Economics Working Paper Series

2025/009

Noise and Bias: The Cognitive Roots of Economic Errors

Carlos Alós-Ferrer, Johannes Buckenmaier and Michele Garagnani

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

LUMS home page: http://www.lancaster.ac.uk/lums/

Noise and Bias: The Cognitive Roots of Economic Errors*

Carlos Alós-Ferrer† Johannes Buckenmaier‡ Michele Garagnani§

This version: April 17, 2025

Abstract

Economic decisions are noisy due to errors and cognitive imprecision. Often, they are also systematically biased by heuristics or behavioral rules of thumb, creating behavioral anomalies which challenge established economic theories. The interaction of noise and bias, however, has been mostly neglected, and recent work suggests that received behavioral anomalies might be just due to regularities in the noise. This contribution formalizes the idea that decision makers might follow a mixture of rules of behavior combining cognitively-imprecise value maximization and computationally simpler shortcuts. The model delivers new testable predictions which we validate in two experiments, focusing on biases in probability judgments and the certainty effect in lottery choice, respectively. Our findings suggest that neither cognitive imprecision nor multiplicity of behavioral rules suffice to explain received patterns in economic decision making. However, jointly modeling (cognitive) noise in value maximization and biases arising from simpler, cognitive shortcuts delivers a unified framework which can parsimoniously explain deviations from normative prescriptions across domains.

JEL Classification: D01 · D81 · D87 · D91

Keywords: Cognitive Imprecision · Strength of Preference · Noise · Decision Biases · Belief Updating · Certainty Heuristic

*We are grateful to Michael Birnbaum, Ido Erev, Ernst Fehr, Antonio Filippin, Graham Loomes, Fabio Maccheroni, Paolo Natenzon, Daniel Read, David Redish, and seminar participants at the Conference on Decision Sciences in Konstanz, the 1st Matterhorn Symposium, the RISLab Workshop on Cognitive Economics, the IAREP conference in Dundee, the ESA Asia-Pacific conference in Osaka, Universidad de Alicante, Warwick Business School, Lancaster University, University of Milano, and Universidad Autónoma de Madrid. Financial support from the German Research Foundation (DFG) through project Al-1169/4, part of the Research Unit "Psychoeconomics" (FOR 1882), is gratefully acknowledged.

†Corresponding author: c.alosferrer@lancaster.ac.uk. Lancaster University Management School. Lancaster, LA1 4YX, United Kingdom.

‡Independent Researcher.

§Centre for Brain, Mind & Markets Laboratory, University of Melbourne, Melbourne, Australia.

1 Introduction

Understanding inconsistencies in human decisions is of fundamental importance for economics and other disciplines. Classical economic theory assumes deterministic and error-free choices, but evidence shows that economic choice is stochastic and error-prone. Human decision makers very often make different choices when confronted with the same set of options repeatedly, but choices are not purely random. On the contrary, individual choice frequencies often display systematic patterns (e.g., Davidson and Marschak, 1959; Tversky, 1969; Camerer, 1989; Hey and Orme, 1994; Agranov and Ortoleva, 2017). This evidence has motivated probabilistic approaches going back to Fechner (1860), Thurstone (1927), Debreu (1958), Luce (1959), and Block and Marschak (1960), and is currently giving new impulse to the microeconomics literature on stochastic choice, which replaces deterministic models of choice with probabilistic systems and Random Utility Models (RUMs; McFadden, 2001; Anderson et al., 1992). The latter are widely used for preference estimation in economics, psychology, marketing (e.g., Baltas and Doyle, 2001), political science (e.g., Nownes, 1992), and other fields.

These models, however, might be insufficient to understand errors in economic decisions. This is because sometimes decisions are also biased in the sense that they exhibit behavioral patterns which are systematic and predictable, but inconsistent with standard economic preferences. The interaction between these two fundamentally different sources of errors in economic decisions has not been taken into account so far. This work proposes and tests a new, simple model of economic choice taking into account that decision errors are ubiquitous, but that there are different sources of errors that interact in complex ways, which we refer to as "noise" and "bias." To integrate these two sources of errors, the model relies on insights from psychology and decision neuroscience ("neuroeconomics"; Camerer et al., 2005; Alós-Ferrer, 2018a), as well as on recent developments in the nascent field of cognitive economics (Caplin, 2025).

