trials even for the same decision task. In this work, we are only concerned with expected response times and not with distributions. Let $R^D = E[t|D]$ and $R^I = E[t|I]$ denote the expected process response times conditional on the response being selected by the more deliberative or the more automatic process, respectively. For simplicity, we assume that expected process response times do not depend on the actually-selected response (but will weaken this assumption in a latter section). However, as we will argue below, this does not imply that the *observed* response times conditional on one or the other alternative are identical.

For instance, Alós-Ferrer (2018) postulates that each of the processes corresponds to a symmetric drift-diffusion model, with different drift rates μ_D, μ_I, but identical boundaries and diffusion parameters. This is a possible microfoundation of the model, but one could assume any other analytical form for processes D and I instead, as long as the assumptions discussed below are fulfilled. With this formalization, it indeed follows that all choice probabilities are strictly positive and that the response times of individual processes are independent of the selected answer (e.g. Ratcliff and Rouder, 1998; Palmer et al., 2005; Ratcliff et al., 2016).

2.1 The Basic Model I: Conflict, Alignment, and a Generalized Stroop Effect

In all tasks and applications we will consider, one of the options is *correct* in a normative sense, while the other is an error, although, in any given paradigm, whether a or b is correct or an error will generally vary from trial to trial. Let $x^D \in \{a, b\}$ denote the correct answer in the given trial. We assume that the deliberative process corresponds to a noisy version of normative thinking and hence selects the correct answer more than half of the time, $P(x^D|D) > 1/2$. Alternatively, for tasks without an objectively-correct answer, one could simply consider the word "correct" to be void of normative content, in the sense that it describes whatever alternative the more deliberative process selects most often. In other words, if the researcher has a clear candidate for the deliberative process, x^D is simply the modal answer of that process, independently of the interpretation. In this sense, the deliberative process *favors* alternative x^D (meaning it selects it more often).

Denote by $P^D = P(x^D|D) > 1/2$ the probability that D selects its own favored alternative. Analogously, let $x^I \in \{a, b\}$ be the alternative favored by the more automatic process. Whether this process selects a or b more often, i.e. whether $x^I = a$ or $x^I = b$, depends on whether it is adaptive or maladaptive for the particular situation at hand. We speak of *alignment* if $x^I = x^D$, i.e. if both processes favor the same option, and of *conflict* if $x^I \neq x^D$, that is, the processes favor different options. Denote by $P^I = P(x^I|I) > 1/2$ the probability that I selects its own favored alternative.

Some paradigms might include trials where the more automatic process I is not relevant, for instance because the cues that should trigger it are absent. Thus, in trials of this type only process D is active. We call such trials *neutral*. Our predictions will concentrate on conflict and alignment, but we will return to the comparison to neutral trials (for the paradigms which allow it) in a latter section.

We now discuss the assumptions and predictions of the model. Naturally, the more automatic process should be faster in expected terms than the more deliberative one. We thus assume

(R) $R^D > R^I$.

For instance, this follows immediately if both processes are symmetric DDMs as described above with $|\mu^I| > |\mu^D|$, i.e. if the more automatic process is viewed as a swifter one, closer to stimulus-response associations and the more deliberative one is a rule-based, more cognitive process.

The first testable prediction of the model concerns the comparison of conflict and alignment. If the researcher has identified the decision processes of interest, then the respective favored options are known *ex ante*. Thus, within a given experiment, some trials might be in conflict and some might be in alignment, and which are which will be known before data collection. Thus, statements conditional on conflict vs. alignment are testable. The following prediction states that the response time of correct responses must be strictly longer in situations of conflict than in situations of alignment. Notice that this prediction arises exclusively from process multiplicity and assumption (R), and hence it is diagnostic for the presence of multiple processes differing in their degree of automaticity.

Theorem 1. *If (R) holds,*

(D1) correct responses are slower in expectation in case of conflict than in case of alignment.

The intuition for Theorem 1 is as follows (all proofs are in the appendix). Independently of whether a given trial corresponds to conflict or alignment, process D delivers the same proportion of correct responses (x^D), which are relatively slow. In case of conflict, process I favors the erroneous answer $x^I \neq x^D$, and hence typically contributes relatively fewer (fast) correct answers. In case of alignment, process I favors the correct response $x^I = x^D$, and hence typically contributes relatively many (fast) correct answers. Hence, one obtains faster correct responses under alignment than under conflict.

A version of Theorem 1 has been previously discussed in Achtziger and Alós-Ferrer (2014) (in the context of reinforcement vs. normative belief updating in decision making) and Alós-Ferrer (2018) (where the processes were assumed to be DDMs). The prediction of this theorem is a generalization of the well-known "Stroop Effect" (Stroop, 1935; MacCleod, 1991), which describes a slow-down of (correct) responses when one is asked to name the color that a word is printed in but that word happens to name a different color (e.g., "Red" printed in blue) compared to when the word names the color it is printed in (e.g., the word "Red" printed in red). However, this effect is usually attributed to central executive functions of the brain related to the detection and resolution of conflict among elementary responses, which tax cognitive resources and require time (Bargh, 1989; Baddeley et al., 2001), but enable the inhibition of automatic responses in case of conflict. The model presented here does not assume such a difference in response times (although, as will be discussed below, it is compatible with this addition), and Theorem 1 holds in its absence.

The model also makes a straightforward prediction for the proportion of correct responses across conflict and alignment. Note that this result is independent of assumption (R), as it does not involve response times.

Theorem 2. *(D2) The proportion of correct responses is strictly smaller in case of conflict than in case of alignment.*

The intuition for Theorem 2 is immediate. In case of alignment, both processes favor the correct response. In case of conflict, process D favors a correct response, but process I favors an error. Even though each process might still select the option favored by the other rule in case of conflict, it does so less often. Hence, in case of alignment the commonly-favored response is selected more often than any of the individual choices in case of conflict.

2.2 The Basic Model II: Errors, Fast and Slow

Since the more automatic process I is closer to stimulus-response associations, it is also natural to assume that it is less noisy and more internally consistent, that is, it selects the own favored answer x^I more often than the more deliberative process D selects x^D. Thus we assume

(P) $P^I > P^D$.

Again, this follows immediately if both processes are symmetric DDMs as described above with $|\mu^I| > |\mu^D|$, since a larger drift rate (in absolute terms) implies both shorter response times and higher consistency. In case of alignment, this implies that process I is objectively better (in the sense of being both faster and more often correct) than process D, and thus can be seen as an efficient cognitive shortcut. In case of conflict, though, process I's favored answer is actually an error, and the process often leads the decision maker astray.

Even with this simple structure, the model already makes nontrivial predictions for the comparison of response times of errors and correct responses, and specifically predicts a non-trivial interaction between responses and cognitive situations or trial types (conflict or alignment). Specifically, the prediction is that errors will be fast in case of conflict and slow in case of alignment.

Theorem 3. *Assume (R).*

(T1) In case of conflict, the expected response time of errors is shorter than the expected response time of correct answers.

(T2) Assume (P). In case of alignment, the expected response time of errors is larger than the expected response time of correct answers.

The intuition behind Theorem 3 is as follows (again, the formal proof is in the Appendix). The (slow) deliberative process always favors the correct response (by definition), and the (fast) automatic process favors either an error or the correct response depending on whether the trial corresponds to conflict or alignment. Thus, in case of conflict, the former delivers relatively many slow, correct responses and the latter contributes relatively many fast errors, leading to on average faster errors. In case of alignment, the two processes favor correct responses, but by (P) the fast, automatic process contributes more of them than the slow, deliberative one, hence in expected terms correct responses end up being on average faster. In other words, in case of alignment,

process I acts as a quick and efficient shortcut to identify the correct answer while the less-consistent process D contributes relatively more (slow) errors. Hence, conditional on an error being observed, it is more likely that the response is generated by the slower process D.

Formally, it is worth noticing that prediction (T1) does not actually require assumption (P). Previous, more specific versions of Theorem 3 have been presented in Achtziger and Alós-Ferrer (2014) and Alós-Ferrer (2018) (where the processes were assumed to be DDMs). In the present context, the importance of Theorem 3 relies on the fact that it predicts when errors should be expected to be (on average) fast or slow, depending only on observables. Specifically, the prediction will be applicable within any experiment or decision task where the researcher has identified two relevant decision processes, one of which can be reliably assumed to be more automatic than the other, and the experimental design allows to derive the most-frequent (favored) answers for each process. Under these circumstances, the researcher can classify trials *ex ante* into conflict and alignment, and aggregate the data conditional on that classification.

2.3 Extended Model I: Conflict Detection and Non-Decision Time

As observed above, Theorem 1 predicts a slow-down of correct responses under conflict compared to alignment which parallels and generalizes the well-known "Stroop Effect" (Stroop, 1935; MacCleod, 1991), but it does so without assuming that this response-time effect arises due to the involvement of time-consuming central executive functions of the brain related to the detection and resolution of conflict. There is, however, evidence for the latter functions, which have been linked to early activity in the Anterior Cingulate Cortex (see, e.g., Nieuwenhuis et al., 2003; De Neys et al., 2008; Achtziger et al., 2014). Thus, it is worth considering how to extend the model to account for these additional factors.

Let $i \in \{A, C\}$ denote alignment or conflict, respectively, and add a non-decision time t_i to the response time which depends on conflict vs. alignment and is such that $t_C \geq t_A$, thus reflecting a stronger involvement of central executive functions in case of conflict. At the same time, since conflict detection enables the inhibition of automatic responses, an extended model should distinguish the probability of the latter depending on conflict or alignment, i.e. replace the process-selection probability Δ with Δ_i while assuming $\Delta_C \leq \Delta_A$. All our previous predictions hold in this extended model (see Appendix).

Theorem 4. *Consider the extended model and assume (R), (P), $t_C \geq t_A$, and $\Delta_C \leq \Delta_A$. Then (D1), (D2), (T1), and (T2) hold.*

The intuition is as follows. First, predictions (T1) and (T2) are conditional on conflict or alignment, respectively, and thus are unaffected by the distinction between Δ_C and Δ_A or the addition of case-specific non-decision times. The generalized-Stroop prediction (D1) still holds because the effect captured in Theorem 1 and the slowdown implied by $t_C \geq t_A$ go in the same direction. The fact that $\Delta_C \leq \Delta_A$ simply reduces the percentage of correct responses which accrue to the fast, automatic process in case of conflict, and hence also contributes to the overall trend. Prediction (D2) does not involve response times, and thus the only change is the additional assumption that $\Delta_C \leq \Delta_A$. However, since the automatic process is more internally consistent than the deliberative one by (P), in case of alignment the larger probability of the former being selected results in a larger percentage of correct responses, thus confirming the original result. In particular, if (R) did not hold because the two processes are indistinguishable in terms of response

times, $R^D = R^I$, and even if (P) does not hold, (D1) (and of course also (D2)) is still predicted as long as $t_C \geq t_A$ holds strictly.

Corollary 1. *Consider the extended model and assume $R^D = R^I$, $t_C > t_A$, and $\Delta_C \leq \Delta_A$. Then (D1) and (D2) hold.*

This extension of the model is straightforward. However, it disciplines the results in sensible ways. For instance, an analogous proof to that of Theorem 1 shows that the expected response time of errors in case of conflict is strictly shorter than the expected response time of correct responses in case of alignment. However, this prediction does not necessarily hold in the extended model, since non-decision times are longer in case of conflict and hence the comparison of total response times would be undetermined. Thus, even though one could have formulated this additional prediction in the original model, we consider it unwarranted in general in the sense that it would not survive natural extensions.

3 Results (Binary Choice)

We examined the recent psychological literature and collected 20 datasets (total $N = 2,792$; 22 individual publications) involving 31 different experimental tasks which can be described in terms of dual, interacting processes and for which (i) choice and response time data was available, and (ii) conflict and alignment could be identified in the dataset. None of the studies was designed to test the predictions of our model. Some of the datasets combine several studies using similar tasks, and some of them contain data on a single study employing several experimental tasks. Hence, we will discuss (and number them) in terms of the paradigms (Dataset 1 to Dataset 21). In several cases, the studies include treatments warranting separate analysis; in those cases, we differentiate the data by adding a letter to the numeral, e.g. 1a–1c.[1] This results in 31 different tests of each of our predictions.

The first six columns of Table 1 list the dataset, research article or articles (authors, year, journal), domain, and broad type of experimental task used (for the treatment if appropriate). The Section "Description of Studies: Binary Choice" below briefly describes each study and how it is encompassed by our model. Five of the datasets (1–5) belong to the area of cognitive control and include Stroop, Simon, and Flanker tasks, plus a Cued Flanker variant and a Hybrid Stroop-Simon design. The next four (6–9) concern attention and cover discrimination tasks employing Gabor patches, kinematograms, random-dot motion, and large flickering checkerboards. Three further datasets (10–12) arise from studies in social cognition and include automatic imitation of bodily gestures and two forms of perspective taking (involving numerosity and spatial orientation, respectively). The next three (13–15) concentrate on memory and involve word recognition, memory associations (conjoint recognition), and availability. Two further datasets (16–17) cover decision-making tasks, namely syllogistic reasoning, base-rate probabilistic judgments, and value-based decisions. Datasets 18–21 concern non-binary choice tasks and are discussed in the Section "Beyond Binary Choice" below (column "Bin." in Table 1 shows whether the dataset involves binary choices or not). Some of the datasets include neutral trials (column "Ne."), which are analyzed in a later section.

[1] Dataset 9 consists of the data of three previous publications, which employ closely-related paradigms and are analyzed as 9a–9c. A third task in Dataset 16 was different enough (different paradigm with non-binary choice) to label it as a separate dataset (21).

D	Authors	Year	Journal	Domain	Task	Bin.	Ne.	N	D1	D2	T1	T2
1a	Liefooghe et al.	2019	JEP: LMC	Cogn. Control	Stroop	Y	N	275	***	***	***	***
1b					Reinforced Associations	Y	N	(275)	***	***	***	***
1c					Derived Associations	Y	N	(275)	***	***	*n.s.*	*n.s.*
2	Gyurkovics et al.	2020	JEP: General		Simon Task	Y	N	108	***	***	*n.s.*	***
3	Weissman	2019	JEP: LMC		Hybrid Stroop-Simon	Y	N	90	***	***	***	***
4	Luna et al.	2020	JEP: HPP		Flanker	Y	N	92	***	***	*	***
5a	White & Curl	2018	Comp.Brain&Beh.		Cued Flanker: No cue	Y	Y	123	***	***	***	*n.s.*
5b					Alerting cue	Y	Y	(123)	***	***	***	***
5c					Orienting cue	Y	Y	(123)	***	***	***	*n.s.*
6	Denison et al.	2018	PNAS	Attention	Gabor Patches	Y	Y	12	***	***	***	***
7	Evans et al.	2017	Sci.Reports		Kinematogram	Y	N	70	***	***	***	***
8a	Heathcote et al.	2019	J.Math.Psy.		Checkerboards: Prob.	Y	Y	32	***	***	***	***
8b					Checkerboards: Reward	Y	Y	(32)	*	**	*n.s.*	***
9	Hu & Rahnev	2019	Cognition									
9a	Bang and Rahnev	2017	Sci.Reports		Gabor Patches	Y	Y	30	***	***	*	***
9b	de Lange et al.	2013	J.Neurosci.		Moving Dots	Y	Y	22	***	**	***	***
9c	Rahnev et al.	2011	J.Neurosci.		Moving Dots	Y	Y	21	***	***	***	***
10a	Ramsey et al.	2019	JEP: HPP	Social Cogn.	Imitation: Low Load	Y	N	172	***	***	***	***
10b					Imitation: High Load	Y	N	(172)	***	***	***	***
11	O'Grady	2020	Quart.J.Exp.Psy.		Perspective Taking	Y	N	30	***	***	**	***
12	Muto et al.	2019	Cognition		Perspective Taking	Y	N	36	***	***	***	*n.s.*
13a	Charoy & Samuel	2020	JEP: LMC	Memory	Word Recognition, Exp.1	Y	N	27	***	***	***	***
13b					Word Recognition, Exp.2	Y	N	(260)	***	***	*	***
14	Brainerd & Lee	2019	JEP: LMC		Conjoint Recognition	Y	N	185	***	***	***	***
15	Glöckner & Bröder	2014	J&DM		Availability Judgments	Y	Y	61	***	***	***	***
16a	Raoelison et al.	2020	Cognition	Dec. Making	Syllogisms	Y	Y	260	***	***	***	***
16b					Base-Rate Questions	Y	Y	(260)	***	***	*	**
17	Fontanesi et al.	2019	Psych.Bull.&Rev.		Value-Based Decisions	Y	Y	27	***	***	*n.s.*	***
18	Steyvers et al.	2019	PNAS	Cogn. Control	Moving Leafs	N	Y	1000	***	***	***	***
19	Dignath et al.	2019	JEP: HPP	Cogn. Control	Stimulus Recognition	N	N	87	***	***	***	***
20	Schmidt & Weissman	2014	PLoS One	Cogn. Control	Prime-Probe Arrow	N	N	32	***	***	***	***
21	Raoelison et al.	2020	Cognition	Dec. Making	Cognitive Reflection	N	Y	(260)	**	***	**	*n.s.*

Table 1: List of datasets (D) analyzed in the manuscript, and summary of the Wilcoxon signed-rank tests (WSR) testing the four predictions of the model. "Bin." indicates whether the task is binary and "Ne." whether it includes neutral trials (both Y/N). N is the dataset size (number of participating subjects; numbers in brackets indicate the same subjects as in other tasks within an experiment). *** $p < .01$, ** $p < .05$, * $p < .1$.

Predictions (D1), (T1), and (T2) involve comparisons of response times, while prediction (D2) compares proportions of correct answers. In all cases, normality assumptions are unwarranted, since proportions belong to the interval $[0, 1]$ and response times are nonnegative. All four predictions, however, can be tested for each individual study or treatment by means of non-parametric Wilcoxon signed-rank (WSR) tests, which do not require distributional assumptions. The last five columns of Table 1 display the size of the dataset (N) and the significance level reached by the corresponding WSR test for each of the four predictions, for a total of 124 tests. The Appendix contains details of the individual tests, robustness analyses, and further comparisons as appropriate for each dataset. As can be seen from Table 1, all four predictions enjoy overwhelming support across the datasets, with the vast majority of tests (104 of 124) being significant at the 1% level. Also, in a few cases where some results are not significant (1c, 8b), those are associated with particularly-weak manipulations (see comments for the individual studies in the next section).

Nonparametric tests, however, lose the information contained in the cardinal variables we use (response times and proportion of correct answers). To recover this information and better illustrate the results, the left-hand side of Figures 1–4 uses a forest plot representation to display the actual difference in response times or proportions of correct answers as given in predictions (D1), (D2), (T1), and (T2), for all 31 studies and treatments. The figures include the traditional 95% confidence intervals (assuming a normal distribution of the difference variables for illustration), and a vertical line at zero for ease of interpretation. Figures 1, 3, and 4 refer to predictions (D1), (T1), and

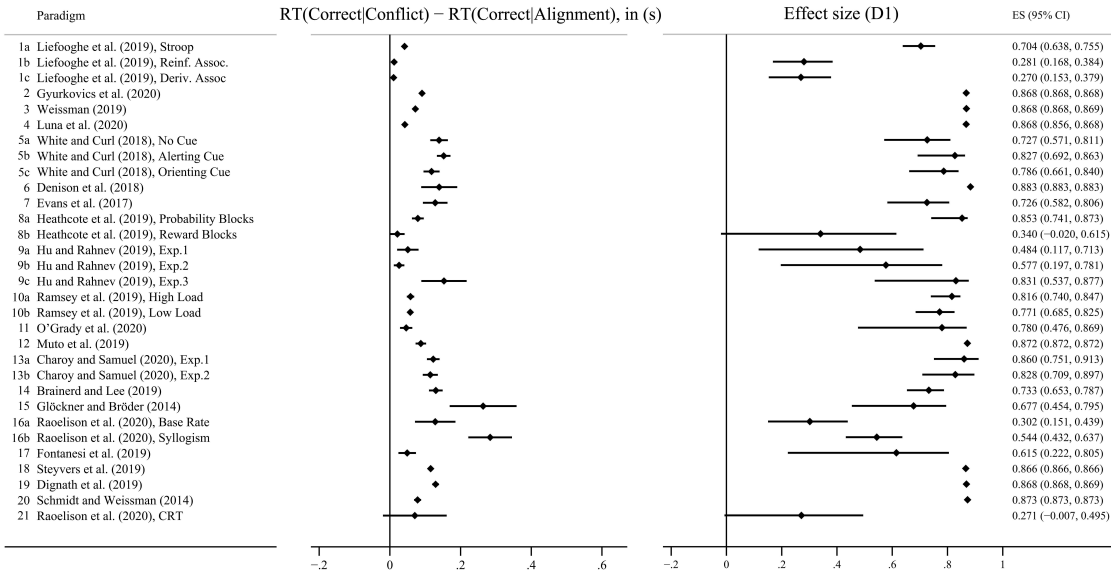

Figure 1: Prediction (D1) across all 31 studies/treatments. Response time of correct choices in conflict vs. alignment and relative effect sizes for the non-parametric tests.

(T2), respectively, and hence the horizontal axis of their left-hand sides corresponds to time differences and is measured in seconds. Figure 2 refers to prediction (D2) and hence the horizontal axis of its left-hand side corresponds to differences in proportions (-1 to 1). In all cases, the representation allows to see the absolute size of the differences in response times or proportions and compare them across studies.

To demonstrate the magnitude of the effects, the right-hand side of Figures 1–4 displays the effect sizes and corresponding 95% confidence intervals for the WSR tests in Table 1. The effect size of a WSR test, r, is considered small for $r \in [0.1, 0.3]$, medium for $r \in [0.3, 0.5]$, and large for $r > 0.5$ (Cohen, 1988; Rosenthal, 1994). To obtain confidence intervals for our non-parametric tests, we follows the bias-corrected-and-accelerated (BCa) bootstrap method of Efron (1987) (see also Kirby and Gerlanc, 2013, for details). The sample sizes of the datasets we consider are generally enough to provide precise confidence intervals (Algina et al., 2006). We set the number of resamples to 5,000, far in excess of the minimum number of 999 recommended in Davison and Hinkley (1997). We remark that, if there is no ambiguity in the test, in particular if all signs in a non-parametric test are positive (or negative), then one obtains point estimates for the effect size instead of a proper interval. This occurs for 9 tests in our dataset (5 of them concerning (D1)).

The figures further illustrate that the datasets lend overwhelming support to our predictions. Figures 1 and 2 summarize the magnitude of the effects for predictions (D1) and (D2), respectively. They show that the RT of correct answers and the error rates are systematically larger under conflict than under alignment, with generally large effect sizes. Figures 3 and 4 display the corresponding results for predictions (T1) and (T2) and show that the expected asymmetry is strongly supported for the data. That is, errors are almost universally faster than correct responses under conflict but slower under alignment, with mostly medium to large effect sizes in both cases.

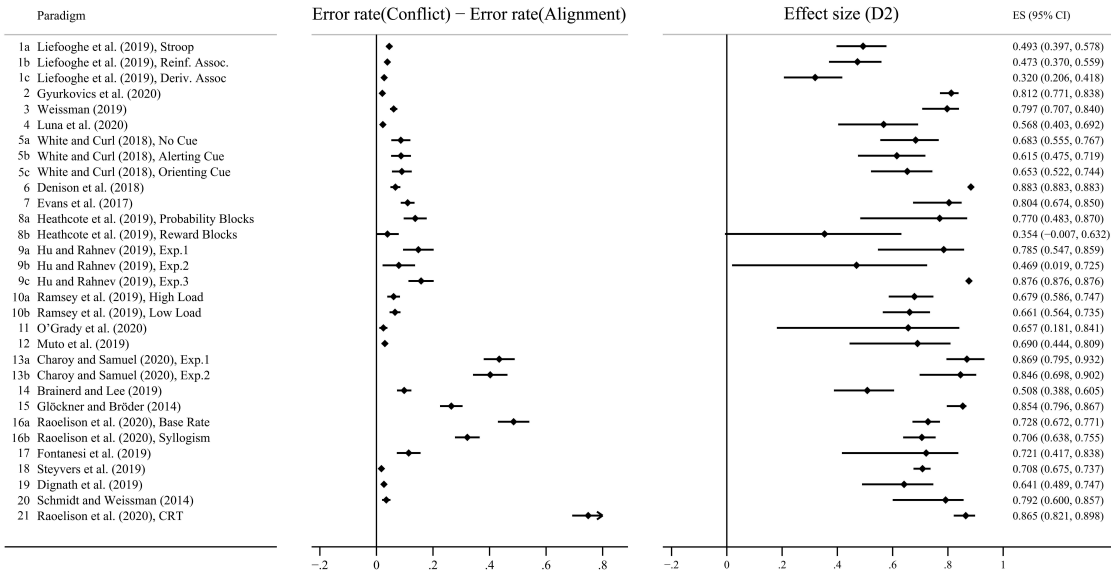

Figure 2: Prediction (D2) across all 31 studies/treatments. Error rates in conflict vs. alignment situations and relative effect sizes for the non-parametric tests.

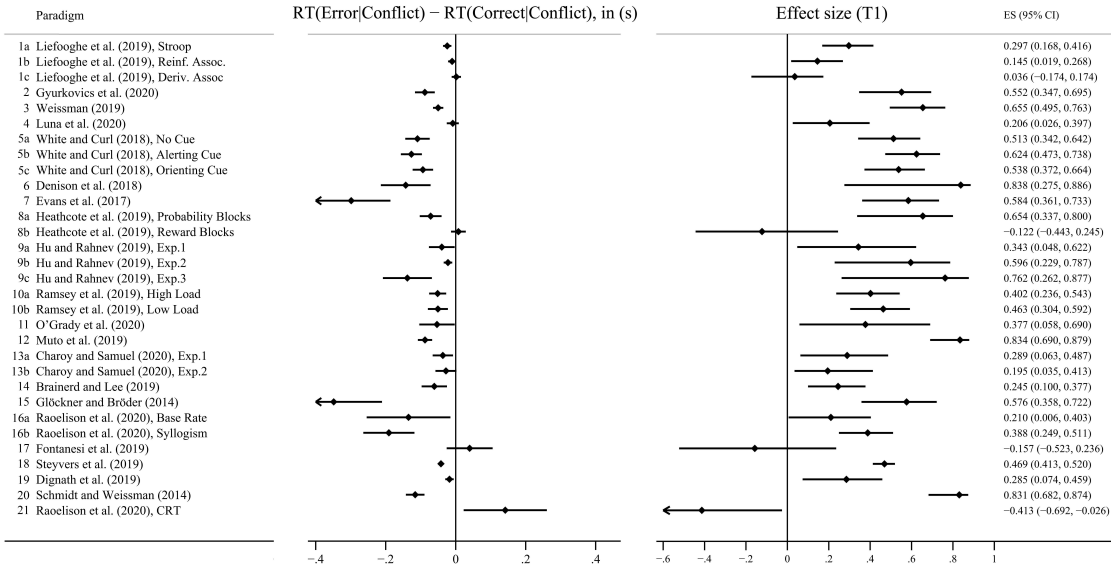

Figure 3: Summary of all the paradigms. Response times of errors vs. correct choices in conflict situations (T1).

4 Description of the Studies: Binary Choice

This section briefly describes the experimental tasks in each of the 21 datasets and how they are encompassed by our model. Details on the individual tests and further robustness analyses are in the Appendix.

4.1 Cognitive Control

Dataset 1: Stroop Effects and Derived Associations. Stroop-like effects can be induced through direct reinforcement of stimuli (i.e., Shiffrin and Schneider, 1977), but it is less clear whether derived, indirect associations (Sidman and Tailby, 1982) produce

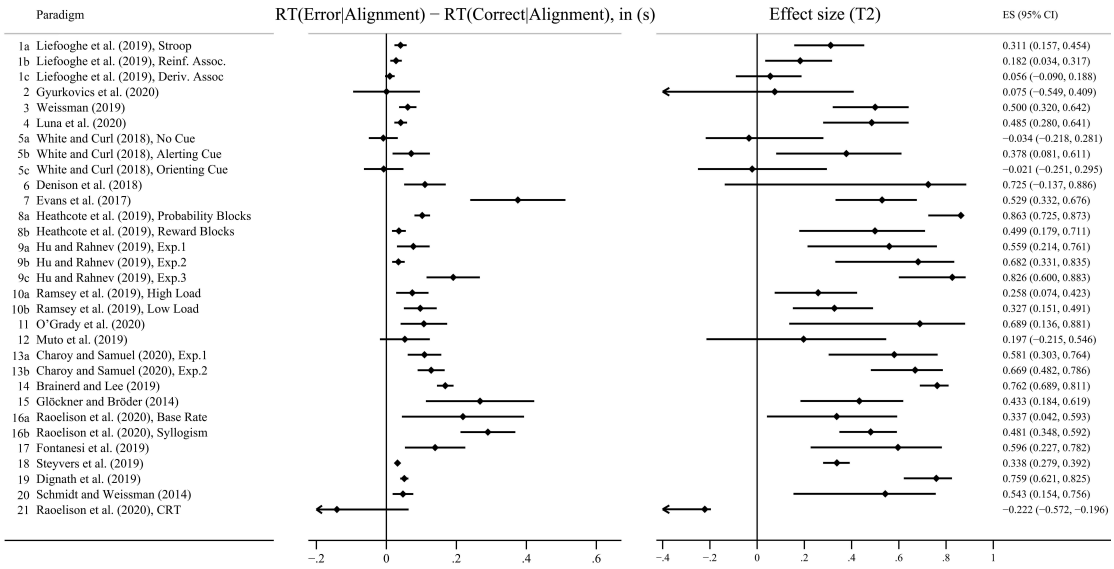

Figure 4: Summary of all the paradigms. Response times of errors vs. correct choices in alignment situations (T2).

similar effects. Liefooghe et al. (2019) showed that Stroop-like effects can be induced through *automatized* processes arising from either directly or indirectly reinforced associations. In five studies ($N = 275$; pooled for our analysis), participants were trained to establish and derive an association between a non-word string of characters and a color-naming word. The non-words, together with actual words, were then used as distractors in a Stroop task (Stroop, 1935). For *reinforced associations*, some non-words were directly reinforced, establishing an association to a color. For instance, participants were trained to select PLESK in the presence of RED and KLAMF in the presence of GREEN. For *derived associations*, other non-words were *indirectly* associated to a color. For instance, participants were trained to select SMELK in the presence of PLESK and GILPT in the presence of KLAMF. A Stroop task followed the association training phases, which used either color names, reinforced associations, or derived associations as distractors. For instance, the word GILPT (associated with KLAMF, in turn associated with GREEN), could be printed in red in an incongruent trial. For each type of stimuli, half of the trials were congruent and the other half were incongruent. An answer is correct if the participant reported the print color of the word, and an error otherwise. In each trial, there were always only two possible answers.

In our terms, the deliberative decision process is to assess the print color of the word. The alternative process in this task is to indicate the meaning of the word, which might be a fully automatic answer (for color names), an automatized one (for reinforced associations), or an indirect, presumably multi-step one (for derived associations). Thus, congruent trials were in alignment and incongruent trials in conflict, although the cognitive characteristics of the involved alternative process obviously differ depending on the distractor.

All predictions hold for normal Stroop trials and also for trials with reinforced associations, demonstrating that the effects can arise with automatized processes as well as with automatic ones. The predicted relations only hold partially for trials with derived associations, where the automatic nature of the alternative process is less clear. Recall that Corollary 1 predicts D1 and D2 (but not T1 and T2) if two different processes are

at work but they do not differ in terms of response times (or internal consistency). This is exactly what happens in this case. Specifically, D1 and D2 hold, but T1 and T2 do not.

Dataset 2: Simon Task. Gyurkovics et al. (2020) investigated how the ability to dynamically adjust cognitive control develops during adolescence. Their research tested the predictions of multiple models of human development (see, e.g., Casey et al., 2005; Steinberg, 2008) which state that cognitive control is still maturing and improving in this stage of life. Participants from four age groups ($12-13$, $14-15$, $18-20$, and $25-27$ years old; $N = 118$) completed variants of the Flanker, Simon, and Go/No-Go tasks. The latter does not deliver response times in the No-Go case, and error rates in the Flanker task were extremely small (less than 0.5% and 1.0% in congruent and incongruent trials, respectively). We reanalyze data from their Simon task, where, interestingly, Gyurkovics et al. (2020) found no substantial changes through adolescence. In this task, participants had to identify the direction that a single arrow was pointing to (left, right, up, or down). The arrow could be presented either above, below, to the left of, or to the right of the center of the screen, but this location was not relevant for the answer. Trials were congruent if the location and the direction coincided, e.g. an arrow pointing right located to the right of the screen, and incongruent otherwise (those were constrained to left-right or up-down cases). An answer is correct if the participant indicated the direction given by the arrow, and an error otherwise.

The deliberative decision process is to indicate the direction actually given by the arrow, while the alternative process is to report the location of the target instead of its direction. In our terms, congruent trials were in alignment, and incongruent trials in conflict. All errors in conflict were intuitive, that is, they followed the location of the arrow instead of its direction (nobody reported "up" on seeing a right-pointing arrow on the left of the screen).

All predictions hold for the Simon task but for response times in alignment (T2). This might be because in alignment the overwhelming majority of participants had zero error rates and hence the sample size for this test is greatly reduced. The comparison testing T2 still shows that errors are slower than correct answers in this case, but it is not statistically significant.

Dataset 3: Hybrid Stroop-Simon Task. There is an active, ongoing debate on whether cognitive control is domain-specific (i.e., Egner, 2008) or domain-general (i.e., Botvinick et al., 2001). Weissman (2019) introduced a hybrid Stroop-Simon task to address this question, and argued in favor of a domain-general interpretation. In this hybrid task, ninety participants (in two pooled experiments) were presented with one of four words naming colors (red, blue, green, and yellow) and were told to identify the color in which the word was printed, while ignoring both its location and meaning. Participants provided answers by pressing one the (color-coded) keys left (red), right (blue), up (green) and down (yellow), which matched the four locations where the stimuli could appear on the screen. Stimuli were paired in the sense that the four combinations of the words red and blue and the print colors red and blue always appeared left or right, and the four combinations of the words green and yellow and the print colors green and yellow always appeared up or down (16 total possible stimuli). An answer is correct if the participant indicated the key corresponding to the print color of the word, and an error otherwise. Due to the pairing of stimuli, all errors in conflict were intuitive, that

is, they followed the location of the stimuli or the meaning of the word (there were no up/down responses for red-blue stimuli or left-right responses for green-yellow stimuli).

The normative decision process is to press the key corresponding to the color of the word, ignoring its position and its meaning. However, there are two alternative (intuitive) processes in this task, corresponding to the meaning of the printed word (Stroop) or its position on-screen (Simon). Trials can be of four types in this hybrid task. Full alignment corresponds to trials where both the on-screen position and the word's meaning align with the print color. Full conflict corresponds to trials where the on-screen position and the word's meaning are aligned, but contradict the print color. Simon conflict corresponds to trials where the on-screen position contradicts the print color, but the latter is aligned with the word's meaning. Last, Stroop conflict corresponds to trials where the word's meaning contradicts the print color, but the latter is aligned with the on-screen position. Note that, in this sense, Simon conflict is Stroop alignment and Stroop conflict is Simon alignment.

The predictions of the model only apply when the two alternative intuitive processes are aligned, and hence can be summarized as one. This corresponds to full alignment and full conflict trials, where we find full support for our hypotheses. For Simon-conflict and Stroop-conflict trials, the model does not apply a priori, since conflict with one alternative process is actually alignment with the other one. Although we had no predictions for those situations, it is still interesting to examine them (see Appendix for details). Simon conflict (hence Stroop alignment) trials behave as one would expect for the case of conflict. These results suggest that the process underlying the Simon effect might be dominating the one responsible for the Stroop effect.

Dataset 4: A Standard Flanker Task The cognitive interference literature has provided mixed evidence on whether simultaneously performing several tasks has a detrimental effect on cognitive control (i.e., Lavie et al., 2004; Salvucci and Taatgen, 2008) or, on the contrary, it can boost performance (i.e., Kim et al., 2005; Gil-Gómez de Liaño et al., 2016). To investigate this issue, Luna et al. (2020) study an otherwise-standard Flanker task with concurrent working memory load, implemented by directing attention to infrequent, displaced on-screen stimuli, and show that interference can both increase and decrease cognitive performance depending on the attentional set. In this task, ninety-two participants (belonging to three similar experiments) had to detect the direction of a central arrow (left/right), flanked by two distracting arrows on each side. In congruent trials, the target and Flankers pointed in the same direction, while in incongruent trials they pointed in opposite ones. In around one third of the trials, the target was slightly displaced either horizontally (leftward/rightward from the central position) or vertically (upward/downward).

An answer is correct if the participant indicated the direction given by the central arrow, and an error otherwise. The deliberative decision process is to indicate the direction of the central arrow, while the alternative process is to report the direction of the Flankers. Congruent trials were in alignment, and incongruent in conflict. Our predictions find full support in this context, independently of the working memory load manipulation.

Dataset 5: Attention and the Flanker Task Smith et al. (2004) show that providing cues can increase the speed of responses and reduce interference from irrelevant stimuli. Following this result White and Curl (2018) study a cued version of the Flanker task (the Attentional Network Test) and show that alerting cues lead to faster encoding,

improved perceptual processing, and increased attentional focusing. Their experiment ($N = 123$) relied on a Flanker task where a central, target arrow might point left or right, and four Flankers (two on each side) might point in the same direction as the target (congruent trials), the opposite direction (incongruent trials), or be absent entirely (neutral trials). The stimuli randomly appeared either up or down, that is, either in the upper or the lower part of the screen. To prompt attention, each trial was preceded by one of three cueing conditions: no cue, an alerting cue in the form of two centrally-positioned asterisks, one up and one down, or an orienting cue in the form of a centrally-placed asterisk shown either up or down, marking the part of the screen that the stimuli would later appear on.

An answer is correct if the participant indicated the direction given by the target stimuli, and an error otherwise. The deliberative decision process is to indicate the direction of the target stimuli, while the alternative process is to report the direction of the Flankers. Congruent trials, where the target and Flankers point in the same direction, were in alignment, incongruent trials were in conflict, and neutral trials correspond to our definition of neutral, where the alternative process is not triggered at all (we will return to neutral trials in a later section). Our hypotheses find full support for the alerting cue condition. In the no cue and orienting cue conditions, all predictions are supported with the exception of (T2), where differences are not significant.

4.2 Attentional Processes

Dataset 6: Attention and Perceptual Decisions Recent work has suggested that optimal choice thresholds in categorization tasks might be adjusted as a function of the uncertainty in the prior distribution (Qamar et al., 2013; Adler and Ma, 2018). Specifically, it has been suggested that humans take into account both sensory measurements and the associated, underlying uncertainty (Knill and Richards, 1996). In support of this view, Denison et al. (2018) showed that perceptual decisions in natural vision are improved by adjusting for attention-dependent uncertainty. In their experiment, $N = 12$ participants were shown drifting Gabor patches and had to decide whether they had been sampled from a narrow Gaussian distribution (mean 0°, $SD = 3°$) or a wider one (with the same mean but $SD = 12°$). They were previously trained to recognize the two categories with an accuracy of at least 70%. The normatively optimal answer, derived from Bayes' rule, is to report the narrow category for stimuli in the interval $[-5.16°; 5.16°]$, and the wide category otherwise (this yields the maximum obtainable accuracy of 80%). We say that an answer is correct if it followed this criterion, and an error otherwise.

In each trial of the actual task, participants were shown four Gabor patches simultaneously for 300 ms, drawn independently and with equal probability from one of the two distributions. After the patches disappeared, one of the four locations (selected randomly) was indicated with a fixation cross and the participant had to categorize the corresponding patch. Attention was manipulated on a trial-by-trial basis by using a cue *before* the patches appeared. In 2/3 the trials, the cue matched the actually-relevant patch (*valid* cue). In 1/6 of the trials, the cue was misleading and pointed to an irrelevant patch (*invalid* cue). In the remaining 1/6 of the trials, there were four cues marking all four locations (*neutral* cues).

The deliberative decision process in this case would be to retrieve the actually-relevant stimulus from memory and categorize it, since the cue is *ex post* irrelevant. This is clearly an effortful strategy heavily relying on working memory. An alternative, simpler decision process would specify to focus on the attentional cue and categorize

the corresponding patch. Reliance on this latter process might be seen as adaptive and efficient, as subjects knew that in the majority of trials the cue would be valid.

In our terms, trials with valid cues are always in alignment, as both decision processes make the same prescription. Invalid trials are in conflict, unless, by chance, both the cued patch and the actually-relevant one would be categorized identically. Thus, we define conflict trials as those where the cue was invalid *and* the two processes actually made different prescriptions. Those are 6.08% of all trials. Further, trials with neutral cues are also neutral in our terms, since the simpler process is not actually triggered. Our predictions find full support in this context.

Dataset 7: Perceptual Decisions and Initial Cues Evans et al. (2017) show that decision processes in perceptual decision making show little evidence of decay in evidence accumulation and are hence more consistent with received drift-diffusion models than with alternative ones where decisions are based on only the most recent evidence (Kiani et al., 2008; Tsetsos et al., 2012). Seventy participants made decisions about the direction of motion (left or right) in a random dot kinematogram (RDK). The stimuli included a brief initial pulse of motion, which was in the direction opposite to the subsequent motion for half of the trials (incongruent) and in the same direction for the rest (congruent trials), but the answer was provided after the kinematogram's movement stopped. An answer is correct if it matched the general motion in the RDK (left or right), and an error otherwise.

The deliberative decision process is to assess the general motion in a Bayesian way, which, since the initial pulse was very brief, should coincide with the direction of movement in the rest of the trial. An alternative process is to report the direction of movement of the initial pulse. This is akin to conservatism in belief updating tasks, since the initial pulse can used to formulate a prior, and conservatism dictates to disproportionally (over)weight the prior. In our terms, congruent trials are in alignment, and incongruent trials in conflict. Our predictions are again fully supported in this context.

Dataset 8: Judging Majority Colors Heathcote et al. (2019) study how beliefs and utilities contribute to the formation of response bias following hypotheses derived from Vickers (1979) as well as using Bayes factors as prescribed in Prince et al. (2012) and (Davis-Stober et al., 2016). Participants ($N = 32$) were presented with large (32×32) checkerboards filled with squares of two possible colors (blue and orange), whose positions changed 20 times per second. They had to indicate the majority color (displayed on either 52% or 54% of the squares). Trials were in three types of blocks. In probability-manipulation blocks, participants were given a prior in the form of which color would correspond to the majority in most (75%) trials. In reward blocks, majority colors occurred equally often, but one color was rewarded with three times as many points as the other if the answer was correct (however, points did not influence rewards, as participant remuneration was not performance-based). In unbiased blocks, majority colors occurred equally often and both colors were associated with the same number of points. An answer is correct if the participant reported the actual majority color for the trial, and an error otherwise.

The deliberative process is to assess which is the actual majority color (which might be particularly noisy in this paradigm). In probability-manipulation blocks, an alternative process is conservatism, which simply focuses on the (asymmetric) prior and reports the modal color in that prior. In reward blocks, an alternative process might be to focus on the most-rewarded color, although this might be speculative since rewards were

hypothetical. In unbiased blocks, there are no candidates for alternative processes. In our terms, trials in unbiased blocks are neutral (and we will return to them in a later section). In probability-manipulation blocks, trials where the actual majority color coincides with the most frequent according to the prior are in alignment, and other trials are in conflict. In reward blocks, trials where the actual majority color coincides with the most-rewarded color are in alignment, and other trials are in conflict.

For probability-manipulated blocks, our predictions find full support. For reward blocks, all predictions are supported except for T1. Since the experiment was not actually incentivised, this might suggest that differences in the magnitude of hypothetical rewards might have played a modest role.

Dataset 9: Predictive Cues and Categorization A large literature has studied response bias in categorization tasks, and specifically how responses depend on individual stimulus sensitivity and the characteristics of the stimuli (i.e., Rahnev and Denison, 2018; Summerfield and De Lange, 2014; Wexler et al., 2015). Hu and Rahnev (2019) show that predictive cues are able to reduce but not to completely eliminate intrinsic response bias in (perceptual) categorization judgements. For this purpose, the authors re-analyze data from three previous experiments (thirty, twenty-two, and twenty-one participants, respectively). In the first experiment (Bang and Rahnev, 2017), subjects indicated whether the overall direction of a series of 30 briefly-presented Gabor patches, with orientations sampled randomly from a normal distribution, had an overall tilt to the left or to the right from the vertical. Two thirds of the trials included a cue (either before or after the main stimulus) indicating which direction was more likely (66.67% probability). In the second experiment (de Lange et al., 2013), subjects indicated the direction of motion (either contracting or expanding) of white dots presented on a black annulus. In half of the trials, cues signaling contraction or expansion were presented. The predictive cues correctly indicated the upcoming motion direction on 75% of the trials. The remaining trials included no cue. In the third experiment (Rahnev et al., 2011), subjects indicated the direction of motion (either left or right) of white dots presented on a black annulus. Two thirds of the trials included cues, which were valid 75% of the time.

In all three experiments, an answer is correct if the participant recognized the direction of the stimuli, and an error otherwise. The deliberative decision process in this case is to accumulate the evidence provided by each patch and extrapolate the overall direction. An alternative process is simply to follow the cue. All trials were cues were valid are in alignment, trials with invalid cues, where the cue indicated the wrong direction, are in conflict, while trials without cues are neutral situations (we will return to those in a later section). Our predictions are confirmed in all three experiments.

4.3 Social Cognition

Dataset 10: Automatic Imitation of Social Gestures Widespread evidence indicates that imitative tendencies are highly automatic, especially when they refer to mimicking bodily gestures of others (see Cracco et al., 2018, for a meta analysis of 226 experiments). Ramsey et al. (2019) further tested the assumption that the mental processes underpinning imitative behavior are relatively automatic by showing that they are unaffected when subjects are placed under cognitive load. In three experiments (58, 55, and 59 subjects, respectively), participants were required to horizontally lift either the index or the middle finger of their right hands in response to seeing the number "1" or "2" on screen, respectively. The numbers were displayed concurrently with a (mirrored)

hand with either finger lifted. An answer is hence correct if the finger corresponding to the displayed number was lifted, and an error otherwise. The task took place during the retention interval of a cognitive load manipulation (memorize an image and report it later), which used different stimuli in each experiment.

The deliberative decision process in this setting is to lift the finger corresponding to the displayed number. The alternative process, which can be assumed to be highly automatic in view of the literature, is to spontaneously imitate the movement of the displayed hand. In our terms, trials where the displayed number and the displayed hand prescribed the same movement (e.g., a lifted index finger next to a "1") are in alignment, while trials where they prescribed different movements (e.g., a lifted middle finger next to a "1") are in conflict. Our predictions are also confirmed in this setting, independently of the cognitive load treatment.

Dataset 11: Perspective Taking (Numerosity) A recent literature has investigated the mentalizing processes behind perspective taking (i.e., Conway et al., 2017; Freundlieb et al., 2016, among others). Among these contributions, O'Grady et al. (2020) show that humans acquire others' perspectives quickly, unconsciously, and involuntarily, but argue that perspective-taking is not completely automatic in the sense of being purely stimulus-driven. We reanalyze data from the "explicit" condition in their Experiment 1 $N = 30$), which contrasted egocentric and altercentric perspectives (the other conditions and experiments lacked this contrast and cannot be analyzed in our terms). In each trial, participants first observed a single-digit number, and then a screen containing a human-like avatar, a set of red balls, and some Lego blocks which might block the avatar's view. Then they were asked whether a previously-seen number matched the number of balls, with some trials referring to the balls visible to the participant, and the rest referring to the balls visible to the avatar. In some trials (congruent), the avatar and the participant could see the same number of balls. In other trials (incongruent), the line of sight of the avatar was partially blocked and the participant could see more balls than the avatar.

An answer is correct if it reflected the prescribed perspective (whether or not the given number corresponded to the number of balls seen by the participant or avatar as prescribed), and an error otherwise. The deliberative decision process is to determine the answer based on the indicated perspective (egocentric or altercentric). An alternative process for this task, however, is to base the answer on the own perspective only. In our terms, congruent trials are in alignment, and incongruent trials in conflict. Our predictions find full support in this context.

Dataset 12: Perspective Taking (Direction) A broad literature has explored the cognitive mechanisms of spatial perspective taking (i.e., Surtees et al., 2013, 2016), but applications to non-human, non-anthropomorphic objects are scarce. Muto et al. (2019) show that whether a non-anthropomorphic object is symmetric or not strongly influences spatial perspective-taking processes in humans. Participants were shown an object (chairs or humanoid figures) surrounded by four pillars, one of them, the target, painted blue (the other three being white). They were asked to indicate the location of a target from the point of view of the object, with some trials asking front vs. back and some asking left vs. right. Stimuli were created by rotating the whole room (reference object and pillars) by different angles, creating situations where the point of view of the reference object coincided with that of the participant (congruent), and situations where they differed, making perspective-taking necessary (incongruent). Muto et al. (2019)

argue that perspective-taking is strongly facilitated when the geometric asymmetry of the reference object provides a frame of reference (e.g., judge front-back for a chair with backrest). We consider their data when this was *not* the case, i.e. when the stimuli was symmetric in the direction required by the question (e.g., left-right but not front-back questions for a chair). This happened in half of the trials of their experiments 1, 4, and 5 (Experiments 2 and 4 contained asymmetric objects only) for a total of thirty-six participants.

An answer is correct if the participant reported the position of the target from the point of view of the reference object, and an error otherwise. The deliberative decision process is to judge the location of the target by taking the perspective of the reference object. An alternative process, however, is to judge the location from an egocentric perspective only. In our terms, congruent trials were in alignment, and incongruent trials in conflict. Our predictions find full support in this context, except for prediction T2 where differences are not significant.

4.4 Memory

Dataset 13: Spoken Word Recognition An important question in perception and linguistics is how the perceptual system maps two acoustically different signals to the same underlying word, and in particular how existing phonological variants of words are processed (Lahiri and Marslen-Wilson, 1991; Gaskell and Marslen-Wilson, 1996). The literature has shown a tight coupling between exposure frequency and the probability of recognizing variations in word pronunciations (e.g., Pitt et al., 2011; Sumner and Samuel, 2009). Charoy and Samuel (2020) study the effect of orthography (in particular omission of sounds when words are pronounced) in the recognition of spoken words.

Seventy-seven participants learned to associate new spoken words with pictures of unusual objects, while hearing the words in a reduced form as typical of conversational speech (meaning some sounds are omitted). The words were paired with either a spelling consistent with the reduced pronunciation, a more canonical spelling, or no spelling. Then they worked through a picture-name matching task, where they indicated whether a spoken word matched a displayed picture or not. In congruent trials, the pictures were presented together with words in the reduced spoken form which had been previously learned. In incongruent trials, words were pronounced in a canonical way (closer to the normative, orthography-based pronunciation) of those words. There were also filler trials with control words.

An answer is correct if the participant recognised the picture-name pair, and an error otherwise. The deliberative decision process in this case is to retrieve from memory the word-picture association while keeping in mind pronunciation rules (i.e., the existence of and differences between reduced vs. canonical forms). This should lead listeners to accept the canonical pronunciation derived from orthographic rules, even though that form of the word has not been previously heard. An alternative process is to retrieve from memory the phonological association only, based on the reduced pronunciation which was actually learned. This should lead listeners to reject the canonical pronunciation. In our terms, congruent trials are in alignment, and incongruent trials in conflict. Our predictions find full support in this context.

Dataset 14: Recollection Processes Widespread evidence suggests that recollection is supported by non-specific feelings of familiarity which are recovered more rapidly than the realistic details that support the recollection of particular items Atkinson and Juola (1973); Mandler (1980). Formalising this notion, Brainerd and Lee (2019) follow

an explicit dual-process approach to compare familiarity and recollection processes in a memory paradigm (a conjoint-recognition task), and show that the relative speeds of different retrieval processes depend on the relation between the stimuli and previously-learned words. Participants (one-hundred and eighty-five participants in six very similar experiments) first studied lists of interrelated words. In the actual task, subjects accepted or rejected new words according to one of three criteria. In *Verbatim trials* subjects should accept previously-learned words and reject all others, independently of whether they were conceptually related to the learned ones or not. In *Gist trials* subjects should accept words related to those learned and reject unrelated ones *and* actually-learned words. In *Verbatim-Gist* trials subjects should accept both previously-learned and related words, and reject new, unrelated ones.

An answer is correct if the participant accepted the word as specified in the active criterion, and an error otherwise. The deliberative decision process in this case is to recollect the studied words and follow the specified criterion to decide whether to accept a word or not. An alternative process is to accept all words related to the studied theme, in particular not distinguishing between actually-learned words and thematically-related ones. In our terms, Verbatim trials were in alignment whenever either previously-learned words (which should be accepted) or new, unrelated words (which should be rejected) were presented, and in conflict if the stimuli were new but related words. Gist trials were in alignment if the stimuli were either new but related words (which should be accepted) or new, unrelated words (which should be rejected), and in conflict if they were previously-learned words. Finally, all Verbatim-Gist trials were in alignment, as the description of the deliberative rule coincides with the associative, automatic process. Our predictions find full support in this context.

Dataset 15: Recognition Heuristic Glöckner and Bröder (2014) compare different decision-making models for judgements involving memory-based information recognition, and in particular target the recognition heuristic (Goldstein and Gigerenzer, 2002; Glöckner and Bröder, 2011). Participants ($N = 61$) decided which of two USA cities had more inhabitants. Two sets of eight mid-sized cities were used, half of which were mainly known (Miami Beach, Charleston, Oklahoma City, Buffalo, Salt Lake City, Richmond, Albany, Orlando), while the other half were mostly unknown (Hialeah, Carson City, Mobile, Tempe, Lansing, Trenton, Topeka, Stockton).

An answer is correct if the participant indicated the city with most inhabitants among the two, and an error otherwise. The deliberative decision process in this case is to retrieve information from memory to decide which city actually has more inhabitants. An alternative process is the recognition heuristic, that is, basing the answer on whether the city is recognized or not, hence equating higher familiarity with larger population. In our terms, trials where a larger, known city was pitted against a smaller, unknown one were in alignment while trials where a known but smaller city was compared to a larger but unknown city were in conflict. Trials where the cities were either both known or both unknown were neutral, and we will revisit them in a later section. Our predictions are again fully supported.

4.5 Decision Making

Dataset 16: Syllogisms and Base Rates Time pressure has been extensively used as a tool to manipulate the cognitive mode in decision making. In two experiments (260 subjects in total), Raoelison et al. (2020) go beyond this approach by studying decisions in reasoning tasks where a first choice given under time pressure (3s limit)

can be corrected afterwards, and find a positive correlation between cognitive capacity and correct answers under time pressure. We reanalyze the first, initial answer in those decisions. The experiments in Raoelison et al. (2020) involved three kinds of tasks: syllogisms, base-rate questions (see, e.g. Alós-Ferrer et al., 2016; Ludwig et al., 2020), and cognitive-reflection items (Frederick, 2005). The first two were presented with binary-answer formats (the third, with non-binary answers, is analyzed in the Section "Beyond Binary Choice" below).

For the syllogisms, participants were asked to evaluate whether or not (yes/no) a statement followed logically from a syllogistic-style reasoning ("All dogs have four legs. Puppies are dogs. Does it follow that all puppies have four legs?" vs. "All dogs have four legs. Puppies have four legs. Does it follow that all puppies are dogs?"). No-conflict items were such that the conclusion followed logically (valid) and was also believable, or the conclusion did not follow logically (invalid) and was unbelievable. Conflict items were such that the conclusion was valid but unbelievable, or invalid but believable. Each participant faced four conflict and four no-conflict items. Experiment 2 also included four neutral items per participant, where the syllogism was stated in abstract terms ("All F are H," etc.) and hence the conclusion elicited no believability evaluation.

The base-rate tasks followed extreme versions of the lawyers-engineers problem of Kahneman and Tversky (1972), where a stereotypical judgment might contradict the stated prior probability ("There are 995 clowns and 5 accountants. Person L is funny. Is Person L more likely to be a clown or an accountant?"). In no-conflict items, base rates and stereotypical information cued the same response, and in conflict items they cued different responses. As in the case of syllogisms, each participant faced four conflict and four no-conflict items. Experiment 2 also included four neutral items per participant, where the stereotypical association applied to both possible responses (e.g., being musical for saxophone and trumpet players).

For syllogisms, an answer is correct if it reflects whether the final statement actually follows logically from the initial ones, and an error otherwise. The deliberative decision process is to evaluate the logical validity of the statements, while the alternative process focuses only on believability of the final one. For base-rate questions, an answer is considered correct if it is in agreement with the extreme prior, and an error otherwise. The deliberative process is to focus on the prior, and the alternative one is to focus on the stereotype. No-conflict items were in alignment, and conflict-items reflect our notion of conflict. Neutral items, where the heuristic was not cued, correspond to our concept of neutral situations and will be examined in a later section. Our predictions hold both for syllogisms and base-rate questions, showing that even for more cognitively demanding (and time-consuming tasks) we find evidence for dual-process effects.

Dataset 17: Reinforcement Learning A small number of recent contributions have explored computational models describing both the processes underlying a single decision and how those are influenced by subjective option values learned over time (e.g., Frank et al., 2015; Pedersen et al., 2017). Among those, Fontanesi et al. (2019) study how reinforcement influences learning in value-based decisions, with participants becoming faster and more accurate for more dissimilar values and generally as the experiment progressed, and decided faster for more attractive (i.e., overall more valuable) pairs of options. During the experiment, each participant ($N = 27$) saw a total of twelve different figures (in three blocks of four each) representing different reward options. Participants chose between two of them in each trial. The payoffs of each option were not fixed but varied and were approximately normally distributed. The mean rewards of the options in

each block were 36, 40, 50, and 54 for options A, B, C, and D, respectively. The standard deviation was 5 for all options. After each choice, participants saw both options' rewards. At the end of the experiment, the accumulated reward was paid to the participants.

An answer is correct if the participant chose the option with the highest expected value, and an error otherwise. The deliberative decision process in this case is to choose the option with the highest expected value. Because monetarily-relevant feedback was provided to participants, a natural alternative process is reinforcement learning. In particular, in its simplest form (win-stay/lose-shift) it prescribes to choose the same option if it was successful (delivered a larger payoff than the alternative in that round) and switch to another if it delivered a bad result (worse than the alternative). In our terms, alignment situations are those where the choice in the last trial was successful and is the correct choice in the current round, or where the previous choice was unsuccessful and the same option is an error in the current trial. Conflict situations are those where the past choice was successful but it is an error in the current trial, or where the previous choice was unsuccessful but it is now the correct option. All those trials where past choices are not among the available alternatives are neutral trials (and we will return to them in a later section). Overall, 30.63% of the observations are classified as alignment, 15.07% as conflict, and 54.30% as neutral. Our predictions find support in this context, except for T1, where differences are not significant.

5 Beyond Binary Choice

In this section, we extend the basic model beyond the binary case to allow for any number of alternatives. First we show that all four predictions can be extended to the general case. Then, we examine a few additional, non-binary paradigms to illustrate the model's applicability.

5.1 Extended Model II: The Multi-Alternative Case

Consider a task with finitely many possible answers. Denote the set of alternatives by $X = \{x_1, \ldots, x_n\}$. As in the binary case, suppose that a more deliberative process D and a more automatic one I codetermine behavior, and let $\Delta > 0$ be the probability that the actual response is selected according to process I. Denote by $P(x|D)$ and $P(x|I)$ the probabilities that an alternative $x \in X$ is selected by process D and I, respectively, so that the actual probability of a response x being selected is

$$P(x) = \Delta P(x|I) + (1 - \Delta)P(x|D).$$

Let x^D and x^I again denote the favored (modal) answers of these processes, respectively. That is, $P^D = P(x^D|D) > P(x|D)$ for all $x \neq x^D$ and $P^I = P(x^I|I) > P(x|I)$ for all $x \neq x^I$. This simply makes explicit that the alternative favored by a process is indeed the process' most frequent selection. Note that, for the multi-alternative case this does not imply that the prescription is selected more than half of the time.

As in the binary case, the interpretation is that the modal (favored) option of the deliberative process, x^D, is normatively correct (or, alternatively, that the word "correct" is used to denote this option). Thus, a trial corresponds to *conflict* if $x^D \neq x^I$, and to *alignment* if $x^D = x^I$. A trial would be *neutral* if the more intuitive process is not cued.

Assumptions (R) and (P) are unchanged in the multi-alternative case. That is, if R^D and R^I are the expected process response times, assumed for simplicity (and as in the binary case) not to depend on the answer actually selected by a process, (R) states

that $R^D > R^I$, capturing that the more automatic process is faster in expected terms. Analogously, (P) states that $P^I > P^D$, that is, the more automatic process is more internally consistent than the more deliberative one in the sense that I selects the own favored response x^I more often than D selects x^D.

The generalized Stroop effect (D1; Theorem 1) also holds in the multi-alternative case. However, it is worth noticing that assumption (P) is now necessary for the proof (it was not required in the binary case).

Theorem 5. *In the multi-alternative case, if (R) and (P) hold,*

(D1) correct responses are slower in expectation in case of conflict than in case of alignment.

Theorem 2 on the proportion of correct responses in conflict and alignment also extends to the multi-alternative case (even if assumptions (R) and (P) do not hold). Further, in this case we obtain an additional prediction. Say that the answer in a trial is the *intuitive choice* if the participant selects the favored answer of the intuitive process, x^I. We then obtain the following generalization (all proofs are in the Appendix).

Theorem 6. *In the multi-alternative case,*

(D2) the proportion of correct responses is strictly smaller in case of conflict than in case of alignment, and

(D2') the proportion of intuitive choices is strictly smaller in case of conflict than in case of alignment (when they are also correct).

The additional prediction (D2') makes particular sense in the multi-alternative case, since many answers might be neither correct (x^D) nor intuitive (x^I). In the binary case, (D2') is an immediate consequence of (D2) if errors are less likely than correct responses, since in conflict intuitive responses are the only errors and hence their frequency is the complementary of the frequency of correct responses. We remark that in binary paradigms with large error rates (above 50%), (D2') would remain an additional prediction, which however would imply (D2).

Predictions (T1) and (T2), on the relative speed of errors and correct responses (Theorem 3) also extend to the multi-alternative case, with one *caveat*. With more than two alternatives, in case of conflict there are two types of errors. *Intuitive errors* are those where the participant selects the option favored by the more automatic process, x^I. Other errors are those where the participant selects some answer not favored by either process, $x \neq x^D, x^I$. In the multi-alternative case, (T1) holds for the comparison of *intuitive* errors and correct responses, and hence it is particularly important to record the actual answers and not only whether they are correct or erroneous. In contrast, in case of alignment, where errors are always answers favored by neither process, (T2) holds unchanged. As in Theorem 3, the proof of (T2) rests on assumption (P), but the proof of (T1) does not.

Theorem 7. *Consider the multi-alternative case and assume (R).*

(T1) In case of conflict, the expected response time of intuitive errors is shorter than the expected response time of correct answers.

(T2) Assume (P). In case of alignment, the expected response time of errors is larger than the expected response time of correct answers.

Last, consider the model extension including a non-decision time and a process selection probability depending on conflict or alignment, t_A vs. t_C and Δ_A vs. Δ_C. As was the case for the binary model (Theorem 4), all predictions above hold for this extended model.

Theorem 8. *Consider the extended model for the multi-alternative case and assume (R), (P), $t_C \geq t_A$, and $\Delta_C \leq \Delta_A$. Then (D1), (D2), (D2'), (T1), and (T2) hold.*

Further, if (R) and (P) did not hold because the two processes are indistinguishable in terms of automaticity, and in particular $R^D = R^I$, predictions (D1), (D2), and (D2') still hold as long as the inequality $t_C \geq t_A$ is strict. This extends Corollary 1 to the multi-alternative case.

Corollary 2. *Consider the extended model for the multi-alternative case and assume $R^D = R^I$, $t_C > t_A$, and $\Delta_C \leq \Delta_A$. Then (D1), (D2), and (D2') hold.*

5.2 Results (Non-Binary Choice)

Datasets 19 to 21 involve four different experimental tasks reporting choice and response time which can be described in terms of dual, interacting processes and for which conflict and alignment could be identified in the dataset. Those are listed in the four bottom rows of Table 1. The Section "Description of Studies: Non-Binary Choice" below briefly describes each study and how it is encompassed by our model. Three of the datasets (18–20) belong to the area of cognitive control, and the last (21) corresponds to a decision-making task (the Cognitive Reflection Test).

As in the binary case, we test our predictions for each individual study or treatment by means of non-parametric Wilcoxon signed-rank (WSR) tests, which do not require distributional assumptions. The results are summarized in the last four columns of Table 1.

All predictions (D1, D2, D2', T1, and T2) hold in all four datasets with Non-Binary Choice, with the exception of (T2) in Dataset 21. The Appendix contains details of the individual tests, robustness analyses, and further comparisons as appropriate for each dataset.

As in the case of binary-choice datasets, the left-hand side of Figures 1–4 give a forest plot representation to display the actual difference in response times or proportions of correct answers as given in predictions (D1), (D2), (T1), and (T2), including the traditional 95% confidence intervals (assuming a normal distribution of the difference variables for illustration), and a vertical line at zero for ease of interpretation. The right-hand side of the figures displays the effect sizes and corresponding 95% confidence intervals for the WSR tests in Table 1, following the same procedure discussed above to obtain confidence intervals.

The effect size of a WSR test, r, is considered small for $r \in [0.1, 0.3]$, medium for $r \in [0.3, 0.5]$, and large for $r > 0.5$ (Cohen, 1988; Rosenthal, 1994). We remind the reader that, if there is no ambiguity in the test, in particular if all signs in a non-parametric test are positive (or negative), then one obtains point estimates for the effect size instead of a proper interval. In the Non-Binary case, this occurs for 3 tests (all of them concerning (D1)).

The figures further illustrate that the datasets also lend overwhelming support to our predictions in the Non-Binary case.

5.3 Description of the Studies: Non-Binary Choice

Dataset 18: Task Switching Across the Lifespan Steyvers et al. (2019) collected a large dataset ($N = 1,000$) from an online cognitive-training platform and showed that practice improves task-switching performance, but persistent costs remain even after extensive practice, and more so in older adults. This work contributes to the literature that tries to identify the characteristics of the cognitive processes underlying task switching (i.e. Rogers and Monsell, 1995; Altmann and Gray, 2008). Stimuli were moving images of leaves which differed along two dimensions: the direction in which the leaf pointed and the direction of movement. Each visual dimension had four feature values (up, down, left, and right), and all 16 combinations were possible. A task cue (the color of the leaf, green or orange) instructed subjects to report either pointing or movement direction. Subjects logged in for multiple sessions. Within each session, the experimental paradigm intermittently alternated between the two task cues, with each task run including a variable number of trials. An answer is correct if the participant indicated the direction that was relevant according to the actual task in the specific trial, and an error otherwise. In this context, an intuitive error is an answer which would have been correct according to the task that was active in the previous trial.

The deliberative decision process is to react to the color cue and focus attention on the pertinent dimension, hence indicating either pointing or movement direction as appropriate. An alternative process, given the structure of the game, is to simply report the same dimension as in the previous trial. Within a given run, all trials except the first were in alignment. The first trial of each run in a session, except for the one of the first run in that session, was in conflict, as new runs changed the task. The first trial in each session was neutral, since there was no previous trial that could cue the alternative process. Our predictions find full support in this context.

Dataset 19: Sequential Conflict Modulation Dignath et al. (2019) showed that episodic memory stores a snapshot of internal attentional states (e.g., focused attention) together with contextual information, hence memory aids actions by automatizing and tailoring them to the situational circumstances. Following Goschke (2000), this work finds that, rather than being orthogonal dimensions, cognitive control and memory retrieval are closely related (see also Scherbaum et al., 2010).

In two experiments ($N = 39$ and $N = 48$, respectively), participants were briefly (139 ms) shown either numbers (integers between 3 and 6 in letters or Arabic digits) in Experiment 1 or colors (4 different ones, displayed either as the corresponding words or as color patches) in Experiment 2, and reported them by pressing the d, f, g, or h key on a keyboard with their right index, middle, ring, or little finger. Crucially, before the target stimuli, a task-irrelevant distractor was also briefly presented for 139 ms. The latter was either a word or a digit, and it was either the same or different from the target that followed (which always appeared in the same format as the distractor). Half of the trials were congruent, meaning that the target and the distractor coincided, while the rest were incongruent, meaning that target and distractor differed. We say that an answer is correct if the participant pressed the key corresponding to the target stimuli, and an error otherwise. Intuitive errors are errors where the distractor stimuli is reported.

The deliberative process is to ignore the distractor and focus on reporting the actual target, which is effortful as the distractor and the stimuli are presented in very fast succession. An alternative decision process is to focus on the distractor, since it appears first. This is a case in which the cognitive differences between the processes might be

arguable, and hence the model becomes a test of the joint hypothesis represented by process multiplicity and which process is to be considered more deliberative. In our terms, congruent trials were in alignment, and incongruent ones were in conflict. Our predictions find full support in this context.

Dataset 20: Prime-Probe Congruency Effects Schmidt and Weissman (2014) conducted two different experiments (16 subjects each) using the prime-probe arrow task of Kunde (2003) to study sequential congruency effects and trial-by-trial attention adjustments without introducing learning effects. The designs used in this work were developed to overcome limitations previously pointed out in the literature (i.e. Mayr et al., 2003; Jiménez and Méndez, 2013). In their experiments, participants were required to indicate a direction (up, down, left, or right) as shown by an arrow (Experiment 1) or the corresponding word (Experiment 2). Before the target stimuli was presented, however, a vertical or horizontal distractor array of five identical arrows (Experiment 1) or three identical words (Experiment 2) was presented for 133 ms, hence priming the participants to give a certain response. The experiments used four congruent (left-left, right-right, up-up, down-down) and four incongruent (left-right, right-left, up-down, down-up) distractor-target pairings. An answer is correct if the participant indicated the direction given by the target stimuli, and an error otherwise.

The deliberative decision process is to indicate the direction actually given by the target stimuli, while the alternative process is to report the direction primed by the distractor array (the most common direction). Obviously, congruent trials, where the target and distractor point in the same direction, were in alignment, while incongruent trials were in conflict. Answers in conflict were intuitive errors if they followed the direction of the distractor (most errors were intuitive in this case). All our predictions hold in this setting.

Dataset 21: Cognitive Reflection The experiments of Raoelison et al. (2020), discussed as Paradigm 16a and 16b, included a third set of questions following the celebrated bat-and-ball item in the Cognitive Reflection Test of Frederick (2005) ("A bat and an ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?"). In these questions, a heuristic cues an intuitive but wrong response (10 cents), while the correct response is not intuitive (5 cents). By changing the involved categories (e.g., pencils and erasers) and numbers, each participant faced four such questions. There were four possible answers: the correct one, the intuitive one, and two fillers. Additionally, each participant faced four no-conflict items (with four possible answers each) where, by eliminating the "more than" comparison, the heuristic cued the correct response. Study 2 also included four neutral items in the form of simple, verbally-stated and framed arithmetic problems, where the heuristic was not cued. As for the other items, participants gave a first answer under time pressure (in this case within 5 s), which they could correct later. We reanalyze that initial answer. An answer is correct if the participant indicated the normative solution to the problem, and an error otherwise. In conflict, intuitive errors correspond to the responses cued by the heuristic.

The deliberative decision process is to solve the analytical problem, while the alternative process is to follow the heuristic. Trials where the heuristic cued the wrong answer were in conflict, and those where it suggested the correct choice were in alignment. Trials where the heuristic was not cued were neutral situations (and we will return to them in a latter section). Again, our predictions hold except for T2. In alignment, the sample

size is drastically reduced since the vast majority of participants made no mistakes. In this case, errors are slower than correct answers, but the difference is not significant.

6 Neither Conflict Nor Alignment: Neutral Trials

Obviously, not all human decisions trigger several, possibly competing processes. Even in paradigms designed to elicit multiple processes, it is also often the case that some trials might lack the cue that elicits the more-automatic process targeted in the particular paradigm. In our terms, such trials are called *neutral*, and offer an opportunity to study the deliberative processes in isolation, and to compare them to situations of either conflict or alignment.

6.1 Neutral Trials in the Considered Datasets

Neutral trials are present in many (but not all) of the datasets we have considered. In the Flanker task of White and Curl (2018) (Dataset 5), in roughly one third of the trials the Flankers were entirely absent. In the Gabor-patch classification task of Denison et al. (2018) (Dataset 6), one sixth of the trials were preceded by strictly neutral cues, creating neither conflict nor alignment. In the task of Heathcote et al. (2019) (Dataset 8), where participants had to judge the majority color in a large, rapidly-flickering checkerboard, 37.5% of the blocks were *unbiased*, meaning that they included neither a prior nor a reward asymmetry, and hence cued no alternative process. In the three categorization experiments reanalyzed by Hu and Rahnev (2019) (Dataset 9), part of the trials included no previous cue and are hence neutral in our terms. In the recognition heuristic experiment of Glöckner and Bröder (2014) (Dataset 15), where participants had to name the most-populated city of two named ones, 46.7% of trials pitted either two well-known or two mostly unknown cities against each other, and hence the recognition heuristic was not cued. In the syllogism tasks of Raoelison et al. (2020) (Dataset 16), each participant faced four neutrally-framed items formulated abstractly, hence not eliciting a believability-based process. In the base-rate tasks from the same article (see again Paradigm 16 above), again four items per each participant were neutral, in this case because the stereotypical association applied to both possible responses. In the reinforcement-learning study of Fontanesi et al. (2019) (Dataset 17), trials where the previous choice was not among the actual alternatives were neutral, because reinforcement could not be triggered. This happened 54.3% of the time. In the dataset of Steyvers et al. (2019) (Dataset 18), where participants had to report either the direction a leaf was pointing to or its direction of movement, the only neutral trials correspond to the very first-trial in each session, since there was no previous trial which could interact with the currently-active criterion. While this is a very small proportion of trials, if still adds up to a large number (47,410 in total) due to the size of the dataset. Last, the Cognitive Reflection Test in Raoelison et al. (2020) (Dataset 21) included neutral items in the form of simple, verbally-stated arithmetic problems which elicited no intuitive response at all.

6.2 Error Rates in Neutral Trials

The first opportunity afforded by the presence of neutral situations in the considered datasets is to compare error rates across possible situations. Prediction (D2) in Theorems 2 and 6 states that the proportion of correct responses should be larger in case of alignment compared to conflict. This prediction also holds for the extended models

(Theorems 4 and 8). The following (immediate) result shows that, in the binary case, error rates for neutral trials should be intermediate between the conflict and alignment cases.

Proposition 1. *Consider the binary-choice case. Assume (P).*

(N1) The proportion of correct responses in neutral trials is strictly smaller than the proportion of correct responses in case of alignment.

(N2) The proportion of correct responses in neutral trials is strictly larger than the proportion of correct responses in case of conflict.

The intuition for these predictions is straightforward. The proportion of correct responses in neutral trials should simply be the expected frequency of the modal answer for the deliberative process, P^D. The proportion of correct responses in case of alignment is a convex combination between P^D and P^I, since in alignment the modal answer of the intuitive process is also correct. Since $P^I > P^D$ by (P), (N1) follows. In case of conflict, the modal answer of the intuitive process is x^I, which is incorrect. The proportion of correct responses in case of conflict is a convex combination between P^D and $1-P^I < 1/2 < P^D$, and (N2) follows (note that this last prediction does not require assumption (P)).

In the multi-alternative case, prediction (N1) follows without change by the same logic, but the second prediction is more subtle.

Proposition 2. *Consider the multi-alternative case. Assume (P).*

(N1) The proportion of correct responses in neutral trials is strictly smaller than the proportion of correct responses in case of alignment.

(N2') The proportion of intuitive choices in neutral trials is strictly smaller than the proportion of intuitive choices in case of conflict.

The intuition for (N2') is also simple. The proportion of intuitive choices in neutral trials is simply the frequency of the intuitive answer under the deliberative process, i.e. $P(x^I|D)$. In case of conflict, the proportion of intuitive choices is a convex combination between P^I and $P(x^I|D)$. By (P), $P^I > P^D$ and the latter is larger that $P(x^I|D)$ since x^D is process D's modal answer. Hence (N2') follows.

These predictions are supported in our datasets. First, in agreement with (N1), our analyses show that it is more likely to observe more errors in neutral than alignment situations. This is summarised in Figure B.1 in Appendix C, which also reports the tests. Specifically, we observe significantly more errors (hence less correct responses) in neutral than alignment situations in twelve out of fifteen studies containing neutral trials, with only one other study displaying the opposite result (Dataset 18). The remaining two studies exhibit no significant differences.

Evidence also suggests that, in agreement with prediction (N2), it is far more likely to observe more (intuitive) errors in conflict than in neutral situations. These results are summarised by Figure B.2 in Appendix C, which also reports the tests. Specifically, we observe significantly more errors in conflict than neutral situations in eleven out of the fifteen studies that allow for this comparison. The opposite effect is never observed; for the remaining four studies, differences are not significant. In the two cases with more than two alternatives including neutral trials, prediction (N2') is supported.

6.3 Response Times in Neutral Trials

Prediction (D1) shows that correct responses should be slower under conflict than in case of alignment. The second comparison of interest involving neutral trials concerns whether correct responses are faster or slower in those trials than in conflict or alignment trials. Assuming no difference between the response times of errors and correct responses within a single process, the expected response time of correct answers in neutral trials is simply R^D, while the expected response time of correct answers in conflict or alignment trials is a convex combination between R^D and R^I, as correct answers might come from either process. Since $R^D > R^I$ by (R), the response time of correct answers should be longer for neutral trials. However, if one takes into account non-decision time as in Extended Model I, it is natural to assume that t_C would be larger than the non-decision time of neutral trials, as the former should capture conflict resolution and process selection, while in neutral trials only one process is active. The two effects go in opposite directions, and hence there is no natural prediction. It is less natural to assume any specific ordering between t_A and non-decision time for neutral trials, but since this does not exclude that the former might be larger, again no natural prediction arises.

Suppose that the time needed for conflict detection and resolution, t_C, is enough to offset the differences in response times among the processes. Suppose, however, that the analogous, shorter time when there is actually no conflict, t_A, does not suffice. In this particular case, we would obtain the prediction that correct responses for neutral trials are faster than those in conflict, but slower than those in alignment.

This is broadly supported by the analysis of our datasets. For the kind of deliberative and intuitive processes involved in the tasks collected here, it is far more likely to observe correct responses in neutral situations being faster than those in conflict but slower than those in alignment.

Specifically, in the fifteen studies that allow for this comparison, we observe slower correct responses in conflict trials compared to neutral ones in nine cases, no significant differences in four cases, and the opposite effect in only one occasion (Dataset 17). We also observe faster correct responses in alignment trials compared to neutral ones in seven cases, no significant differences in seven other cases, and the opposite effect in only one occasion (Dataset 18). These results are summarised in Figures B.3 (neutral vs. conflict) and B.4 (neutral vs. alignment) in Appendix C, which also contains the actual tests.

6.4 Slow Errors in Neutral Trials

The last comparison of interest afforded by the presence of neutral situations in our datasets concerns the relative speed of favored (modal) answers and other answers in the absence of dual-process effects. The original, symmetric-boundaries drift-diffusion model of Ratcliff (1978) predicts identical response times distributions for either response, unless trial-by-trial variability in either drift rates or starting points is assumed. For simplicity, the model variants discussed so far assume no differences in the response times conditional on given answers, *for a fixed process*, and show that asymmetries (the relative speed of errors) can arise simply due to the interaction of several processes (however, the subsection Extended Model III below relaxes this assumption). Still, the considered datasets allow us to empirically evaluate whether there is actual evidence pointing to an asymmetry in response times already within a single process.

Overall, evidence suggests that for the kind of deliberative processes involved in the tasks collected here, and in the absence of a different, more automatic process, either

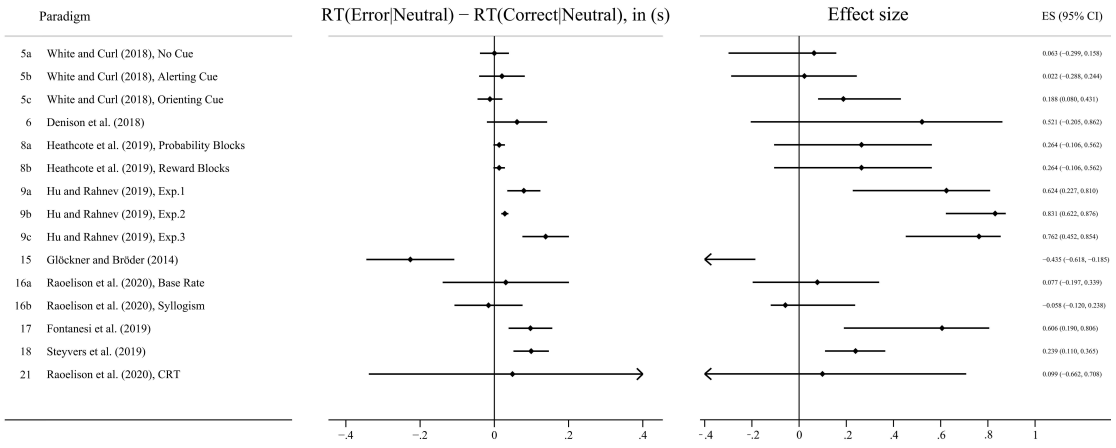

Figure 5: Comparison of the response times of correct choices and errors in neutral trials, for all datasets containing this type of trials. Effect sizes are for the non-parametric tests. Bootstrapped confidence intervals using 5,000 repetitions.

the response times of modal and non-modal responses are not significantly different, or, if a difference exists, it goes in the direction of modal answers being faster. Note that, since neutral trials involve deliberative processes only, and a correct answer is defined as the modal answer of the deliberative process, this is the same as observing slower errors in neutral trials.

These results are summarised by Figure 5. We observe no significant differences in four of the datasets enumerated above (5, 8, 16, and 21), encompassing seven different comparisons. We observe significantly slower errors (compared to correct responses) in the neutral trials of four other datasets (6, 9, 17, and 18), encompassing six different comparisons. Faster errors in neutral trials were only observed in one dataset (15), and only for a specific subset of comparisons (when both cities were categorized as unknown). All tests are in Appendix C.

7 Extended Model III: Option-Dependent Process Response Times

In this section, we relax assumption (R) to reflect the possibility that expected process response times differs depending on the actually-selected response. In the previous section, we examined neutral trials, where only one (deliberative) process should be active, and found that, in general, either there was no difference in response times conditional on the selected alternative, or correct answers were observed to be slower. In the presence of a single, deliberative process D, this would correspond to the property that the response time of the favored (modal) answer x^D is shorter than the response time of other possible answers.

Consider the more general, multi-alternative case as in Extended Model II. That is, the set of alternatives is $X = \{x_1, \ldots, x_n\}$. All derivations and results apply to the binary case simply by considering two alternatives, $n = 2$. Further, we allow for a non-decision time and a process selection probability depending on conflict or alignment as in Extended Model I, t_A vs. t_C and Δ_A vs. Δ_C. Again all derivations and results apply to the basic case with $t_A = t_C$ and $\Delta_A = \Delta_C = \Delta$

Let $R^D(x) = E[t|x, D]$ and $R^I(x) = E[t|x, I]$ be the expected response times conditional on the response being selected by process D or I, respectively, and on the alternative x being selected. Assumption (R) is now weakened to the following property.

(R') $R^D(x^D) > R^I(x^I)$.

That is, the more automatic process is a swifter one only in the sense that it selects its own favored option faster on average than the more deliberative process selects its respective favored option.

We then require an additional assumption relating the response times of favored and non-favored alternatives for a given process. In view of the evidence in the previous section, we postulate that favored options are chosen faster than non-favored ones.

(R-D) $R^D(x^D) \leq R^D(x)$ for all $x \neq x^D$.
(R-I) $R^I(x^I) \leq R^I(x)$ for all $x \neq x^I$.

These assumptions are weaker that those in the Basic Model in the sense that, if expected response times do not depend on the selected alternative, (R-D) and (R-I) become vacuous, and (R) implies (R'). For process D, assumption (R-D) corresponds to "slow errors" within a single process. This is, however, different from a statement of whether observed errors are fast or slow, as the latter arise from the interaction of both processes and, except for neutral trials, it is not possible to observe which process actually selects the response. x In this more general case, all previous predictions still hold, with the exception of (T1).

Theorem 9. *Consider the extended model for the multi-alternative case and alternative-dependent response times. Assume (R'), (R-D), (R-I), (P), $t_C \geq t_A$, and $\Delta_C \leq \Delta_A$. Then (D1), (D2), (D2'), and (T2) hold.*

Hence, with the exception of (T1), our predictions do not depend on the assumption the expected response times do not depend on the selected alternative. Prediction (T1), i.e. fast errors in case of conflict, fails to obtain under the weaker assumptions above. This is because, without quantitative assumptions expected response time differences, assumption (R-D) runs exactly against property (T1). Empirically, we would of course still expect (T1) to hold as long as within-process response time differences as a function of selected alternatives are not too large. Formally, it is possible to prove a stronger result. (T1) will also hold in general if the response times of the deliberative process are option-dependent, as long as the selection of intuitive options (x^I) by the deliberative process is not much slower than the selection of the own favored responses (x^D). This is reflected by the following result.

Theorem 10. *Consider the extended model for the multi-alternative case and alternative-dependent response times. Assume (R'), (R-I), (P), $t_C \geq t_A$, and $\Delta_C \leq \Delta_A$. Then,*

(a) If the process response times of process D are option-independent (as in the Basic Model), (T1) holds.

(b) If the process response times of process D are not option-independent, but $R^D(x^I)$ is not much larger that $R^D(x^D)$, (T1) also holds. That is, given all parameters of the processes D and I except for $R^D(x^I)$, there exists $\bar{R} > R^D(x^D)$ such that, if $R^D(x^I) < \bar{R}$, then (T1) holds.

However, the fact that (T1) depends on assumptions not needed for any other prediction signals that this particular prediction might be less stable (across tasks and cognitive processes) than others. This broadly reflects our empirical results.

It is also possible to show a result analogous to Corollaries 1 and 2 in this case. Suppose that two processes are indistinguishable in terms of automaticity in the weaker sense that the expected response times of their respective favored responses do not differ. Then, predictions (D1), (D2), and (D2') still hold, as long as the inequality $t_C \geq t_A$ is strict.

Corollary 3. *Consider the extended model for the multi-alternative case and alternative-dependent response times. Assume (R-I),* $R^D(x^D) = R^I(x^I)$ *,* $t_C > t_A$*, and* $\Delta_C \leq \Delta_A$*. Then (D1), (D2), and (D2') hold.*

8 General Discussion

We presented and tested a simple formal nonparametric model which predicts that errors will be slower or faster than correct responses depending on whether two underlying processes are in alignment or in conflict, respectively. This corresponds to congruent and incongruent trials in many cognitive-control conflict tasks (as Stroop, Flanker, Simon, etc.), but extends to many other paradigms in the domains of attention, social cognition, memory, and decision making. Crucially, the model delivers predictions which are *ex ante* valid, i.e., do not depend on any specific values of model parameters.

The model also predicts a generalized Stroop effect, i.e., correct responses must be slower in case of conflict compared to alignment. It also predicts larger error rates in case of conflict. All predictions are shown to enjoy overwhelming support in 31 experimental tasks from 20 datasets from the recent literature.

A number of extensions show that the predictions are robust to considering more than two alternatives, non-decision times, process-selection probabilities, and differentiating conflict and alignment. The predictions hold even though the individual processes might exhibit no alternative-dependent response-time differences. That is, the directional predictions in response times arise from the dual-process structure which models the interaction of two different processes, and not from specific response-time differences associated with the alternatives themselves.

The limitations of the model arise from its very nonparametric nature. One could argue that the proposed framework is not a (computational) model in the sense usually associated with the word in cognitive psychology, because it operates at a different level of abstraction. The model cannot be fit to specific datasets, since it lacks any parameters to fit. Thus, it cannot be directly compared to other, parametric models. The model, however, is fully falsifiable (e.g., Jones and Dzhafarov, 2014), since the predictions do not depend on any specific parameter values or distributional assumptions.

8.1 Relationship to Single-Process Models

The model remains agnostic on the nature of the individual processes (deliberative and intuitive). In particular, one can further specify the framework by assuming specific models for those individual processes, without losing the general predictions. For example, one can assume the deliberative and intuitive processes to be independent Drift-Diffusion Models with different drift rates, as long as the intuitive one has a larger drift rate (in absolute value) and is hence swifter and more internally consistent than the deliberative one, hence fulfilling the general assumptions. In particular, this can be done assuming simple, symmetric DDMs without inter-trial variability in either drift rates or starting points. It is well-know that such simple DDMs predict identical response times for errors and correct responses, which has motivated more complex versions of the DDM. Our

model can generate fast and slow errors building upon two DDMs without any inter-trial variability, because the predictions arise from the interaction between the processes.

This is related to the same intuition that slow errors can already arise in a DDM with two different, randomly-selected drift rates (Ratcliff and Rouder, 1998). The difference is that, in our model, the two processes react to different dimensions of the stimuli, which allows to capture conflict and alignment. A detailed DDM-microfoundation of the model would be equivalent to a DDM specification where the drift rate is randomly sampled from two positive values for conflict trials, $\mu_1 > \mu_2 > 0$, and from two values with different signs for alignment trials, μ_2 and $-\mu_1$. Such a model was considered in Alós-Ferrer (2018).

Alternatively, one can assume any other process-model for the deliberative and intuitive processes as long as the general assumptions of any of the model extensions are fulfilled. For example, Poisson counter models (Townsend and Ashby, 1983; Smith and Van Zandt, 2000; Townsend and Liu, 2020) are known to predict slow errors, and have been criticized because fast errors are also observed empirically. Suppose that the deliberative and intuitive processes in our model are captured by two different Poisson counter models, one being swifter and more internally consistent than the other. Each process will entail longer response times for its own non-modal responses, i.e., slow within-process "errors." As we have seen, this is frequently observed in neutral trials, where only one process should be involved. The results in our Extended Model III then show that all our predictions continue to hold. That is, our framework can be immediately used to extend Poisson counter models (or any other class of models predicting slow errors) while capturing empirically-received response time asymmetries.

8.2 Other RT Asymmetries

Our model's most important prediction concerns the relative speed of errors depending on (exogenously observable) conflict or alignment between the underlying processes. The literature has reported other response time asymmetries. The first and most well-known one is the speed-accuracy tradeoff, where errors are observed to be faster if speed is emphasized, and slower if accuracy is emphasized.

Our model can explain this asymmetry as follows. Suppose individual processes exhibit the slow error property in the sense that the respective process modal response is on average faster than other responses (as in the case of Poisson counter models). Emphasizing accuracy is a manipulation which should shift the balance toward deliberative processes. In terms of the model, Δ should be closer to zero if accuracy is emphasized. The model hence approaches the single-process case where all decisions come from the same (deliberative) process. Hence, errors will tend to be slower. If speed is emphasized, however, the balance should be shifted toward the faster, more intuitive process. As long as the task still involves conflict, the value of Δ will be intermediate compared to the accuracy condition, resulting in the model's prediction of fast errors for conflict trials.

Other empirically reported asymmetries could also be explained by the model. It is sometimes reported that faster subjects tend to produce errors faster than correct responses, while slower subjects exhibit the opposite pattern (Ratcliff et al., 2004, p. 165). At the same time, slower subjects seem to be more deliberative in the sense that they make fewer errors. Suppose again that individual processes exhibit faster modal responses compared to other responses. Slower subjects might be slow because they rely on their deliberative processes more often (which explains the lower error rates). Again, this corresponds to a value of Δ closer to zero, and the model for those subjects approaches the single-process case, where errors will tend to be slower. In contrast, faster

subjects might be fast because they more often (but not exclusively) rely on intuitive processes, resulting in intermediate values of Δ, leading to the model's prediction of fast errors for conflict trials.

8.3 Comparison to Other Dual-Process Models

While there are many dual-process models in the literature, few are explicitly formalized. There are, however, a few notable exceptions which allow for an explicit comparison. Ulrich et al. (2015) considers a Diffusion Model for Conflict Tasks (DMC) which assumes two processes proceeding in parallel. Decisions follow from evidence accumulation driven mainly by a controlled (deliberative) process. However, an automatic process works on task-irrelevant stimuli and spills over by influencing the controlled process' drift rate. The sign of the spillover depends on whether a trial is congruent or not, and hence the correct response is facilitated (inhibited) by the automatic process in congruent (incongruent) trials, corresponding to our concept of alignment (conflict).

In contrast with our model, the DMC operates by aggregating both processes in a single accumulator, while the processes in our model operate independently. The DMC is analytically intractable in the sense that closed-form predictions are not feasible. Ulrich et al. (2015) focus on explaining negative-going delta functions, which plot the quantile difference for the RT distributions in incongruent and congruent conditions as a function of the quantiles' average, while we aim to predict the relative speed of errors compared to correct answers.

Diederich and Trueblood (2018) consider a formal, serial dual-process model for decisions under risk, where evidence accumulation is mainly driven by an intuitive process at the beginning of each trial, and by a more deliberative process toward the end of the trial. The switching point is randomly determined. That is, as in Ulrich et al. (2015), both processes contribute to a single accumulator, while in our model the processes operate independently, with random process selection. The model of Diederich and Trueblood (2018) is closer to dual-stage models of evidence accumulation, as, e.g., the one of Hübner et al. (2010). Diederich and Trueblood (2018) concentrate on biases in risky decision making and do not examine the relative speed of errors.

8.4 Conclusions

We presented a dual-process model which predicts slow or fast errors depending on an exogenously-observable classification of trials in a large variety of tasks. The model is nonparametric in the sense that the predictions do not depend on any specific values of parameters, and hence can fully falsify the model.

The predictions, which also include effects on error rates and a generalized Stroop effect, should hold whenever a task elicits two clearly-differentiated cognitive processes, one more deliberative than the other. This includes all classical conflict tasks (Stroop, Flanker, Simon, etc.), but also many others in the domains of cognitive control, attention, social cognition, memory, and decision making.

We tested the model in 31 experimental tasks from 20 different datasets and found overwhelming support for its predictions. These results are encouraging and suggest that the model captures a general structural relation between the interaction of cognitive processes and observable features of human behavior, very especially the relative speed of errors.

References

Achtziger, A. and Alós-Ferrer, C. (2014). Fast or Rational? A Response-Times Study of Bayesian Updating. *Management Science*, 60(4):923–938.

Achtziger, A., Alós-Ferrer, C., Hügelschäfer, S., and Steinhauser, M. (2014). The Neural Basis of Belief Updating and Rational Decision Making. *Social Cognitive and Affective Neuroscience*, 9(1):55–62.

Adler, W. T. and Ma, W. J. (2018). Comparing Bayesian and non-Bayesian Accounts of Human Confidence Reports. *PLoS Computational Biology*, 14(11):e1006572.

Algina, J., Keselman, H., and Penfield, R. D. (2006). Confidence Interval Coverage for Cohen's Effect Size Statistic. *Educational and Psychological Measurement*, 66(6):945–960.

Alós-Ferrer, C. (2018). A Dual-Process Diffusion Model. *Journal of Behavioral Decision Making*, 31(2):203–218.

Alós-Ferrer, C., Garagnani, M., and Hügelschäfer, S. (2016). Cognitive Reflection, Decision Biases, and Response Times. *Frontiers in Psychology*, 7 (1402):1–21.

Altmann, E. M. and Gray, W. D. (2008). An Integrated Model of Cognitive Control in Task Switching. *Psychological Review*, 115(3):602–639.

Atkinson, R. C. and Juola, J. F. (1973). Factors Influencing Speed and Accuracy of Word Recognition. *Attention and Performance IV*, pages 583–612.

Baddeley, A. D., Chincotta, D., and Adlam, A. (2001). Working Memory and the Control of Action: Evidence From Task Switching. *Journal of Experimental Psychology: General*, 130(4):641–657.

Bang, J. W. and Rahnev, D. (2017). Stimulus Expectation Alters Decision Criterion but Not Sensory Signal in Perceptual Decision Making. *Scientific Reports*, 7(1):1–12.

Bargh, J. A. (1989). Conditional Automaticity: Varieties of Automatic Influences in Social Perception and Cognition. In Uleman, J. S. and Bargh, J. A., editors, *Unintended Thought*, pages 3–51. Guilford.

Blurton, S. P., Kyllingsbæk, S., Nielsen, C. S., and Bundesen, C. (2020). A Poisson Random Walk Model of Response Times. *Psychological Review*, 127(3):362–411.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict Monitoring and Cognitive Control. *Psychological Review*, 108(3):624–652.

Brainerd, Charles J. and, K. and Lee, W.-F. (2019). Recollection Is Fast and Slow. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(2):302.

Brown, S. D. and Heathcote, A. (2005). A Ballistic Model of Choice Response Time. *Psychological Review*, 112(1):117–128.

Brown, S. D. and Heathcote, A. (2008). The Simplest Complete Model of Choice Response Time: Linear Ballistic Accumulation. *Cognitive Psychology*, 57:153–178.

Casey, B. J., Galvan, A., and Hare, T. A. (2005). Changes in Cerebral Functional Organization During Cognitive Development. *Current Opinion in Neurobiology*, 15(2):239–244.

Charoy, J. and Samuel, A. G. (2020). The Effect of Orthography on the Recognition of Pronunciation Variants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(6):1121–1145.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., and Bird, G. (2017). Submentalizing or Mentalizing in a Level 1 Perspective-Taking Task: A Cloak and Goggles Test. *Journal of Experimental Psychology: Human Perception and Performance*, 43(3):454–465.

Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., Radkova, I., Deschrijver, E., and Brass, M. (2018). Automatic Imitation: A Meta-Analysis. *Psychological Bulletin*, 144(5):453–500.

Davis-Stober, C. P., Morey, R. D., Gretton, M., and Heathcote, A. (2016). Bayes Factors for State-Trace Analysis. *Journal of Mathematical Psychology*, 72:116–129.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge, UK.

de Lange, F. P., Rahnev, D. A., Donner, T. H., and Lau, H. (2013). Prestimulus Oscillatory Activity Over Motor Cortex Reflects Perceptual Expectations. *Journal of Neuroscience*, 33(4):1400–1410.

De Neys, W., Vartanian, O., and Goel, V. (2008). Smarter than We Think: When Our Brains Detect That We Are Biased. *Psychological Science*, 19(5):483–489.

Denison, R. N., Adler, W. T., Carrasco, M., and Ma, W. J. (2018). Humans Incorporate Attention-Dependent Uncertainty into Perceptual Decisions and Confidence. *Proceedings of the National Academy of Sciences*, 115(43):11090–11095.

Diederich, A. and Trueblood, J. S. (2018). A Dynamic Dual Process Model of Risky Decision Making. *Psychological Review*, 125(2):270.

Dignath, D., Johannsen, L., Hommel, B., and Kiesel, A. (2019). Reconciling Cognitive-Control and Episodic-Retrieval Accounts of Sequential Conflict Modulation: Binding of Control-States into Event-Files. *Journal of Experimental Psychology: Human Perception and Performance*, 45(9):1265–1270.

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, 82(397):171–185.

Egner, T. (2008). Multiple Conflict-Driven Control Mechanisms in the Human Brain. *Trends in Cognitive Sciences*, 12(10):374–380.

Evans, J. S. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, 59:255–278.

Evans, N. J., Hawkins, G. E., Boehm, U., Wagenmakers, E.-J., and Brown, S. D. (2017). The Computations that Support Simple Decision-Making: A Comparison Between the Diffusion and Urgency-Gating Models. *Scientific Reports*, 7(1):1–13.

Fontanesi, L., Gluth, S., Spektor, M. S., and Rieskamp, J. (2019). A Reinforcement Learning Diffusion Decision Model for Value-Based Decisions. *Psychonomic Bulletin & Review*, 26(4):1099–1121.

Frank, M. J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T. V., Cavanagh, J. F., and Badre, D. (2015). fMRI and EEG Predictors of Dynamic Decision Parameters During Human Reinforcement Learning. *Journal of Neuroscience*, 35(2):485–494.

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4):25–42.

Freundlieb, M., Kovács, Á. M., and Sebanz, N. (2016). When Do Humans Spontaneously Adopt Another's Visuospatial Perspective? *Journal of Experimental Psychology: Human Perception and Performance*, 42(3):401–412.

Gaskell, M. G. and Marslen-Wilson, W. D. (1996). Phonological Variation and Inference in Lexical Access. *Journal of Experimental Psychology: Human Perception and Performance*, 22(1):144–158.

Gil-Gómez de Liaño, B., Stablum, F., and Umiltà, C. (2016). Can Concurrent Memory Load Reduce Distraction? A Replication Study and Beyond. *Journal of Experimental Psychology: General*, 145(1):e1–e12.

Glöckner, A. and Bröder, A. (2011). Processing of Recognition Information and Additional Cues: A Model-Based Analysis of Choice, Confidence, and Response Time. *Judgment and Decision Making*, 6(1):23–42.

Glöckner, A. and Bröder, A. (2014). Cognitive Integration of Recognition Information and Additional Cues in Memory-Based Decisions. *Judgment and Decision Making*, 9(1):35–50.

Goldstein, D. G. and Gigerenzer, G. (2002). Models of Ecological Rationality: The Recognition Heuristic. *Psychological Review*, 109(1):75–90.

Goschke, T. (2000). Intentional Reconfiguration and Involuntary Persistence in Task set Switching. In *Control of Cognitive Processes: Attention and Performance*, pages 331—-335. MIT Press, Cambridge, MA.

Gyurkovics, M., Stafford, T., and Levita, L. (2020). Cognitive Control Across Adolescence: Dynamic Adjustments and Mind-Wandering. *Journal of Experimental Psychology: General*, 149(6):1017–1031.

Hawkins, G. E. and Heathcote, A. (2021). Racing Against the Clock: Evidence-Based Versus Time-Based Decisions. *Psychological Review*, 128(2):222.

Heathcote, A., Holloway, E., and Sauer, J. (2019). Confidence and Varieties of Bias. *Journal of Mathematical Psychology*, 90:31–46.

Hu, M. and Rahnev, D. (2019). Predictive Cues Reduce but Do Not Eliminate Intrinsic Response Bias. *Cognition*, 192:1–8.

Hübner, R., Steinhauser, M., and Lehle, C. (2010). A Dual-Stage Two-Phase Model of Selective Attention. *Psychological Review*, 117(3):759–784.

Jiménez, L. and Méndez, A. (2013). It Is Not What You Expect: Dissociating Conflict Adaptation from Expectancies in a Stroop Task. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1):271–284.

Jones, M. and Dzhafarov, E. N. (2014). Unfalsifiability and Mutual Translatability of Major Modeling Schemes for Choice Reaction Time. *Psychological Review*, 121(1):1–32.

Kahneman, D. and Tversky, A. (1972). Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology*, 3:430–454.

Kiani, R., Hanks, T. D., and 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):3017–3029.

Kim, S.-Y., Kim, M.-S., and Chun, M. M. (2005). Concurrent Working Memory Load Can Reduce Distraction. *Proceedings of the National Academy of Sciences*, 102(45):16524–16529.

Kirby, K. N. and Gerlanc, D. (2013). BootES: An R Package for Bootstrap Confidence Intervals on Effect Sizes. *Behavior Research Methods*, 45(4):905–927.

Knill, D. C. and Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge University Press.

Kunde, W. (2003). Sequential Modulations of Stimulus-Response Correspondence Effects Depend on Awareness of Response Conflict. *Psychonomic Bulletin & Review*, 10(1):198–205.

Lahiri, A. and Marslen-Wilson, W. (1991). The Mental Representation of Lexical Form: A Phonological Approach to the Recognition Lexicon. *Cognition*, 38(3):245–294.

Laming, D. R. J. (1968). *Information Theory of Choice-Reaction Times*. Academic Press, New York, NY.

Lavie, N., Hirst, A., De Fockert, J. W., and Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General*, 133(3):339–354.

Liefooghe, B., Hughes, S., Schmidt, J. R., and De Houwer, J. (2019). Stroop-Like Effects of Derived Stimulus–Stimulus Relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(2):327–349.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford University Press, New York.

Ludwig, J., Ahrens, F. K., and Achtziger, A. (2020). Errors, Fast and Slow: An Analysis of Response Times in Probability Judgments. *Thinking and Reasoning*, 26(4):627–639.

Luna, F. G., Telga, M., Vadillo, M. A., and Lupiáñez, J. (2020). Concurrent Working Memory Load May Increase or Reduce Cognitive Interference Depending on the Attentional Set. *Journal of Experimental Psychology: Human Perception and Performance*, 46(7):667–680.

MacCleod, C. M. (1991). Half a Century of Research on the Stroop Effect: An Integrative Review. *Psychological Bulletin*, 109(2):163–203.

Mandler, G. (1980). Recognizing: The Judgment of Previous Occurrence. *Psychological Review*, 87(3):252–271.

Mayr, U., Awh, E., and Laurey, P. (2003). Conflict Adaptation Effects in the Absence of Executive Control. *Nature Neuroscience*, 6(5):450–452.

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., and Forstmann, B. U. (2012). Bias in the Brain: A Diffusion Model Analysis of Prior Probability and Potential Payoff. *Journal of Neuroscience*, 32(7):2335–2343.

Muto, H., Matsushita, S., and Morikawa, K. (2019). Object's Symmetry Alters Spatial Perspective-Taking Processes. *Cognition*, 191:103987.

Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., and Ridderinkhof, K. R. (2003). Electrophysiological Correlates of Anterior Cingulate Function in a Go/No-Go Task: Effects of Response Conflict and Trial Type Frequency. *Cognitive, Affective, and Behavioral Neuroscience*, 3(1):17–26.

O'Grady, C., Scott-Phillips, T., Lavelle, S., and Smith, K. (2020). Perspective-Taking is Spontaneous but Not Automatic. *Quarterly Journal of Experimental Psychology*, 73(10):1605–1628.

Palmer, J., Huk, A. C., and Shadlen, M. N. (2005). The Effect of Stimulus Strength on the Speed and Accuracy of a Perceptual Decision. *Journal of Vision*, 5:376–404.

Pedersen, M. L., Frank, M. J., and Biele, G. (2017). The Drift Diffusion Model as the Choice Rule in Reinforcement Learning. *Psychonomic Bulletin & Review*, 24(4):1234–1251.

Pitt, M. A., Dilley, L., and Tat, M. (2011). Exploring the Role of Exposure Frequency in Recognizing Pronunciation Variants. *Journal of Phonetics*, 39(3):304–311.

Prince, M., Brown, S., and Heathcote, A. (2012). The Design and Analysis of State-Trace Experiments. *Psychological Methods*, 17(1):78–99.

Qamar, A. T., Cotton, R. J., George, R. G., Beck, J. M., Prezhdo, E., Laudano, A., Tolias, A. S., and Ma, W. J. (2013). Trial-to-Trial, Uncertainty-Based Adjustment of Decision Boundaries in Visual Categorization. *Proceedings of the National Academy of Sciences*, 110(50):20332–20337.

Rahnev, D. and Denison, R. N. (2018). Suboptimality in Perceptual Decision Making. *Behavioral and Brain Sciences*, 41.

Rahnev, D., Lau, H., and De Lange, F. P. (2011). Prior Expectation Modulates the Interaction Between Sensory and Prefrontal Regions in the Human Brain. *Journal of Neuroscience*, 31(29):10741–10748.

Ramsey, R., Darda, K. M., and Downing, P. E. (2019). Automatic Imitation Remains Unaffected Under Cognitive Load. *Journal of Experimental Psychology: Human Perception and Performance*, 45(5):601–615.

Raoelison, M. T., Thompson, V. A., and De Neys, W. (2020). The Smart Intuitor: Cognitive Capacity Predicts Intuitive Rather than Deliberate Thinking. *Cognition*, 204:1–14.

Ratcliff, R. (1978). A Theory of Memory Retrieval. *Psychological Review*, 85:59–108.

Ratcliff, R. (2002). A Diffusion Model Account of Response Time and Accuracy in a Brightness Discrimination Task: Fitting Real Data and Failing to Fit Fake but Plausible Data. *Psychonomic Bulletin & Review*, 9(2):278–291.

Ratcliff, R., Gomez, P., and McKoon, G. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review*, 111(1):159.

Ratcliff, R. and McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, 20(4):873–922.

Ratcliff, R. and Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, 9(5):347–356.

Ratcliff, R. and Rouder, J. N. (2000). A Diffusion Model Account of Masking in Two-Choice Letter Identification. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1):127–140.

Ratcliff, R., Smith, P. L., Brown, S. D., and McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Science*, 20:260–281.

Ratcliff, R., Van Zandt, T., and McKoon, G. (1999). Connectionist and Diffusion Models of Reaction Time. *Psychological Review*, 106(2):261–300.

Rogers, R. D. and Monsell, S. (1995). The Costs of a Predictable Switch Between Simple Cognitive Tasks. *Journal of Experimental Psychology: General*, 124(2):207–231.

Rosenthal, R. (1994). Parametric Measures of Effect Size. In Cooper, H. and Hedges, L. V., editors, *The Handbook of Research Synthesis*, pages 231–244. Russell Sage Foundation, New York.

Salvucci, D. D. and Taatgen, N. A. (2008). Threaded Cognition: An Integrated Theory of Concurrent Multitasking. *Psychological Review*, 115(1):101–130.

Scherbaum, S., Dshemuchadse, M., Fischer, R., and Goschke, T. (2010). How Decisions Evolve: The Temporal Dynamics of Action Selection. *Cognition*, 115(3):407–416.

Schmidt, J. R. and Weissman, D. H. (2014). Congruency Sequence Effects Without Feature Integration or Contingency Learning Confounds. *PLoS One*, 9(7):e102337.

Shiffrin, R. M. and Schneider, W. (1977). Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending and a General Theory. *Psychological Review*, 84(2):127–190.

Sidman, M. and Tailby, W. (1982). Conditional Discrimination vs. Matching to Sample: An Expansion of the Testing Paradigm. *Journal of the Experimental Analysis of Behavior*, 37(1):5–22.

Smith, P. L., Ratcliff, R., and Wolfgang, B. J. (2004). Attention Orienting and the Time course of Perceptual Decisions: Response Time Distributions with Masked and Unmasked Displays. *Vision Research*, 44(12):1297–1320.

Smith, P. L. and Van Zandt, T. (2000). Time-Dependent Poisson Counter Models of Response Latency in Simple Judgment. *British Journal of Mathematical and Statistical Psychology*, 53(2):293–315.

Steinberg, L. (2008). A Social Neuroscience Perspective on Adolescent Risk-Taking. *Developmental Review*, 28(1):78–106.

Steyvers, M., Hawkins, G. E., Karayanidis, F., and Brown, S. D. (2019). A Large-Scale Analysis of Task Switching Practice Effects Across the Lifespan. *Proceedings of the National Academy of Sciences*, 116(36):17735–17740.

Stroop, J. R. (1935). Studies of Interference in Serial Verbal Reactions. *Journal of Experimental Psychology*, 35:643–662.

Summerfield, C. and De Lange, F. P. (2014). Expectation in Perceptual Decision Making: Neural and Computational Mechanisms. *Nature Reviews Neuroscience*, 15(11):745–756.

Sumner, M. and Samuel, A. G. (2009). The Effect of Experience on the Perception and Representation of Dialect Variants. *Journal of Memory and Language*, 60(4):487–501.

Surtees, A., Apperly, I., and Samson, D. (2013). Similarities and Differences in Visual and Spatial Perspective-Taking Processes. *Cognition*, 129(2):426–438.

Surtees, A., Samson, D., and Apperly, I. (2016). Unintentional Perspective-Taking Calculates Whether Something Is Seen, but Not How It Is Seen. *Cognition*, 148:97–105.

Swensson, R. G. (1972). The Elusive Tradeoff: Speed vs. Accuracy in Visual Discrimination Tasks. *Perception & Psychophysics*, 12(1):16–32.

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

Townsend, J. T. and Liu, Y. (2020). Can the Wrong Horse Win: The Ability of Race Models to Predict Fast or Slow Errors. *Journal of Mathematical Psychology*, 97:1–12.

Tsetsos, K., Gao, J., McClelland, J. L., and Usher, M. (2012). Using Time-Varying Evidence to Test Models of Decision Dynamics: Bounded Diffusion vs. the Leaky Competing Accumulator Model. *Frontiers in Neuroscience*, 6:79.

Ulrich, R., Schröter, H., Leuthold, H., and Birngruber, T. (2015). Automatic and Controlled Stimulus Processing in Conflict Tasks: Superimposed Diffusion Processes and Delta Functions. *Cognitive Psychology*, 78:148–174.

Vickers, D. (1979). *Decision Processes in Visual Perception*. Academic Press.

Weber, E. U. and Johnson, E. J. (2009). Mindful Judgment and Decision Making. *Annual Review of Psychology*, 60:53–85.

Weissman, D. H. (2019). Interacting Congruency Effects in the Hybrid Stroop–Simon Task Prevent Conclusions Regarding the Domain Specificity or Generality of the Congruency Sequence Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(5):945–967.

Wexler, M., Duyck, M., and Mamassian, P. (2015). Persistent States in Vision Break Universality and Time Invariance. *Proceedings of the National Academy of Sciences*, 112(48):14990–14995.

White, C. N. and Curl, R. (2018). Cueing Effects in the Attentional Network Test: A Spotlight Diffusion Model Analysis. *Computational Brain & Behavior*, 1(1):59–68.

White, C. N., Ratcliff, R., and Starns, J. J. (2011). Diffusion Models of the Flanker Task: Discrete Versus Gradual Attentional Selection. *Cognitive Psychology*, 63(4):210–238.

APPENDICES

A Proofs of Theoretical Results

For each prediction in the article, we proceed to prove the most general result and then comment how the results in the different Basic and Extended Models derive from the results in this Appendix.

In particular, the analysis below refers to the multi-alternative case where the set of alternatives is given by $X = \{x_1, \ldots, x_n\}$. Results for the binary case are immediately obtained by setting $n = 2$. The analysis also consider non-decision times depending on alignment on conflict, t_A vs. t_C, and a process selection probability also depending on alignment or conflict, Δ_A vs. Δ_C. Results for the Basic Models a obtained setting $t_A = t_C$ and $\Delta_A = \Delta_C = \Delta$.

A.1 Prediction D1: Generalized Stroop Effect

We prove this result in the setting of Extended Model III. That is, expected response times might differ depending on the selected alternative, as long as (R'), and (R-I) hold. Note that the result does not require assumptions (R-D) or (P).

Theorem A.1. *Consider the multi-alternative case allowing for alternative-dependent expected response times. Assume (R'), (R-I), $t_C \geq t_A$, and $\Delta_C \leq \Delta_A$.*

(D1) Correct responses are slower in expectation in case of conflict than in case of alignment.

Proof. The expected response time of correct responses in case of alignment ($x^D = x^I$) is

$$E(t|x^D, \text{Alignment}) = t_A + \frac{(1-\Delta_A)P^D R^D(x^D) + \Delta_A P^I R^I(x^I)}{(1-\Delta_A)P^D + \Delta_A P^I} \tag{1}$$

and the expected response time of correct responses in case of conflict ($x^D \neq x^I$) is

$$E(t|x^D, \text{Conflict}) = t_C + \frac{(1-\Delta_C)P^D R^D(x^D) + \Delta_C P(x^D|I) R^I(x^D)}{(1-\Delta_C)P^D + \Delta_C P(x^D|I)}. \tag{2}$$

Since $R^I(x^I) \leq R^I(x)$ for all $x \neq x^I$ by (R-I), it follows that

$$E(t|x^D, \text{Conflict}) \geq t_C + \frac{(1-\Delta_C)P^D R^D(x^D) + \Delta_C P(x^D|I) R^I(x^I)}{(1-\Delta_C)P^D + \Delta_C P(x^D|I)}. \tag{3}$$

Thus, a sufficient condition for $E(t|x^D, \text{Conflict}) > E(t|x^D, \text{Alignment})$ is that the right-hand side of the last inequality is strictly larger than the right-hand side of (1). Since $t_C \geq t_A$, this holds if the fraction in the right-hand side of the last inequality is strictly larger than the fraction in the right-hand side of (1). After some straightforward computations, this latter condition holds if and only if

$$P^D \cdot \left[R^D(x^D) - R^I(x^I)\right] \cdot \left[\Delta_A(1-\Delta_C)P^I - (1-\Delta_A)\Delta_C P(x^D|I)\right] > 0$$

Since $R^D(x^D) > R^I(x^I)$ by (R'), the last inequality is fulfilled if

$$\frac{\Delta_A}{(1-\Delta_A)} P^I > \frac{\Delta_C}{(1-\Delta_C)} P(x^D|I)$$

which follows because $\Delta_A \geq \Delta_C$ and $P^I > P(x^D|I)$. □

If process response times are assumed not to depend on selected alternatives, assumption (R-I) holds vacuously and (R) implies (R'). Hence, in Basic Model I the result above implies Theorem 1 by setting $n = 2$, $t_A = t_C$ and $\Delta_A = \Delta_C = \Delta$. In Extended Model I, it implies that (D1) holds in Theorem 4, by setting $n = 2$. Note that, in Theorem 4, (D1) does not require assumption (P). In Extended Model II (the multi-alternative case), the result above implies Theorem 5 by setting $t_A = t_C$ and $\Delta_A = \Delta_C = \Delta$. It also implies that (D1) holds in Theorem 8. Again, in this case, assumption (P) is not needed. Last, for Extended Model III, the result above implies that (D1) holds in 9. Note, however, that neither (R-D) nor (P) are needed for this implication.

Corollary 4. *Consider the multi-alternative case allowing for alternative-dependent expected response times. Assume $R^D(x^D) = R^I(x^I)$, $t_C > t_A$, and $\Delta_C \leq \Delta_A$. Then (D1) holds.*

Proof. Since $R^D(x^D) = R^I(x^I)$, the computations in the proof of Theorem A.1 show that the fractions on the right-hand side of (1) and (5) are equal. Since $t_C > t_A$, the strict inequality follows. □

This Corollary shows that (D1) holds in Corollary 3 in Extended Model III. It also implies that property (D1) holds in Corollary 1 in Extended Model I by setting $n = 2$, and in Corollary 2 in Extended Model II.

A.2 Predictions D2 and D2': The Frequency of Errors

Again, we first prove these results in the setting of Extended Model III and then we show how this result implies properties (D2) and (D2') in all other models. Please note that (D2) and (D2') do not involve any statements about response times, and hence assumptions (R), (R'), (R-I), and (R-D), as well as conditions on t_C and t_A, have no bearing here.

Theorem A.2. *Consider the multi-alternative case allowing for alternative-dependent expected response times. Assume (P) and $\Delta_C \leq \Delta_A$.*

(D2) The proportion of correct responses is strictly smaller in case of conflict than in case of alignment.

(D2') The proportion of intuitive choices is strictly smaller in case of conflict than in case of alignment (when they are also correct).

Proof. We first prove (D2). The proportion of correct responses in case of alignment ($x^D = x^I$) is

$$P(x^D|\text{Alignment}) = (1 - \Delta_A) \cdot P^D + \Delta_A \cdot P^I,$$

and in case of conflict ($x^D \neq x^I$) it is

$$P(x^D|\text{Conflict}) = (1 - \Delta_C) \cdot P^D + \Delta_C \cdot P(x^D|I) < (1 - \Delta_C) \cdot P^D + \Delta_C \cdot P^I,$$

where the last inequality holds because $P^I > P(x^D|I)$ when $x^D \neq x^I$.

Hence, property (D2) holds if

$$(1 - \Delta_C) \cdot P^D + \Delta_C \cdot P^I \leq (1 - \Delta_A) \cdot P^D + \Delta_A \cdot P^I$$

or, equivalently,

$$(\Delta_A - \Delta_C)P^D \leq (\Delta_A - \Delta_C)P^I.$$

This last property holds because $\Delta_A \geq \Delta_C$ and, by (P), $P^I > P^D$.

We now turn to (D2'). The proportion of intuitive choices (x^I) in case of alignment ($x^D = x^I$) is the same as the proportion of correct responses in this case, given above. The proportion of intuitive choices in case of conflict ($x^D \neq x^I$) is

$$P(x^I|\text{Conflict}) = (1 - \Delta_C) \cdot P(x^I|D) + \Delta_C \cdot P^I < (1 - \Delta_C) \cdot P^D + \Delta_C \cdot P^I,$$

where the last inequality holds because $P^D > P(x^I|D)$ when $x^D \neq x^I$.

Hence, property (D2') reduces to the same computation as in the proof of (D2). □

This result implies Theorem 3 in Basic Model I setting $n = 2$ and $\Delta_C = \Delta_A$, as that result only requires assumption (P). It also implies that (D2) holds in Theorem 4 setting $n = 2$. The implication also holds in Corollary 1, as the additional conditions on response times do not affect this prediction. Theorem A.2 above also implies Theorem 7 in Extended Model II setting $\Delta_C = \Delta_A$, and directly implies that properties (D2) and (D2') hold in Theorem 8 and Corollary 2. Last, it again implies these properties in Theorem 9 and Corollary 3 as the only difference between those results and previous ones are assumptions on response times, which have no bearing on (D2) and (D2').

A.3 Predictions N1, N2, and N2': Neutral Trials

We now turn to Propositions 1 and 2. The following result collects the needed proofs.

Proposition 3. *Consider neutral trials.*

(N1) Assume (P). Both in the binary and in the multi-alternative cases, the proportion of correct responses in neutral trials is strictly smaller than the proportion of correct responses in case of alignment.

(N2) In the binary case, the proportion of correct responses in neutral trials is strictly larger than the proportion of correct responses in case of conflict.

(N2') Assume (P). In the multi-alternative case, the proportion of intuitive choices in neutral trials is strictly smaller than the proportion of intuitive choices in case of conflict.

Proof. We first prove (N1). The proportion of correct responses in neutral trials is P^D. In case of alignment ($x^D = x^I$), it is

$$P(x^D|\text{Alignment}) = (1 - \Delta_A) \cdot P^D + \Delta_A \cdot P^I.$$

Since $P^I > P^D$ by (P), (N1) follows.

We now turn to (N2). In the binary case, the proportion of correct responses in case of conflict ($x^D \neq x^I$) it is

$$P(x^D|\text{Conflict}) = (1 - \Delta_C) \cdot P^D + \Delta_C \cdot (1 - P^I) < (1 - \Delta_C) \cdot P^D + \Delta_C \cdot P^D = P^D,$$

where the inequality holds because $1 - P^I < 1/2 < P^D$.

Last, we prove (N2'). The proportion of intuitive choices in neutral trials is $P(x^I|D)$. In case of conflict ($x^D \neq x^I$), it is

$$P(x^I|\text{Conflict}) = (1 - \Delta_C) \cdot P(x^I|D) + \Delta_C \cdot P^I.$$

Hence (N2') holds if $P^I > P(x^I|D)$. This is true because $P^I > P^D$ by (P) and $P^D > P(x^I|D)$ because x^D is process D's modal answer. □

Proposition 1 follows from statements (N1) and (N2) in the result above, while Proposition 2 follows from statements (N1) and (N2').

A.4 Prediction T2: Slow Errors in Case of Alignment

As in previous sections, we establish this prediction for the most general Extended Model III and then we show how that result implies property (T2) for all other models in the main text. Note that assumptions on t_C vs. t_A and Δ_C vs. Δ_A are inconsequential for this result, as all variables involved refer to the case of alignment only.

Theorem A.3. *Consider the multi-alternative case allowing for alternative-dependent expected response times. Assume (R'), (R-D), (R-I), and (P).*

(T2) In case of alignment, the expected response time of errors is larger than the expected response time of correct answers.

Proof. The expected response time of correct responses in case of alignment ($x^D = x^I$) is as given in (1), and the expected response time of errors is

$$\begin{aligned} E(t|x \neq x^D, \text{Alignment}) &= t_A + \frac{(1-\Delta_A)\sum_{x\neq x^D} P^D(x)R^D(x) + \Delta_A \sum_{x\neq x^I} P^I(x)R^I(x)}{(1-\Delta_A)(1-P^D) + \Delta_A(1-P^I)} \\ &\geq t_A + \frac{(1-\Delta_A)(1-P^D)R^D(x^D) + \Delta_A(1-P^I)R^I(x^I)}{(1-\Delta_A)(1-P^D) + \Delta_A(1-P^I)}. \end{aligned}$$

where the inequality follows from (R-D) and (R-I). Thus, a sufficient condition for $E(t|x \neq x^D, \text{Alignment}) > E(t|x^D, \text{Alignment})$ is that the right-hand side of the last inequality is strictly larger than the right-hand side of (1). After some straightforward computations, this condition holds if and only if

$$(R^D(x^D) - R^I(x^I)) \cdot (P^I - P^D) > 0.$$

This inequality holds because $R^D(x^D) > R^I(x^I)$ by (R') and $P^I > P^D$ by (P). □

This result implies that (T2) holds in Theorem 3 because assumption (R) implies (R') and (R-D), (R-I) hold trivially in Basic Model II, where expected process response times are assumed not to depend on the chosen option. It also implies this prediction in Theorem 4 in Extended Model I, as the assumptions on t_C vs. t_A and Δ_C vs. Δ_A have no bearing on (T2). For the same reasons, Theorem A.3 implies (T2) in Theorems 7 and 8 in Extended Model II, as the only difference is that multiple alternatives are allowed in the those results. Last, the result directly implies that (T2) holds in Theorem 9 in Extended Model III.

A.5 Prediction T1: Fast Errors in Case of Conflict

We establish this prediction for Extended Model III and then we show how that result implies property (T1) for previous models. Note that, as in the previous subsection, assumptions on t_C vs. t_A and Δ_C vs. Δ_A are inconsequential for this result, as all variables involved refer to the case of conflict only.

Theorem A.4. *Consider the multi-alternative case, and assume (R') and (R-I). Further assume that, for the deliberative process D, expected process response times do not depend on chosen options.*

(T1) In case of conflict, the expected response time of intuitive errors is shorter than the expected response time of correct answers.

Proof. The expected response time of intuitive errors (choosing x^I) in case of conflict ($x^D \neq x^I$) is

$$E(t|x^I, \text{Conflict}) = t_C + \frac{(1-\Delta_C)P(x^I|D)R^D + \Delta_C P^I R^I(x^I)}{(1-\Delta_C)P(x^I|D) + \Delta_C P^I}. \tag{4}$$

The expected response time of correct responses in case of conflict is as given in (2), replacing $R^D(x^D)$ with R^D. By (R-I), $R^I(x) \geq R^I(x^I)$ for all $x \neq x^I$, and hence

$$E(t|x^D, \text{Conflict}) \geq t_C + \frac{(1-\Delta_C)P^D R^D + \Delta_C P(x^D|I)R^I(x^I)}{(1-\Delta_C)P^D + \Delta_C P(x^D|I)} \tag{5}$$

Thus, a sufficient condition for $E(t|x^D, \text{Conflict}) > E(t|x^I, \text{Conflict})$ is that the right-hand side of inequality (5) is strictly larger than the expression in (4). After some straightforward computations, this condition holds if and only if

$$\left[R^D - R^I(x^I)\right] \cdot \left[P^D P^I - P(x^I|D)P(x^D|I)\right] > 0.$$

This inequality holds because $P^D > P(x^I|D)$ and $P^I > P(x^D|I)$ (as $x^D \neq x^I$), and $R^D > R^I(x^I)$ by (R'). □

This result proves Theorem 10(a), and immediately implies (T1) in Theorems 3 (Basic Model) and 7 (Extended Model II). Since assumptions on t_C vs. t_A and Δ_C vs. Δ_A have no impact on (T1), the prediction also follows immediately in Theorems 4 and 8.

The following result proves Theorem 10(b), which captures the idea that (T1) still holds provided that intuitive options selected by the deliberative process D are not too slow. We remark that, formally, this result also implies Theorem A.4, but we keep the proof of that result explicit for clarity.

Theorem A.5. *Consider the multi-alternative case, and assume (R') and (R-I). Fix all process parameters except for $R^D(x^I)$. Then, there exists $\bar{R} > R^D(x^D)$ such that (T1) as in Theorem A.4 holds whenever $R^D(x^I) < \bar{R}$.*

Proof. The expected response time of intuitive errors (choosing x^I) in case of conflict ($x^D \neq x^I$) is

$$E(t|x^I, \text{Conflict}) = t_C + \frac{(1-\Delta_C)P(x^I|D)R^D(x^I) + \Delta_C P^I R^I(x^I)}{(1-\Delta_C)P(x^I|D) + \Delta_C P^I}.$$

The expected response time of correct responses in case of conflict ($x^D \neq x^I$) is

$$\begin{aligned} E(t|x^D, \text{Conflict}) &= t_C + \frac{(1-\Delta_C)P^D R^D(x^D) + \Delta_C P(x^D|I)R^I(x^D)}{(1-\Delta_C)P^D + \Delta_C P(x^D|I)} \\ &\geq \frac{(1-\Delta_C)P^D R^D(x^D) + \Delta_C P(x^D|I)R^I(x^I)}{(1-\Delta_C)P^D + \Delta_C P(x^D|I)}, \end{aligned}$$

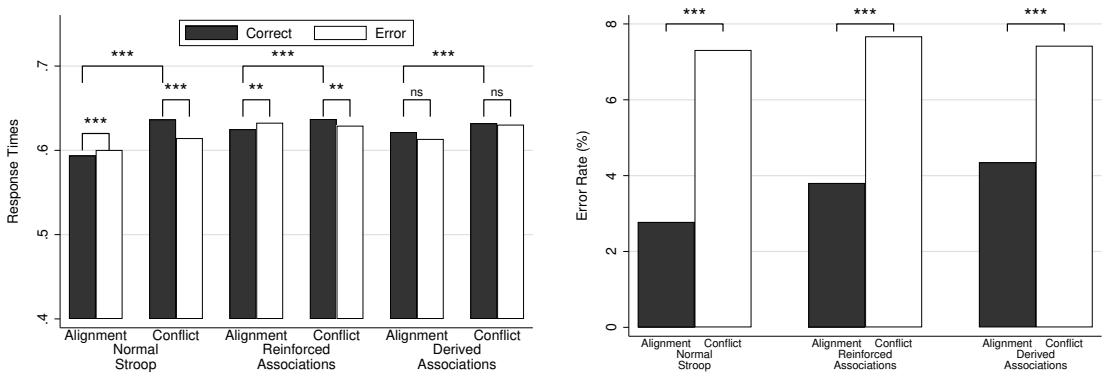

Figure A.1: Analyses of the Stroop task by Liefooghe et al. (2019). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$, ** $p < .05$.

where the inequality follows from (R-I) as in the proof of Theorem A.4 Thus, a sufficient condition for $E(t|x^D, \text{Conflict}) > E(t|x^I, \text{Conflict})$ is that

$$\frac{(1-\Delta_C)P^D R^D(x^D) + \Delta_C P(x^D|I) R^I(x^I)}{(1-\Delta_C)P^D + \Delta_C P(x^D|I)} > \frac{(1-\Delta_C)P(x^I|D) R^D(x^I) + \Delta_C P^I R^I(x^I)}{(1-\Delta_C)P(x^I|D) + \Delta_C P^I}.$$

Straightforward but cumbersome computations show that the last inequality is equivalent to

$$R^D(x^I) < R^D(x^D) + \frac{\Delta_C \left(R^D(x^D) - R^I(x^I)\right) \left[P^D P^I - P(x^I|D)P(x^D|I)\right]}{P(x^I|D)\left[\Delta_C P(x^D|I) + (1-\Delta_C)P^D\right]}$$

hence the statement of the Theorem holds taking the right-hand side of this inequality as $\bar{R}$. Note that the denominator of the fraction is always strictly positive, and the numerator is also strictly positive by (R') and because $P^D > P(x^I|D)$ and $P^I > P(x^D|I)$ (as $x^D \neq x^I$). Hence, $\bar{R} > R^D(x^D)$ as claimed. □

B Detailed Analysis of Predicted Effects for the Individual Datasets

B.1 Cognitive Control

Dataset 1: Stroop Effects and Derived Associations. Liefooghe et al. (2019) reported five experiments encompassing 57, 54, 59, 49, and 56 subjects, respectively. The experiments differed in details of the training phases and the implementation. For instance, Experiment 2 included Go/No-Go trials intermixed with the Stroop ones, and Experiments 3–5 used four colors, mapped however into two keys. Participants went through 360, 480, 576, 576, and 768 trials, respectively. For the sake of brevity, we report our analysis pooling all five experiments ($N = 275$). Figure A.1 summarizes the results for normal Stroop trials, reinforced associations, and derived associations. For normal Stroop trials, all predictions hold, although response time differences are small due to the fast nature of the task. Correct answers are slower in conflict than in alignment (Prediction D1, 0.637 vs. 0.594 s; $N = 273$, $z = 11.625$, $p < 0.001$, $r = 0.704$), and error rates are larger in conflict than in alignment (Prediction D2, 7.31% vs. 2.78%;

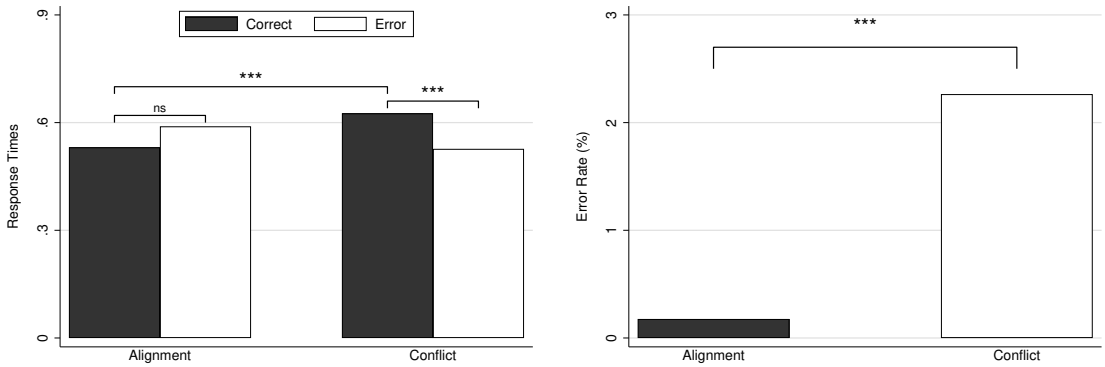

Figure A.2: Analyses of the Simon task by Gyurkovics et al. (2020). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

$N = 274$, $z = 8.160$, $p < 0.001$, $r = 0.493$; Figure A.1, right panel). Further, errors in conflict are faster than correct answers (Prediction T1, 0.615 vs. 0.637 s; $N = 217$, $z = -4.375$, $p < 0.001$, $r = 0.300$), while the opposite is true in alignment, where errors are slower than correct choices (Prediction T2, 0.600 vs. 0.594 s; $N = 151$, $z = 3.827$, $p < 0.001$, $r = 0.311$).

All predictions also hold for trials with reinforced associations. Again, correct answers are slower and error rates are larger in conflict than in alignment (D1, 0.637 vs. 0.625 s; $N = 273$, $z = 4.636$, $p < 0.001$, $r = 0.281$; D2, 7.67% vs. 3.81%; $N = 273$, $z = 7.812$, $p < 0.001$, $r = 0.473$). Also, errors in conflict are faster than correct answers (T1, 0.629 vs. 0.637 s; $N = 238$, $z = -2.241$, $p = 0.025$, $r = 0.145$), but the opposite is true in alignment, (T2, 0.633 vs. 0.625; $N = 181$, $z = 2.454$, $p = 0.014$, $r = 0.182$).

The predicted relations only hold partially for trials with derived associations, suggesting that the two processes do not differ in terms of response times (Corollary 1). Specifically, D1 and D2 hold: correct answers are slower and error rates are higher in conflict than in alignment (D1, 0.632 vs. 0.622 s; $N = 273$, $z = 4.454$, $p < 0.001$, $r = 0.270$; D2, 7.42% vs. 4.35%; $N = 273$, $z = 5.281$, $p < 0.001$, $r = 0.320$). However, T1 and T2 do not: there are no significant differences in response times between errors and correct answers neither in conflict (T1, 0.630 vs. 0.632 s; $N = 215$, $z = -0.525$, $p = 0.599$, $r = 0.036$) nor in alignment (T2, 0.613 vs. 0.622 s; $N = 207$, $z = 0.802$, $p = 0.4226$, $r = 0.056$).

Dataset 2: Simon Task. The dataset of Gyurkovics et al. (2020) comprises $N = 118$ participants who completed 291 trials of the Simon task each. Figure A.2 illustrates the results. Correct answers are slower and error rates are larger in conflict than in alignment (D1, 0.645 vs. 0.555 s; $N = 118$, $z = 9.427$, $p < 0.001$, $r = 0.868$; D2, 2.26% vs. 0.18%; $N = 118$, $z = 8.822$, $p < 0.001$, $r = 0.812$; Figure A.2, right panel). Errors are faster in conflict than correct answers (T1, 0.558 vs. 0.645 s; $N = 88$, $z = -5.176$, $p < 0.001$, $r = 0.552$). In alignment, the overwhelming majority of participants had zero error rates and hence the sample size is greatly reduced. The comparison still shows that errors are slower than correct answers in this case, but the difference is not significant (T2, 0.588 vs. 0.555 s; $N = 17$, $z = 0.308$, $p = 0.782$, $r = 0.190$).

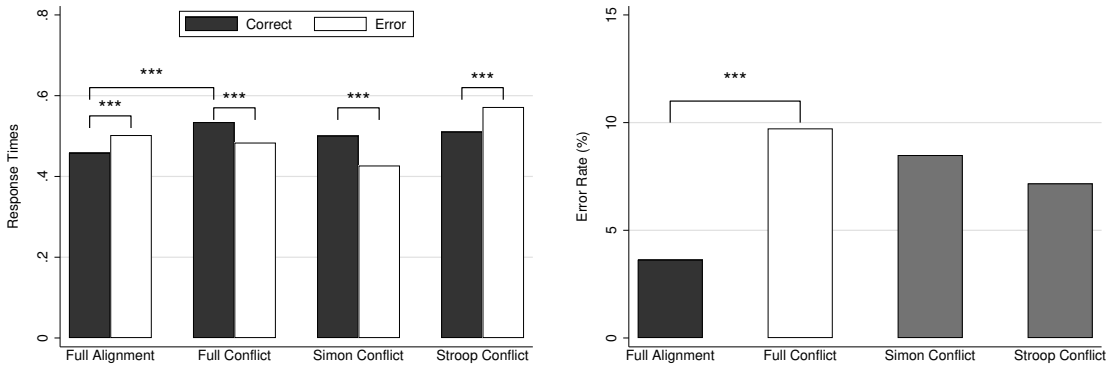

Figure A.3: Analyses of the hybrid Stroop-Simon task task by Weissman (2019). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

Dataset 3: Hybrid Stroop-Simon Task. Ninety participants in the two experiments of Weissman (2019) ($N = 58$ and $N = 32$) completed 768 trials each. We again report our analysis pooling them together, but results are qualitatively unchanged when we look at the experiments separately. Figure A.3 illustrates the results. The predictions of the model only apply when the two alternative processes are aligned, and hence can be summarized as one. This corresponds to full alignment and full conflict trials. Indeed, correct answers are slower and error rates are larger in (full) conflict than in (full) alignment (D1, 0.519 vs. 0.447 s; $N = 90$, $z = 8.239$, $p < 0.001$, $r = 0.868$; D2, 9.73% vs. 3.63%; $N = 90$, $z = 7.563$, $p < 0.001$, $r = 0.797$; Figure A.3, right panel). Also, errors are faster than correct answers in (full) conflict, and the opposite is true in (full) alignment (T1, 0.469 vs. 0.519 s; $N = 90$, $z = -6.211$, $p < 0.001$, $r = 0.655$; T2, 0.508 vs. 0.447 s $N = 88$, $z = 4.689$, $p < 0.001$, $r = 0.500$).

For Simon-conflict and Stroop-conflict trials, the model does not apply *a priori*, since conflict with one alternative process is actually alignment with the other one. Although we had no predictions for those situations, it is still interesting to examine them. Simon conflict (hence Stroop alignment) trials behave as one would expect for the case of conflict, with correct answers being slower than errors (T1, 0.488 vs. 0.420 s; $N = 90$, $z = 6.794$, $p < 0.001$, $r = 0.716$). Stroop conflict (hence Simon alignment) behave as one would expect for the case of alignment, with correct answers being faster than errors (T2, 0.497 vs. 0.560 s; $N = 89$, $z = 6.041$, $p < 0.001$, $r = 0.640$). These results suggest that the process underlying the Simon effect dominates the one responsible for the Stroop effect. However, error rates are larger both for Simon conflict (8.49%) and for Stroop conflict (7.18%) compared to full alignment (3.63%; $N = 90$; Simon conflict, $z = 7.357$, $p < 0.001$, $r = 0.775$; Stroop conflict, $z = 7.162$, $p < 0.001$, $r = 0.755$), further attesting the active influence of both processes in this context.

Dataset 4: A Standard Flanker Task The three experiments of Luna et al. (2020) comprise a total of ninety-two participants ($N = 48$, 20, and 24, respectively) who completed 480, 320, and 416 trials each, respectively. For the sake of brevity, we report our analysis pooling the three experiments. Results are qualitatively unchanged when we consider the experiments separately (horizontal vs. vertical displacement was implemented between subjects in the first and within in the second and third). Participants were also instructed to withhold action if a time counter (in milliseconds) appeared on-

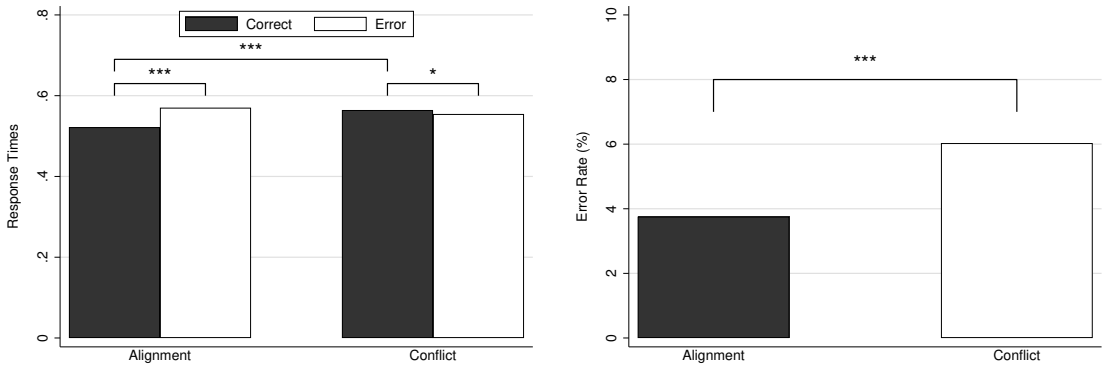

Figure A.4: Analyses of the Flanker data from Luna et al. (2020). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$, * $p < .1$

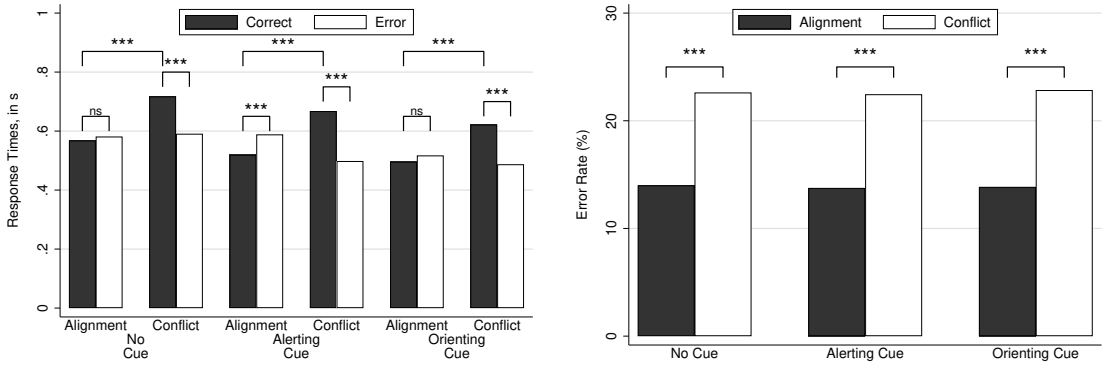

Figure A.5: Analyses of the cued-Flanker data from White and Curl (2018). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

screen, which happened in around 11.69% of the trials. We exclude these trials from our analysis.

Figure A.4 shows that our predictions find full support in this context. Correct answers are slower and error rates are larger in conflict than in alignment (D1, 0.564 vs. 0.522 s; $N = 92$, $z = 8.321$, $p < 0.001$, $r = 0.868$; D2, 6.03% vs. 3.76%; $N = 92$, $z = 5.444$, $p < 0.001$, $r = 0.568$, Figure A.4, right panel). Errors are faster than correct answers in conflict (T1, 0.555 vs. 0.564 s; $N = 87$, $z = -1.917$, $p = 0.055$, $r = 0.206$), and slower in alignment (T2, 0.570 vs. 0.522 s; $N = 76$, $z = 4.230$, $p < 0.001$, $r = 0.485$).

Dataset 5: Attention and the Flanker Task The data of White and Curl (2018) includes 123 participants, each of which went through 576 trials, although some trials were coded as "no response" if the answer was not given within 1.5 s. Figure A.5 illustrates the results. Correct answers are slower in conflict than in alignment in all cueing conditions (D1, no cue: 0.720 vs. 0.529 s; $N = 122$, $z = 8.027$, $p < 0.001$, $r = 0.727$; alerting cue: 0.680 vs. 0.529 s; $N = 118$, $z = 8.981$, $p < 0.001$, $r = 0.827$; orienting cue: 0.634 vs. 0.514 s; $N = 118$, $z = 8.541$, $p < 0.001$, $r = 0.786$). Error rates are larger in conflict than in alignment (D2, no cue: 22.62% vs. 14.00%; $N = 123$,

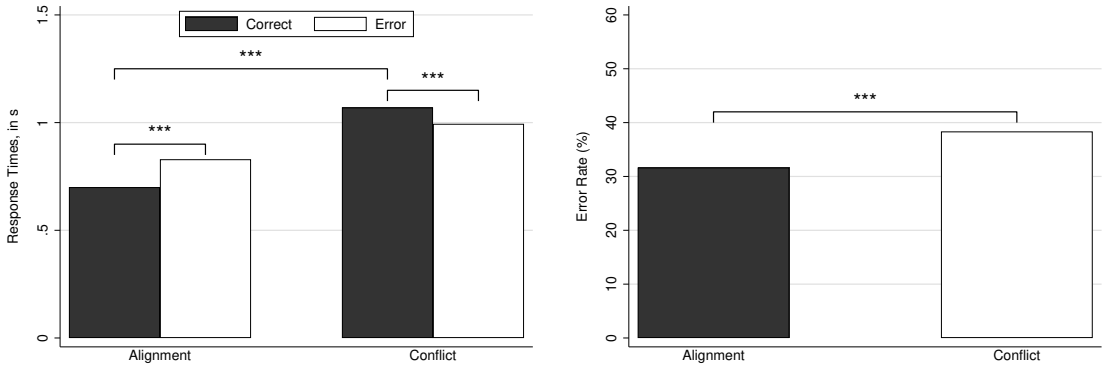

Figure A.6: Analysis of the visual attention task of Denison et al. (2018). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

$z = 7.577$, $p < 0.001$, $r = 0.683$; alerting cue: 22.46% vs. 13.76%; $N = 123$, $z = 6.824$, $p < 0.001$, $r = 0.615$; orienting cue: 22.85% vs. 13.86%; $N = 123$, $z = 7.241$, $p < 0.001$, $r = 0.653$; Figure A.5, right panel). Errors are faster in conflict than correct answers (T1, no cue: 0.611 vs. 0.720 s, $N = 116$, $z = 5.524$, $p < 0.001$, $r = 0.513$; alerting cue: 0.557 vs. 0.680 s; $N = 118$, $z = -6.576$, $p < 0.001$, $r = 0.605$; orienting cue: 0.542 vs. 0.634 s; $N = 111$, $z = -5.667$, $p < 0.001$, $r = 0.538$). In alignment errors are slower than correct answers for alerting cues (T2, 0.604 vs. 0.529 s; $N = 46$, $z = 2.562$, $p = 0.009$, $r = 0.378$). However, there were no significant differences for the other two cueing conditions (T2, no cue: 0.585 vs. 0.579 s; $N = 61$, $z = 0.269$, $p = 0.792$, $r = 0.034$; orienting cue: 0.531 vs. 0.514 s; $N = 51$, $z = 0.150$, $p = 0.886$, $r = 0.021$).

B.2 Attentional Processes

Dataset 6: Attention and Perceptual Decisions The dataset of Denison et al. (2018) includes $N = 12$ participants with about 2,000 trials each (collected in five different test sessions). Figure A.6 shows that our predictions find full support in this context. Correct answers are slower and error rates are larger in conflict than in alignment (D1, 1.157 vs. 0.759 s; $N = 12$, $z = 3.059$, $p < 0.001$, $r = 0.883$; D2, 38.36% vs. 31.65%; $N = 12$, $z = 3.059$, $p < 0.001$, $r = 0.883$; Figure A.6, right panel). Errors are faster than correct answers in conflict (T1, 1.083 vs. 1.157 s; $N = 12$, $z = -2.903$, $p = 0.001$, $r = 0.838$) and slower in alignment (T2, 0.869 vs. 0.759 s; $N = 12$, $z = 2.510$, $p = 0.009$, $r = 0.725$).

Dataset 7: Perceptual Decisions and Initial Cues In the experiment of Evans et al. (2017), seventy participants completed 480 trials each. Figure A.7 shows that our predictions are again fully supported. Correct answers are slower and error rates are larger in conflict than in alignment (D1, 2.144 vs. 2.015 s; $N = 70$, $z = 4.427$, $p < 0.001$, $r = 0.529$; D2, 25.88% vs. 14.84%; $N = 70$, $z = 6.727$, $p < 0.001$, $r = 0.804$). Errors are faster than correct answers in conflict (T1, 1.845 vs. 2.144 s; $N = 70$, $z = -6.072$, $p < 0.001$, $r = 0.726$) and slower in alignment (T2, 2.391 vs, 2.015 s; $N = 70$, $z = 4.427$, $p < 0.001$, $r = 0.529$).

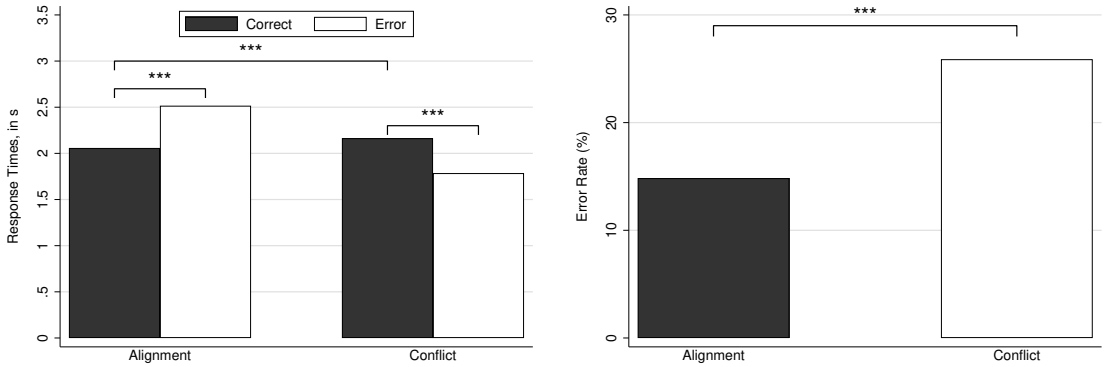

Figure A.7: Analyses of the direction of motion task by Evans et al. (2017). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

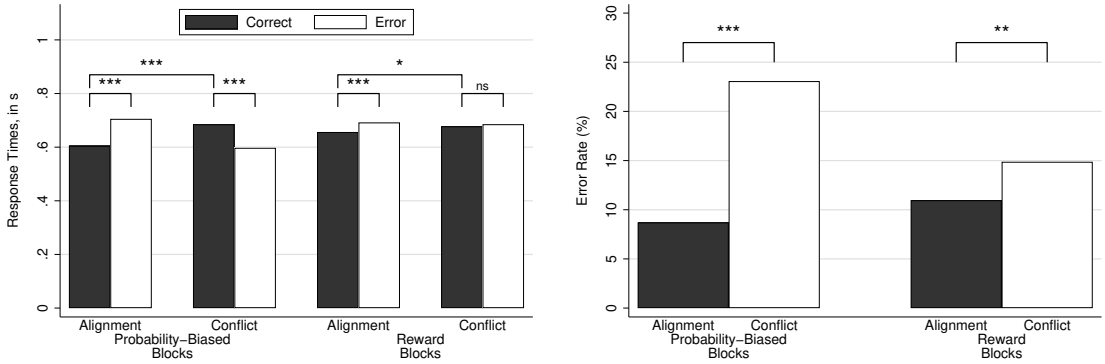

Figure A.8: Analyses of the perceptual binary decision task by Heathcote et al. (2019). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$, ** $p < .05$, * $p < .1$.

Dataset 8: Judging Majority Colors Thirty-two participants in Heathcote et al. (2019) completed 320 trials each, divided among unbiased, reward, and probability-biased blocks. Figure A.8 shows that, for probability-manipulated blocks, our predictions find full support. Correct answers are slower and error rates are larger in conflict than in alignment (D1, 0.685 vs. 0.605 s; $N = 32$, $z = 4.824$, $p < 0.001$, $r = 0.853$; D2, 23.08% vs. 8.70%; $N = 32$, $z = 4.357$, $p < 0.001$, $r = 0.770$; Figure A.8, right panel). Errors are faster than correct answers in conflict (T1, 0.597 vs. 0.685 s; $N = 32$, $z = -4.432$, $p < 0.001$, $r = 0.783$) and slower in alignment (T2, 0.705 vs. 0.605 s; $N = 32$, $z = 4.824$, $p < 0.001$, $r = 0.853$).

For reward blocks we find only partial support for our predictions, in line with the consideration that the experiment was not actually incentivised, hence differences in the magnitude of purely hypothetical rewards might actually have played a modest role. Correct answers are marginally significantly slower in conflict than in alignment (D1, 0.678 vs. 0.656 s; $N = 32$, $z = 1.926$, $p = 0.055$, $r = 0.340$). Errors are larger in conflict than in alignment (D2, 14.87% vs. 10.97%; $N = 32$, $z = 2.001$, $p = 0.045$, $r = 0.354$). In conflict, response times for errors and correct answers are not significantly different

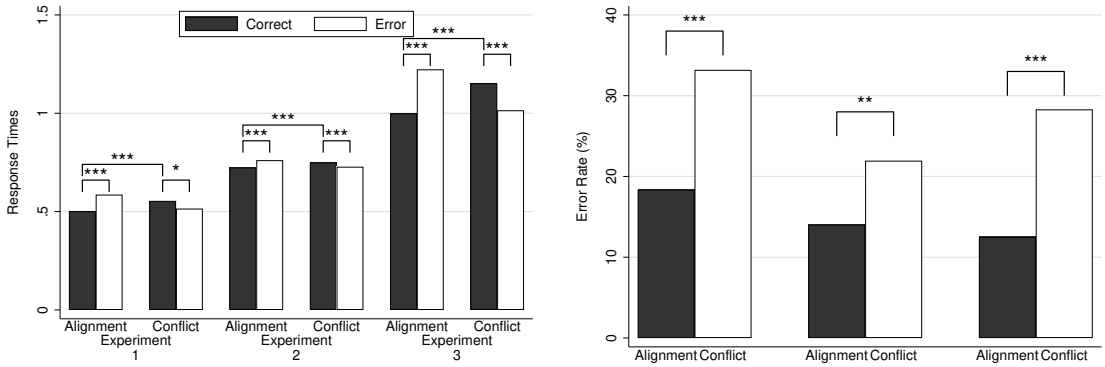

Figure A.9: Analyses of the three categorization judgment tasks by Hu and Rahnev (2019). Left: Average response times (in seconds) for errors and correct answers conditional on a trial being in alignment or in a neutral case. Right: Error rates in alignment vs. neutral. WRS tests: *** $p < .01$, ** $p < .05$, * $p < .1$.

(T1, 0.685 vs. 0.678 s; $N = 32$, $z = -0.692$, $p = 0.500$, $r = 0.122$). In alignment errors are slower than correct answers (T2, 0.692 s vs. 0.656 s; $N = 32$, $z = 2.824$, $p = 0.004$, $r = 0.499$).

Dataset 9: Predictive Cues and Categorization In the experiments collected in Hu and Rahnev (2019), thirty, twenty-two, and twenty-one participants completed 480, 672, and 864 trials, respectively. Since they used different tasks, we analyze them separately. Figure A.9 shows that our predictions are confirmed in all three experiments. Correct answers are slower in conflict than alignment (D1, Bang and Rahnev (2017): 0.554 vs. 0.503 s; $N = 29$, $z = 2.606$, $p < 0.001$, $r = 0.484$; de Lange et al. (2013): 0.750 vs. 0.724 s; $N = 23$, $z = 2.768$, $p = 0.004$, $r = 0.577$; Rahnev et al. (2011): 1.153 vs. 1.000 s; $N = 21$, $z = 3.806$, $p < 0.001$, $r = 0.831$). Error rates are larger in conflict than in alignment (D2, Bang and Rahnev (2017): 33.19% vs. 18.36%; $N = 30$, $z = 4.299$, $p < 0.001$, $r = 0.785$; de Lange et al. (2013): 21.94% vs. 14.03%; $N = 23$, $z = 2.251$, $p = 0.023$, $r = 0.469$; Rahnev et al. (2011): 28.30% vs. 12.51%; $N = 21$, $z = 4.015$, $p < 0.001$, $r = 0.876$). Errors are faster in conflict than correct answers (T1, Bang and Rahnev (2017): 0.515 vs. 0.554 s; $N = 29$, $z = -1.849$, $p = 0.065$, $r = 0.343$; de Lange et al. (2013): 0.728 vs. 0.750 s; $N = 23$, $z = -2.859$, $p = 0.003$, $r = 0.596$; Rahnev et al. (2011): 1.015 vs. 1.153 s; $N = 21$, $z = -3.493$, $p < 0.001$, $r = 0.762$). Last, in alignment, errors are slower than correct answers (T2, Bang and Rahnev (2017): 0.586 vs. 0.503 s; $N = 28$, $z = 2.960$, $p = 0.002$, $r = 0.559$; de Lange et al. (2013): 0.761 vs. 0.724 s; $N = 23$, $z = 3.198$, $p < 0.001$, $r = 0.667$; Rahnev et al. (2011): 1.223 vs. 1.000 s; $N = 20$, $z = 3.696$, $p < 0.001$, $r = 0.826$).

B.3 Social Cognition

Dataset 10: Automatic Imitation of Social Gestures The three experiments in Ramsey et al. (2019) yielded data for 58, 55, and 59 subjects, respectively. Each participant went through 256 trials, half of them under each load treatment. For the sake of brevity, we report our analysis pooling all three experiments ($N = 172$), while keeping the cognitive load treatments (low and high load) separate. Results are qualitatively unchanged when we look at the three experiments separately. Figure A.10 shows that

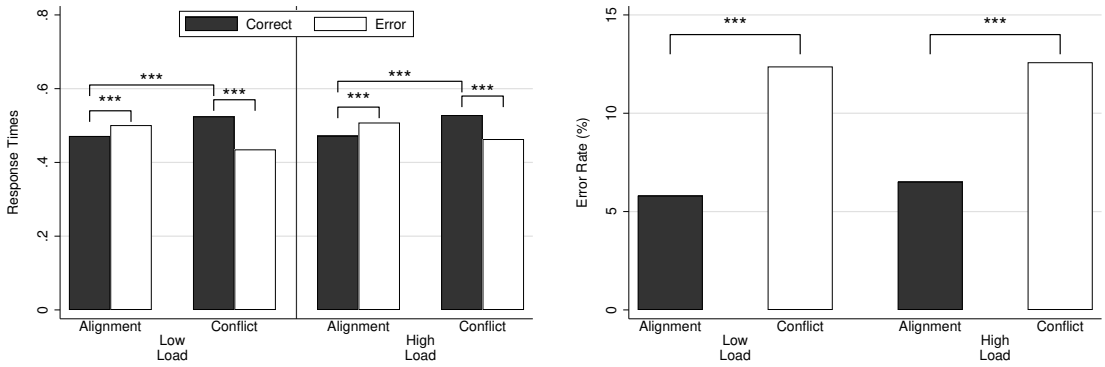

Figure A.10: Analyses of the three experiments on automatic imitation by Ramsey et al. (2019). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict further conditional on the cognitive load manipulation. Right: Error rates in alignment vs. conflict conditional on the cognitive load manipulation. WRS tests: *** $p < .01$.

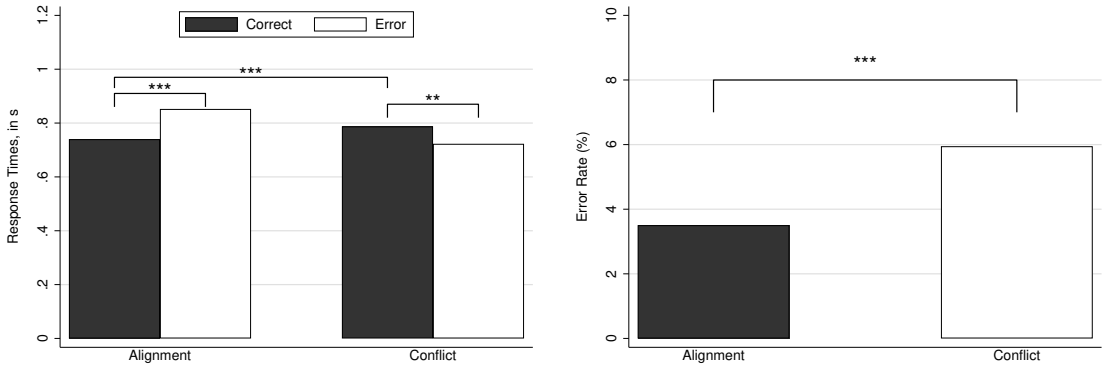

Figure A.11: Analyses of the perspective-taking task by O'Grady et al. (2020). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$, ** $p < .05$.

all our predictions are also confirmed in this setting, independently of the cognitive load treatment. Specifically, under low cognitive load, correct answers are slower and error rates are larger in conflict than in alignment (D1, 0.551 vs. 0.488 s; $N = 165$, $z = 9.909$, $p < 0.001$, $r = 0.771$; D2, 12.45% vs. 5.84%; $N = 172$, $z = 8.672$, $p < 0.001$, $r = 0.661$; Figure A.10, right panel). Errors are faster than correct answers in conflict (T1, 0.498 vs. 0.551 s; $N = 140$, $z = 5.483$, $p < 0.001$, $r = 0.463$) but slower in alignment (T2, 0.579 vs. 0.488 s; $N = 111$, $z = 3.448$, $p < 0.001$, $r = 0.327$). All results hold also under high cognitive load. Again, correct answers are slower and error rates are larger in conflict than in alignment (D1, 0.551 vs. 0.493 s; $N = 169$, $z = 10.608$, $p < 0.001$, $r = 0.816$; D2, 12.66%vs. 6.51%; $N = 172$, $z = 9.115$, $p < 0.001$, $r = 0.695$). Errors are faster than correct answers in conflict (T1, 0.499 vs. 0.551 s; $N = 142$, $z = -4.790$, $p < 0.001$, $r = 0.402$) but slower in alignment (T2, $N = 169$, $z = 10.608$, $p < 0.001$, $r = 0.816$).

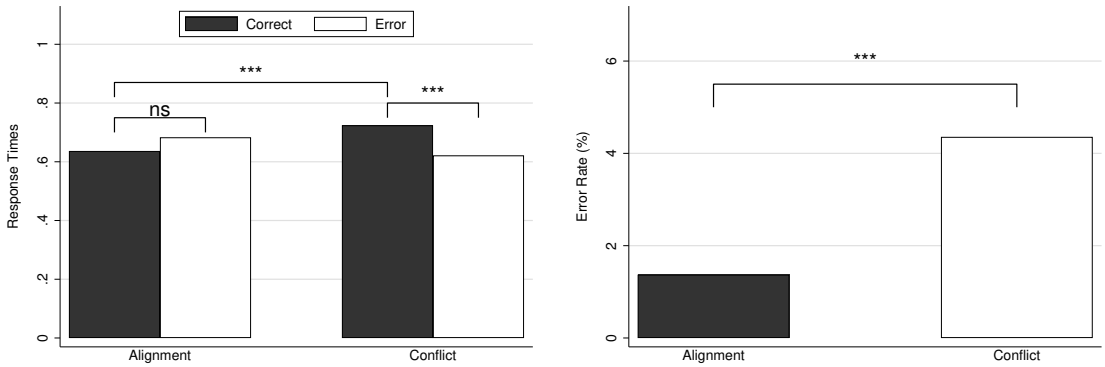

Figure A.12: Analyses of the perspective-taking task by Muto et al. (2019). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: ** $p < .05$, *** $p < .01$.

Dataset 11: Perspective Taking (Numerosity) The relevant condition of O'Grady et al. (2020) encompassed 30 participants who completed 256 trials each. Figure A.11 shows that our predictions find full support in this context. Correct answers are slower and error rates are larger in conflict than in alignment (D1, 0.787 vs. 0.740 s; $N = 30$, $z = 3.893$, $p < 0.001$, $r = 0.711$; D2, 5.95% vs. 3.50%; $N = 30$, $z = 3.422$, $p = 0.002$, $r = 0.625$; Figure A.11, right panel). Errors are faster than correct answers in conflict (T1, 0.723 vs. 0.787 s; $N = 29$, $z = -2.451$, $p = 0.013$, $r = 0.455$) and slower in alignment (T2, 0.852 vs. 0.740 s; $N = 21$, $z = 2.669$, $p = 0.005$, $r = 0.582$).

Dataset 12: Perspective Taking (Direction) For the sake of brevity, we report our analysis pooling the three target experiments in Muto et al. (2019). Results are qualitatively unchanged when we consider the experiments separately. Each experiment comprised 12 subjects with 320 trials each, of which 160 involved symmetric objects in the sense described above. Figure A.12 shows that our predictions find full support in this context, but for prediction T2. Correct answers are slower and error rates are larger in conflict than in alignment (D1, 0.724 vs. 0.636 s; $N = 36$, $z = 11.625$, $p < 0.001$, $r = 0.704$; D2, 4.36% vs. 1.37%; $N = 36$, $z = 4.140$, $p < 0.001$, $r = 0.690$; Figure A.12, right panel). Errors are faster than correct answers in conflict (T1, 0.621 vs. 0.724 s; $N = 34$, $z = -4.864$, $p < 0.001$, $r = 0.834$). and slower in alignment, but the latter difference is not statistically significant (T2, 0.683 vs. 0.636 s; $N = 26$, $z = 11.625$, $p < 0.001$, $r = 0.704$).

B.4 Memory

Dataset 13: Spoken Word Recognition In Experiment 1 of Charoy and Samuel (2020), seventy-seven participants completed 32 relevant trials (control trials with nonwords or familiar words are not relevant for our analysis). For our analysis, we ignore the type of written spelling which accompanied the spoken stimuli. Experiment 2 included a 48-hour delay between training associations and actual testing. Figure A.13 shows that our predictions find full support in this context. For Experiment 1, correct answers are slower and error rates are larger in conflict than in alignment (D1, 0.933 vs. 0.891 s; $N = 77$, $z = 7.548$, $p < 0.001$, $r = 0.860$; D2, 48.21% vs. 4.79%; $N = 77$, $z = 7.722$,

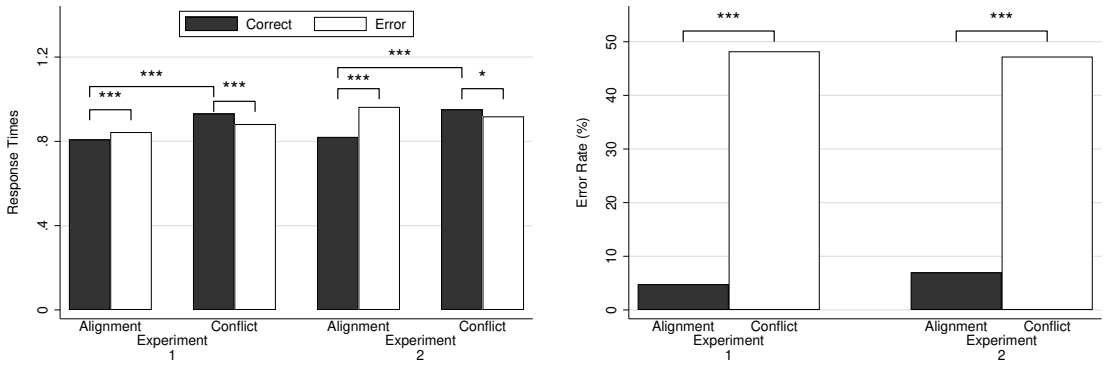

Figure A.13: Analyses of the recognition task by Charoy and Samuel (2020). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$ and * $p < .1$.

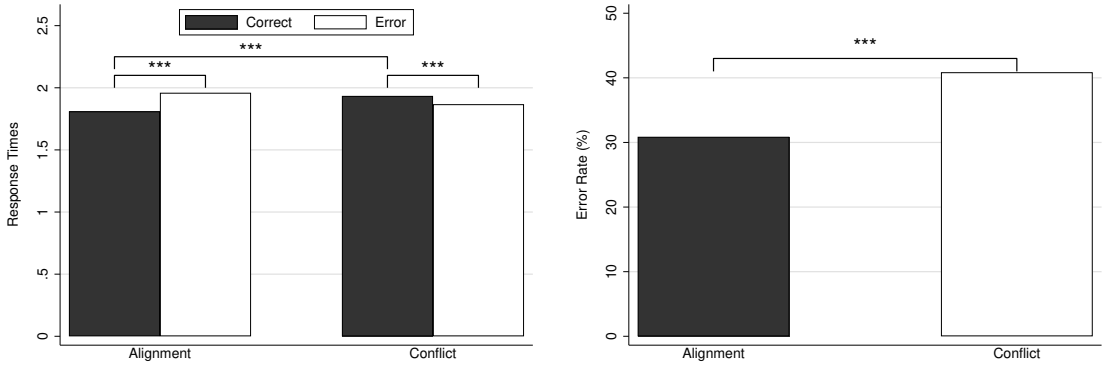

Figure A.14: Analyses of the false-memory task by Brainerd and Lee (2019). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

$p < 0.001$, $r = 0.880$; Figure A.13, right panel). Errors are faster than correct answers in conflict (T1, 0.893 vs. 0.933 s; $N = 77$, $z = -2.536$, $p = 0.010$, $r = 0.289$) and slower in alignment (T2, 0.807 vs. 0.891 s; $N = 46$, $z = 3.939$, $p < 0.001$, $r = 0.581$). For Experiment 2, again correct answers are slower and error rates are larger in conflict than in alignment (D1, 0.961 vs. 0.843 s; $N = 74$, $z = 7.125$, $p < 0.001$, $r = 0.828$; D2, 47.23% vs. 6.99%; $N = 76$, $z = 7.373$, $p < 0.001$, $r = 0.846$). Errors are marginally faster than correct answers in conflict (T1, 0.928 vs. 0.961 s; $N = 74$, $z = -1.678$, $p = 0.094$, $r = 0.195$), and slower in alignment (T2, 0.978 vs. 0.843 s; $N = 69$, $z = 5.557$, $p < 0.001$, $r = 0.669$).

Dataset 14: Recollection Processes In Brainerd and Lee (2019), one-hundred and eighty-five participants in six very similar experiments completed 225 trials each. We again report our analysis pooling the six experiments and all trial types together, but looking at them separately reveals qualitatively similar patterns. Figure A.14 shows that our predictions find full support in this context. Correct answers are slower and error rates are larger in conflict than in alignment (D1, 1.934 vs. 1.752 s; $N = 183$, $z = 9.968$,

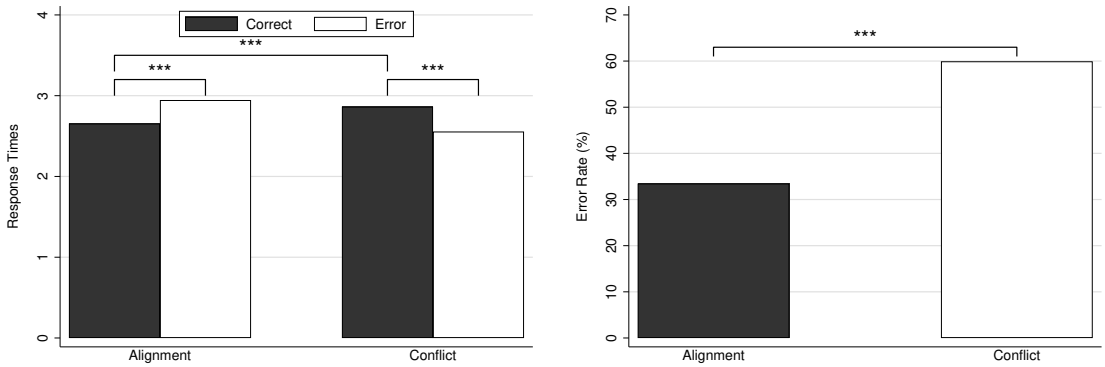

Figure A.15: Analyses of the judgment task by Glöckner and Bröder (2014). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

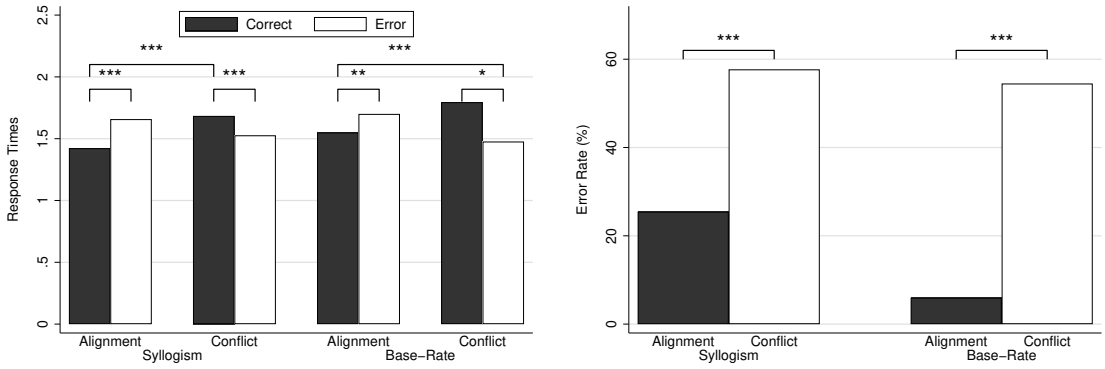

Figure A.16: Analyses of the decision tasks (syllogisms and base-rate questions) of Raoelison et al. (2020). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

$p < 0.001$, $r = 0.737$; D2, 40.86% vs. 30.55%; $N = 185$, $z = 6.261$, $p < 0.001$, $r = 0.460$; Figure A.14, right panel). Errors are faster than correct answers in conflict (T1, 1.867 vs. 1.934 s; $N = 183$, $z = -3.612$, $p < 0.001$, $r = 0.267$) but slower in alignment (T2, 1.941 vs. 1.752 s; $N = 183$, $z = 8.632$, $p < 0.001$, $r = 0.638$).

Dataset 15: Recognition Heuristic In Glöckner and Bröder (2014), sixty-one participants completed 120 trials, 40 each in alignment, conflict, or neutral cases. Figure A.15 shows that our predictions are again fully supported. Correct answers are slower are error rates are larger in conflict than in alignment (D1, 3.178 vs. 2.914 s; $N = 61$, $z = 5.290$, $p < 0.001$, $r = 0.677$; D2, 59.94% vs. 33.46%; $N = 61$, $z = 6.669$, $p < 0.001$, $r = 0.854$; Figure A.13, right panel). Errors are faster than correct answers in conflict (T1, 2.829 vs. 3.178 s; $N = 61$, $z = -4.500$, $p < 0.001$, $r = 0.576$) and slower in alignment (T2, 3.181 vs. 2.914 s; $N = 61$, $z = 3.379$, $p < 0.001$, $r = 0.433$).

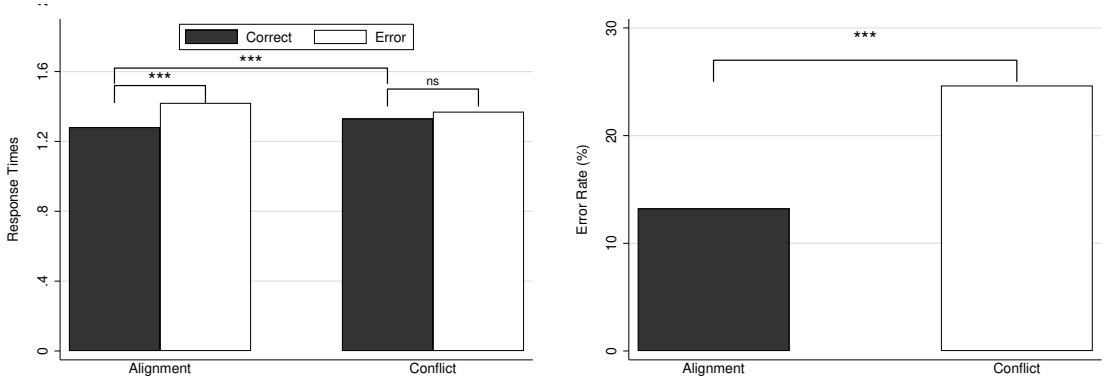

Figure A.17: Analyses of the value-based learning task by Fontanesi et al. (2019). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

B.5 Decision Making

Dataset 16: Syllogisms and Base Rates The dataset in Raoelison et al. (2020) comprises 260 subjects in two experiments ($N = 100$ and 160, respectively). We again report our analysis pooling both experiments, but results are qualitatively unchanged when considering them separately. Our analysis refers to the initial answers collected in the experiments, provided they were given within the specified 3-second interval.

Our predictions hold both for syllogisms and base-rate questions, as Figure A.16 illustrates. For Syllogisms, correct answers are slower and error rates are larger in conflict than in alignment (D1, 1.684 vs. 1.423 s; $N = 203$, $z = 7.752$, $p < 0.001$, $r = 0.544$; D2, 57.69% vs. 25.45%; $N = 259$, $z = 11.358$, $p < 0.001$, $r = 0.706$; Figure A.16, right panel). Errors are faster than correct answers in conflict (T1, 1.527 vs. 1.684 s; $N = 177$, $z = -5.167$, $p < 0.001$, $r = 0.388$) and slower in alignment (T2, 1.658 vs. 1.423 s; $N = 168$, $z = 6.232$, $p < 0.001$, $r = 0.481$). For Base Rate judgements, again correct answers are slower and error rates are larger in conflict than in alignment (D1, 1.795 vs. 1.550 s; $N = 162$, $z = 3.842$, $p < 0.001$, $r = 0.302$; D2, 54.59% vs. 5.98%; $N = 258$, $z = 11.692$, $p < 0.001$, $r = 0.728$; Figure A.16, right panel). Also, errors are faster than correct answers in conflict (T1, 1.477 vs. 1.795 s; $N = 85$, $z = -1.939$, $p = 0.052$, $r = 0.210$) and slower in alignment (T2, 1.700 vs. 1.550 s; $N = 41$, $z = 2.158$, $p = 0.030$, $r = 0.337$).

Dataset 17: Reinforcement Learning In Fontanesi et al. (2019), twenty-seven participants completed 240 trials. Figure A.17 shows that our predictions find support in this context, with all predictions except T1 confirmed. Specifically, correct answers are slower and error rates are larger in conflict than in alignment (D1, 1.330 vs. 1.281 s; $N = 27$, $z = 3.195$, $p < 0.001$, $r = 0.615$; D2, 24.65 vs. 13.24%; $N = 27$, $z = 3.748$, $p < 0.001$, $r = 0.721$; Figure A.17, right panel). In conflict, there are no significant differences in the response times of errors and correct answers (T1, 1.370 vs. 1.330 s; $N = 27$, $z = -0.817$, $p = 0.427$, $r = 0.157$). In alignment, errors are slower than correct answers as predicted (T2, 1.420 vs. 1.281 s; $N = 27$, $z = 3.099$, $p = 0.001$, $r = 0.596$).

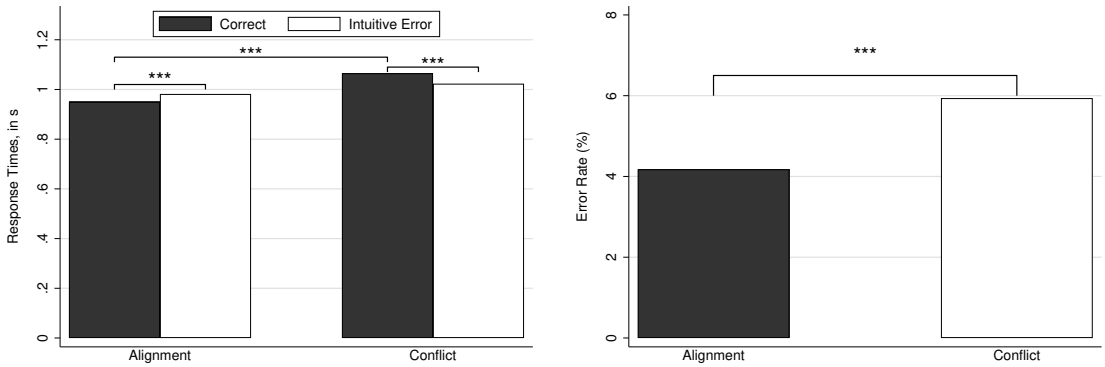

Figure A.18: Analysis of the dataset from Steyvers et al. (2019). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

B.6 Non-Binary Choice

Dataset 18: Task Switching Across the Lifespan We focus on the largest dataset of Steyvers et al. (2019), which includes 1,000 users who played for up to 60 sessions each, for a total of 46,470 sessions and 2,881,161 trials (other datasets are restricted to most-active users or older adults, both with at least a thousand sessions). The 1,000 users in the dataset were selected so that five age groups were approximately equally represented (21-30, 31-40, 41-50, 51-60, 61-70, and 71-80 years). Figure A.18 shows that our predictions, as given in Theorems 5–8, find full support in this context. Correct answers are slower in conflict than in alignment (D1; 1.065 s vs. 0.950 s, $N = 1000$, $z = 27.391$, $p < 0.001$, $r = 0.866$). Error rates were larger in conflict than in alignment (D2; 5.94% vs. 4.18%; $N = 1000$, $z = 22.391$, $p < 0.001$, $r = 0.708$; Figure A.18, right panel). The new prediction (D2') also holds: for *all* 1,000 subjects, the proportion of correct responses in alignment (average 95.82%) exceeds the proportion of intuitive errors in conflict (5.94%). In conflict, intuitive errors are faster than correct answers (T1; 1.023 s vs. 1.065 s; $N = 986$, $z = -14.732$, $p < 0.001$, $r = 0.469$), and in case of alignment errors are slower than correct responses (T2; 0.982 s vs. 0.950 s; $N = 999$, $z = 10.677$, $p < 0.001$, $r = 0.338$), although the differences are admittedly small.

Dataset 19: Sequential Conflict Modulation Thirty-nine and forty-eight subjects participated in Experiments 1 and 2 of Dignath et al. (2019), respectively. Each participant worked through 1,152 trials in 24 blocks of 48 trials each. For brevity, we report the analysis pooling both experiments ($N = 87$). However, results are unchanged if we look at them separately. The results are also unchanged if we examine the data separately depending on the type of stimuli representation.

Figure A.19 shows that our predictions also find full support in this context. Correct answers are slower and error rates are larger in conflict than in alignment (D1, 0.712 s vs. 0.583 s, $N = 87$, $z = 2.925$, $p = 0.003$, $r = 0.314$; D2, 10.79% vs. 8.12%, $N = 87$, $z = 5.982$, $p < 0.001$, $r = 0.641$; Figure A.19, right panel). Regarding (D2'), for all 87 subjects the proportion of correct responses in alignment (average 91.88%) exceeds the proportion of intuitive errors in conflict (10.79%). Last, intuitive errors are faster than correct answers in conflict (T1, 0.695 s vs. 0.712 s; $N = 87$, $z = -2.662$, $p = 0.007$,

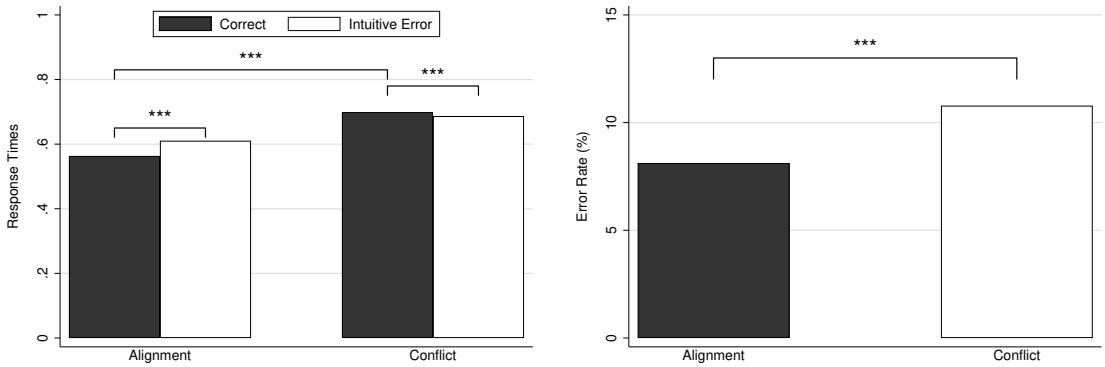

Figure A.19: Analysis of the experiments in Dignath et al. (2019). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

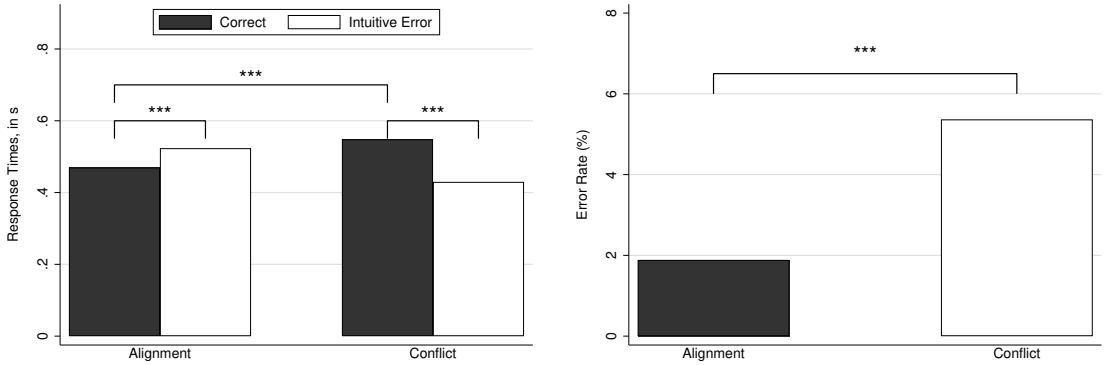

Figure A.20: Analyses of the congruency data from the two experiments of Schmidt and Weissman (2014). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Under conflict, only intuitive errors are considered. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

$r = 0.285$), but errors are slower than correct answers in alignment (T2, 0.634 s vs. 0.583 s; $N = 87$, $z = -7.077$, $p < 0.001$, $r = 0.759$).

Dataset 20: Prime-Probe Congruency Effects Each experimental dataset in Schmidt and Weissman (2014) comprises 16 subjects, each of which went through 768 trials. We again report our analysis pooling both experiments, but results are qualitatively unchanged when we look at the experiments separately. All our predictions hold in this setting, as illustrated in Figure A.20. Correct answers are slower and error rates are higher in conflict than in alignment (D1, 0.548 s vs. 0.470 s, $N = 32$, $z = 4.937$, $p < 0.001$, $r = 0.873$; D2, 5.37% vs. 1.89%; $N = 32$, $z = 4.479$, $p < 0.001$, $r = 0.792$; Figure A.20, right panel). For (D2'), for all 32 subjects the proportion of correct responses in alignment (average 98.11%) exceeds the proportion of intuitive errors in conflict (5.37%). In conflict, intuitive errors are faster than correct answers (T1, 0.430 s vs. 0.548 s; $N = 31$, $z = -4.625$, $p < 0.001$, $r = 0.831$). In alignment, errors are slower than correct answers (T2, 0.523 s vs. 0.470 s; $N = 30$, $z = 2.972$, $p = 0.002$, $r = 0.543$).

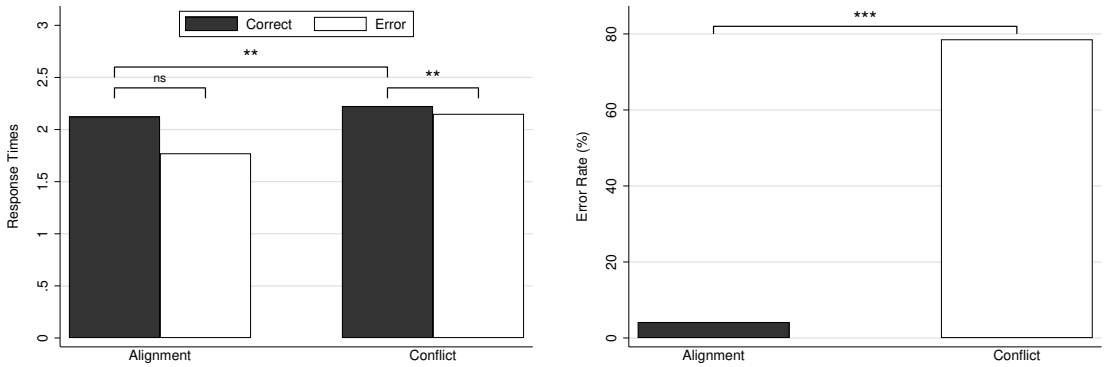

Figure A.21: Analyses of the Bat-and-Ball questions of Raoelison et al. (2020). Left: Average response times (in seconds) for errors and correct answers conditional on alignment vs. conflict. Under conflict, only intuitive errors are considered. Right: Error rates in alignment vs. conflict. WRS tests: *** $p < .01$.

Dataset 21: Cognitive Reflection As in the analysis of Paradigm 16, we pool the two experiments of Raoelison et al. (2020) for a total of 260 subjects. Again, our predictions hold, as illustrated in Figure A.21. Correct answers are slower and error rates are higher in conflict than in alignment (D1, 2.225 s vs. 1.772 s, $N = 59$, $z = 2.083$, $p = 0.037$, $r = 0.271$; D2, 79.79% vs. 4.28%; $N = 224$, $z = 12.940$, $p < 0.001$, $r = 0.865$; Figure A.16, right panel). Unfortunately, Raoelison et al. (2020) did not record the actual responses, but merely whether they were correct or not, and hence we cannot distinguish intuitive errors and other errors. Thus, for predictions (D2') and (T1), we conduct the tests as if all errors had been intuitive. Although this is probably not the case, since many errors in the CRT are intuitive, it might be a reasonable approximation. The proportion of correct responses in alignment is indeed larger than the proportion of errors in conflict (D2', 95.72% vs. 79.79%; $N = 260$, $z = 5.876$, $p < 0.001$, $r = 0.364$). In conflict, errors are faster than correct answers (T1, 2.150 s vs. 2.225 s; $N = 28$, $z = -2.186$, $p = 0.028$, $r = 0.413$). In alignment, the sample size is drastically reduced since the vast majority of participants made no mistakes. In this case, errors are slower than correct answers, but the difference is not significant (T2, 2.125 s vs. 1.772 s; $N = 24$, $z = 1.086$, $p = 0.286$, $r = 0.222$).

C Analysis of Neutral Trials in the Individual Datasets

In this section we report the analyses of neutral trials in the paradigms which include this particular type of situation. By definition, neutral trials are those where the cue triggering the intuitive/impulsive process is absent, and hence only the deliberative process is active. Predictions D1, D2 and D2' explicitly compare conflict trials to alignment trials, while predictions T1 and T2 refer only to conflict or alignment trials, respectively. Hence, none of those predictions apply to neutral trials. Propositions 1 and 2 deliver predictions for error rates in neutral trials compared to conflict and alignment trials, which we test below. As discussed in the main text, response times in these trials can also be used (in a more exploratory sense) to study the properties of the involved deliberative processes.

Error rates in neutral trials In agreement with prediction N1, error rates where almost always larger in neutral trials compared to alignment trials. We observe this effect in twelve of the fifteen studies containing neutral trials. This occurs in Dataset 5 (No cue: $N = 123, z = 2.784, p = 0.005$, $r = 0.251$; Alerting cue: $N = 123, z = 2.431, p = 0.015$, $r = 0.219$), Dataset 6 ($N = 12, z = 2.903, p = 0.002$, $r = 0.839$), Dataset 8 (probability blocks: $N = 32$, $z = 4.151$, $p < 0.001$, $r = 0.733$; reward blocks $N = 32$, $z = 2.431$, $p = 0.014$, $r = 0.430$), Dataset 9 (Experiment 1: $N = 30$, WSR, $z = 2.180$, $p = 0.029$, $r = 0.400$; Experiment 3: $N = 21$, WSR, $z = 2.485$, $p = 0.011$, $r = 0.543$), Dataset 15 ($N = 61$, $z = 6.777$, $p < 0.001$, $r = 0.868$), Dataset 16 (Base Rate: $N = 160$, $z = 4.796$, $p < 0.001$, $r = 0.376$; Syllogism: $N = 159$, $z = 4.879$, $p < 0.001$, $r = 0.384$), Dataset 17 ($N = 27$, $z = 3.099$, $p = 0.001$, $r = 0.600$), and Dataset 21 ($N = 58$, $z = 3.456$, $p < 0.001$, $r = 0.454$). We observe no significant differences in two datasets. This occurs in Dataset 5 (Orienting cue: $N = 123, z = 1.333, p = 0.184$, $r = 0.120$) and Dataset's 9 Experiment 2 ($N = 23$, WSR, $z = 0.578$, $p = 0.580$, $r = 0.120$). The opposite effect, i.e. significantly more errors in alignment than neutral trials, is only observed in Dataset 18 ($N = 999$, $z = 22.166$, $p < 0.001$, $r = 0.700$). We remark again that, as for the comparison of response times of correct choices versus alignment situations, in Dataset 18 neutral situations corresponded to the first trials of the different sessions, hence they might not be completely comparable to the other trials.

In agreement with predictions N2 and N2', error rates where generally lower in neutral trials compared to conflict trials. Specifically, we observe significantly more errors in conflict than neutral situations in eleven of the fifteen studies containing neutral trials (Figure B.2). This occurs in Dataset 5 (No cue: conflict 22.62% vs. neutral 14.72%; $N = 123, z = 2.784, p = 0.005$, $r = 0.251$; Alerting cue: conflict 22.46% vs. neutral 14.35%; $N = 123, z = 2.431, p = 0.015$, $r = 0.219$; Orienting cue: conflict 22.85% vs. neutral 14.30%; $N = 123, z = 1.333, p = 0.184$, $r = 0.120$), Dataset's 8 probability blocks (conflict 23.08% vs. 14.73% neutral; $N = 32$, $z = -3.927$, $p < 0.001$, $r = 0.694$), experiment 1 of Dataset 9 (conflict 33.19% vs. neutral 22.68%; $N = 30$, WSR, $z = -3.703$, $p = 0.001$, $r = 0.676$), Paradigm 15 (conflict 59.94% vs. neutral 56.31%; $N = 61$, $z = 2.816$, $p = 0.004$, $r = 0.361$), Dataset 17 (conflict 24.65% vs. neutral 18.55%; $N = 27$, $z = -2.763$, $p = 0.004$, $r = 0.532$), and Dataset 18 (conflict 5.94% vs. neutral 1.21%; $N = 999$, $z = -24.403$, $p < 0.001$, $r = 0.772$). In the last two datasets, the tests correspond to intuitive errors in multi-alternative choice as described in (N2').

The opposite effect is never observed. For the remaining four of the fifteen studies with neutral trials, we observe no significant differences. This occurs in Dataset 6 (conflict 38.36% vs. 37.65 neutral; $N = 12, z = 0.628, p = 0.569$, $r = 1.182$), Dataset's 8 reward blocks (conflict 14.87% vs. 14.73% neutral; $N = 32$, $z = 0.337$, $p = 0.746$, $r = 0.060$), and two experiments of Dataset 9 (Experiment 2: conflict 21.94% vs. neutral 21.55%, $N = 23$, WSR, $z = -0.091$, $p = 0.941$, $r = 0.020$; Experiment 3: conflict 28.30% vs. neutral 28.99%, $N = 21$, WSR, $z = 0.087$, $p = 0.946$, $r = 0.020$).

Comparing response times of correct responses between neutral and other trials The second comparison of interest concerns the speed of correct responses in neutral trials compared to conflict or alignment trials. As discussed in the main text, if the time needed for conflict detection and resolution is enough to offset the differences in response times among the processes, we would obtain the prediction that correct responses for neutral trials are faster than those in conflict, but slower than those in alignment.

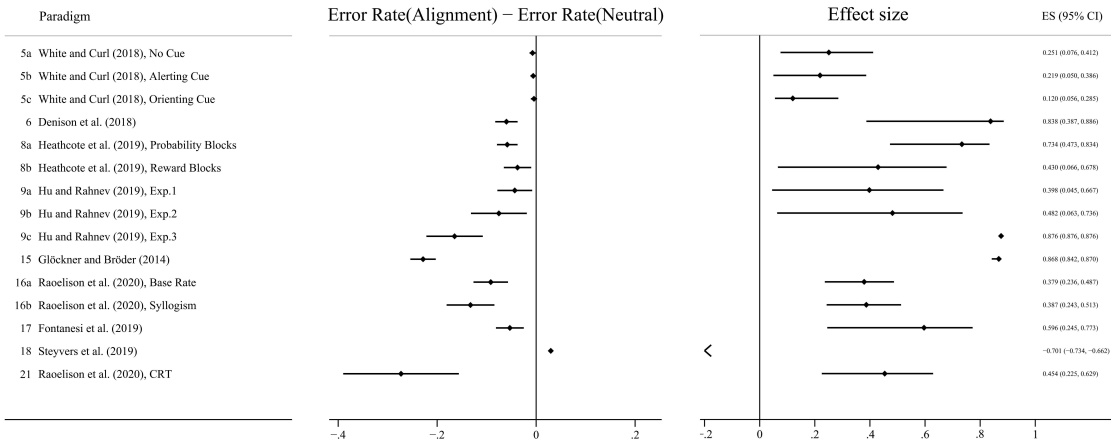

Figure B.1: Comparison of error rates between neutral trials and alignment trials, for all datasets containing neutral trials. Effect sizes are for the non-parametric tests. For studies 18 and 21 error rates are calculated using intuitive errors.

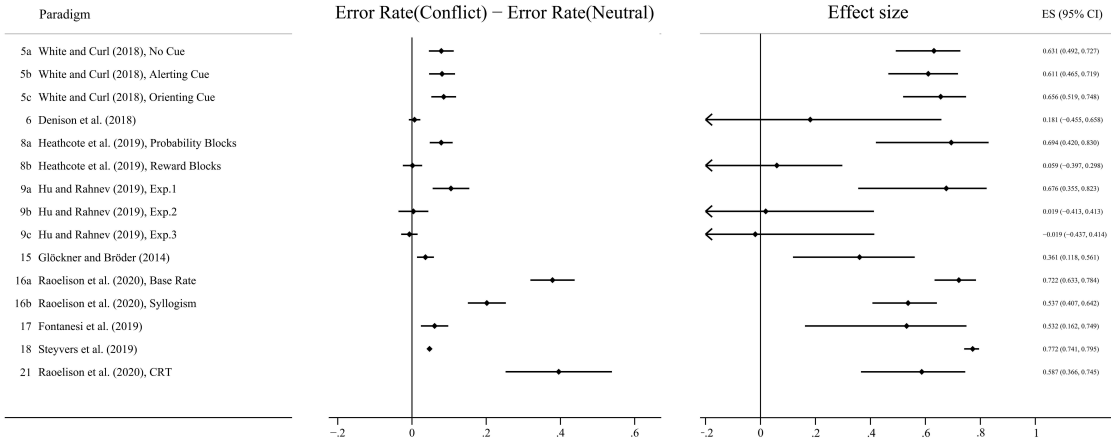

Figure B.2: Comparison of error rates between neutral trials and conflict trials, for all datasets containing neutral trials. Effect sizes are for the non-parametric tests. For studies 18 and 21 error rates are calculated using intuitive errors.

This is generally confirmed by the data. First, we observe significantly slower correct choices in conflict than neutral in nine comparisons (Figure B.3). This occurs in Dataset 5 (no-cue condition: $N = 119, z = -8.605, p < 0.001$, $r = 7.789$; alerting-cue condition: $N = 117, z = -8.231, p < 0.001$, $r = 0.761$; orienting-cue condition: $N = 118, z = -8.928, p < 0.001$, $r = 0.822$), Dataset 8's probability blocks ($N = 32$, $z = -2.655$, $p = 0.007$, $r = 0.469$), Dataset 9 (Experiment 1: $N = 29$, WSR, $z = -2.606$, $p = 0.008$, $r = 0.483$; Experiment 2: $N = 23$, WSR, $z = -2.555$, $p = 0.009$, $r = 0.532$; Experiment 3: $N = 21$, WSR, $z = -2.555$, $p = 0.009$, $r = 0.526$), Dataset 16's Base Rates ($N = 107$, $z = -3.465$, $p < 0.001$, $r = 0.335$), and Dataset 18 ($N = 999$, $z = -25.747$, $p < 0.001$, $r = 0.812$). The opposite effect, i.e., significantly faster correct choices in conflict than neutral situations, only occurs in Dataset 17 ($N = 27$, $z = -3.724$, $p < 0.001$, $r = 0.716$). In the remaining five of the fifteen studies containing neutral trials, we observe no significant differences. This occurs in Dataset 6 ($N = 12, z = -1.255, p = 0.233$, $r = 0.884$), the reward condition of Dataset 8 ($N = 32$, $z = 1.103$, $p = 0.278$, $r = 0.195$), Dataset 15 ($N = 61$, $z = 0.758$, $p = 0.453$, $r = 0.097$), Dataset 16's Syllogism ($N = 121$,

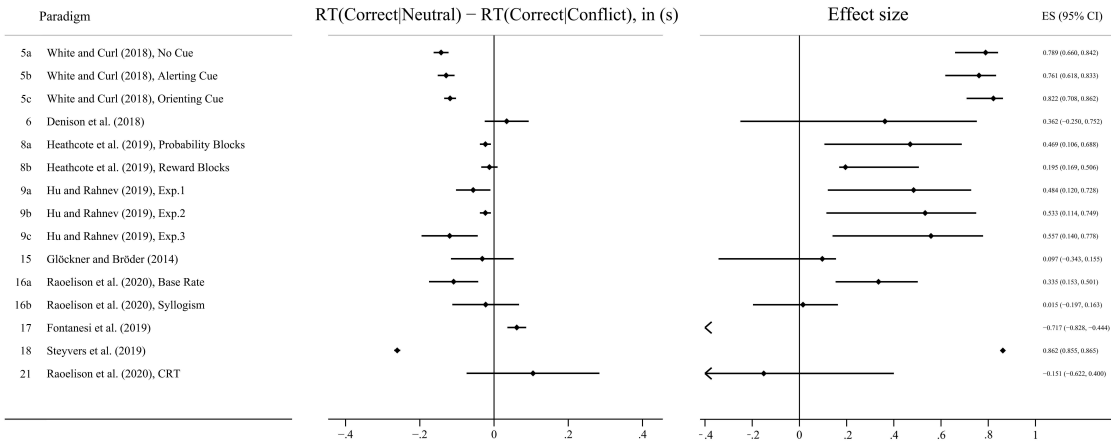

Figure B.3: Comparison of response times of correct choices between neutral and conflict trials, for all datasets containing neutral trials. Effect sizes refer to the non-parametric tests.

$z = -0.167$, $p = 0.869$, $r = 0.015$) and Dataset 21's CRT ($N = 14$, $z = 0.565$, $p = 0.604$, $r = 0.151$).

Second, we also observe significantly slower correct choices in neutral than alignment in seven comparisons (Figure B.4). This occurs in two of Dataset 5's conditions (Alerting cue: $N = 116, z = 4.105, p < 0.001$, $r = 0.381$; 0.514; $N = 118, z = 2.859, p = 0.004$, $r = 0.263$), Dataset 6 ($N = 12, z = 3.059, p = 0.005$, $r = 0.884$), Dataset 8's probability blocks ($N = 32$, $z = 4.488$, $p < 0.001$, $r = 0.766$), Dataset 9' Experiment 3 ($N = 21$, WSR, $z = 2.485$, $p = 0.011$, $r = 0.543$), Dataset 15 ($N = 61$, $z = -4.040$, $p < 0.001$, $r = 0.517$), and Dataset 17 ($N = 27$, $z = 4.445$, $p < 0.001$, $r = 0.854$). The opposite effect, i.e., significantly slower correct choices in alignment than neutral situations, only occurs in Dataset 18 ($N = 999$, $z = 27.244$, $p < 0.001$, $r = 0.862$) where neutral situations corresponded only to the first trials of the different sessions. In the remaining seven of the fifteen studies containing neutral trials, we observe no significant differences. This occurs in Dataset 5's no cue condition ($N = 120, z = -0.686, p = 0.494$, $r = 0.063$), Dataset 8's reward blocks ($N = 32$, $z = -0.916$, $p = 0.369$, $r = 0.162$), two of the experiments in Dataset 9 (Experiment 1: $N = 30$, WSR, $z = -0.874$, $p = 0.393$, $r = 0.159$; Experiment 2: $N = 23$, WSR, $z = -0.578$, $p = 0.580$, $r = 0.120$), Dataset 16 (Base Rate: $N = 20$, $z = -0.299$, $p = 0.784$, $r = 0.067$; Syllogism $N = 102$, $z = -0.018$, $p = 0.986$, $r = 0.002$), and Dataset 21 ($N = 8$, $z = -1.400$, $p = 0.195$, $r = 0.495$).

Relative speed of errors in neutral trials Last, we compare the speed of errors and correct responses in neutral trials for datasets which include those (Figure 5 in the main text). We observe that either the response times of errors and correct responses are not significantly different, or, if a difference exists, it goes in the direction of errors being slower (all tests are WSR tests).

There are no significant differences between the response times of errors and correct responses in the neutral trials in Dataset 5 (no-cue condition: errors 0.591 s vs. correct answers 0.577 s, $N = 73$, $z = 0.542$, $p = 0.592$, $r = 0.063$; alerting-cue condition: 0.585 vs. 0.545 s, $N = 53, z = 0.164, p = 0.874$, $r = 0.023$; orienting-cue condition: 0.520 vs. 0.514 s, $N = 58, z = 1.428, p = 0.155$, $r = 0.189$), Dataset 8 (errors 0.678 s vs. correct answers 0.664 s, $N = 32$, $z = -1.496$, $p = 0.139$, $r = 0.264$), Dataset 16 (syllogisms: errors 1.662 s vs. correct answers 1.646 s, $N = 118$, $z = 0.632$, $p = 0.529$, $r = 0.058$;

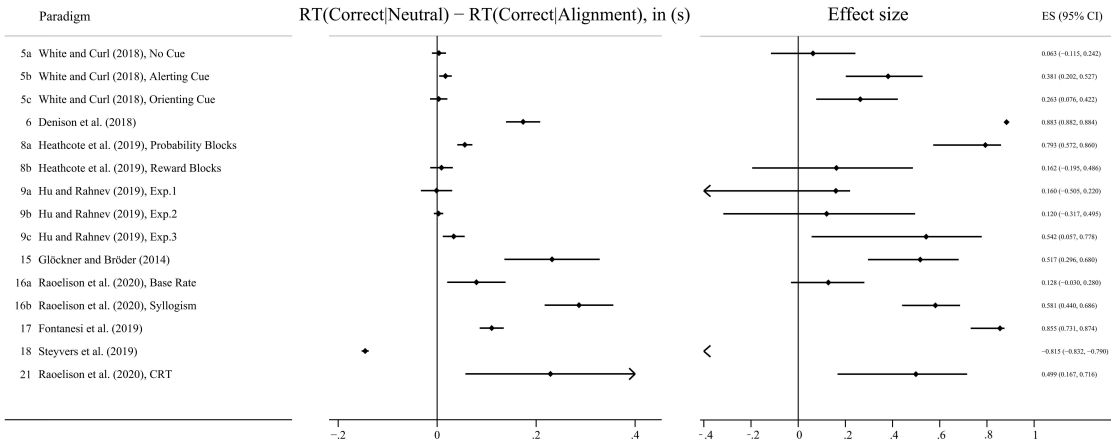

Figure B.4: Comparison of response times of correct choices between neutral and alignment, for all datasets containing neutral trials. Effect sizes refer to the non-parametric tests.

base rates: 1.622 vs. 1.570 s, $N = 51$, $z = -0.553$, $p = 0.586$, $r = 0.077$), and Dataset 21 (2.268 vs. 1.937 s, $N = 8$, $z = -0.280$, $p = 0.843$, $r = 0.100$).

We observe significantly slower errors (compared to correct responses) in the neutral trials of four other datasets, encompassing six different comparisons. This occurs in Dataset 6 (errors 0.994 s vs. correct answers 0.932 s, $N = 12, z = 1.804, p = 0.077$, $r = 0.810$), Dataset 9 (Experiment 1: errors 0.577 s vs. correct answers 0.501 s, $N = 29$, $z = 3.362$, $p < 0.001$, $r = 0.624$; Experiment 2: 0.755 vs. 0.727 s, $N = 23$, $z = 3.984$, $p < 0.001$, $r = 0.831$; Experiment 3: 1.172 vs. 1.033 s, $N = 21$, WSR, $z = 3.493$, $p = 0.001$, $r = 0.762$), Dataset 17 (errors 1.489 s vs. correct answers 1.391 s, $N = 27$, $z = 3.147$, $p = 0.001$, $r = 0.606$), and Dataset 18 (0.985 vs. 0.805 s, $N = 217$, $z = 3.524$, $p < 0.001$, $r = 0.239$).

We only observe significantly faster errors in the most-populated-city question of Dataset 15 (errors 2.920 s vs. correct answers 3.146 s, $N = 61$, $z = 3.394$, $p < 0.001$, $r = 0.435$). However, even in this case the effect only holds if both cities were categorized as unknown (2.773 s vs. 3.269 s; $N = 61$, $z = 4.435$, $p < 0.001$, $r = 0.568$). If both are categorized as known, the difference is not significant (3.120 vs. 3.115 s, $N = 61$, $z = 0.018$, $p = 0.989$, $r = 0.002$).