The first source of errors can be thought of as (cognitive) *noise*. A recent literature in economics argues that noise in economic decisions arises from *cognitive imprecision*. Essentially, some anomalies in economic decision-making might occur because decision values are perceived or computed in an imprecise way.[1] One possible reason is perceptual: decision errors might be due to noise because decision stimuli are imperfectly perceived and the human brain functions as a decoder of an efficiently-coded but noisy signal (Polanía et al., 2019; Khaw et al., 2021; Frydman and Jin, 2022; Vieider, 2024). Another reason might be computational: the complexity of decisions or of the available options might result in noisier value computations due to cognitive limitations and hence generate errors (Enke and Graeber, 2023; Oprea, 2024; Enke et al., 2025). This literature is aligned with earlier, extensive evidence from perceptual tasks in psychophysics,

[1] More generally, the view in decision neuroscience is that noise arises because the computation of decision values in the human brain, linked to the dopaminergic system, is subject to neuronal noise that creates choice inconsistencies at the individual level, even for simple tasks with objective answers (see, e.g., Shadlen and Kiani, 2013; Shadlen and Shohamy, 2016).

which has identified psychometric and chronometric regularities also known as "strength of preference" phenomena: errors are more frequent and response times are longer when the decision values are closer in the relevant scale (Cattell, 1893, 1902; Dashiell, 1937; Laming, 1985; Klein, 2001; Wichmann and Hill, 2001). In a ground-breaking, classical work, Mosteller and Nogee (1951) showed that these effects also appear in economic decisions and that, for those, preferences might sometimes deliver the relevant scale. Specifically, strength-of-preference phenomena occur in decisions under risk with estimated utilities. Recent work has further shown that these phenomena are relevant for (estimated, subjective) preferences in economic decisions (Chabris et al., 2009; Alós-Ferrer and Garagnani, 2022a,b), and that these regularities can be used to improve both preference revelation (Alós-Ferrer et al., 2021) and prediction out of sample (Alós-Ferrer and Garagnani, 2024b).

The second source of errors is the systematic *bias* that can arise from the multiplicity of decision processes in the human brain. A large literature in psychology has developed *dual-process theories*, which postulate that decisions result from the interaction of multiple processes which differ along an automaticity continuum (see, among others, Kahneman, 2003, 2011; Evans, 2008; Weber and Johnson, 2009; Alós-Ferrer and Strack, 2014). Some processes are more *automatic* (or intuitive): fast, unconscious, and requiring few cognitive resources. They capture impulsive reactions and behavior along the lines of stimulus-response schemes. Other processes are more *deliberative* (or controlled): slow, consuming cognitive resources, and reflected upon (partly) consciously. Relatively automatic processes might include many heuristics studied in behavioral economics (e.g., the certainty heuristic, representativeness, etc.; Kahneman and Tversky, 1972, 1979; Grether, 1980), and many boundedly-rational *behavioral rules* studied in game theory (e.g., imitation, Alós-Ferrer and Weidenholzer, 2008, 2014; myopic best reply, Kandori and Rob, 1995; aspiration levels, Oechssler, 2002). A number of dual-process (or dual-self) models and applications have also explored this idea in economics, following up on the analogy with the distinction between full and bounded rationality (Thaler and Shefrin, 1981; Benhabib and Bisin, 2005; Fudenberg and Levine, 2006, 2012; Gennaioli and Shleifer, 2010; Heller et al., 2017; Cerigioni, 2021). While most dual-process models are essentially "as if," previous work has shown that process data (response times, eye tracking, and electroencephalography) can be used as diagnostic tools to reveal process multiplicity (e.g., Achtziger and Alós-Ferrer, 2014; Achtziger et al., 2014, 2015; Alós-Ferrer et al., 2021), although it has done so without taking into account possible interactions with strength of preference effects. Further, empirically-estimated Hidden Markov Models suggest that decision makers switch dynamically between alternative decision processes even for simple economic decisions (Alós-Ferrer and Garagnani, 2023).

If human decisions are affected by both noise and bias, ignoring one or the other might create severe confounds. First, as shown by Krajbich et al. (2015) and Evans et al. (2015), *apparent* "fast and slow" effects might be observed even in the absence of multiple processes because some decisions are fast or slow by virtue of being easy or hard (large or small utility differences, respectively). Second, a recent literature has shown that some behavioral anomalies that are often explained through heuristics might simply arise from standard preference models and regularities in noisy choice. This includes the efficient coding literature mentioned above (Khaw et al., 2021; Frydman and Jin, 2022; Vieider, 2024), which links noise to cognitive imprecision, and recent work on inconsistencies generated by the subjective perception of complexity (Enke and Graeber, 2023; Oprea, 2024).

Third, strength of preference might be enough to explain some apparent anomalies which have been extensively accepted to reveal behavioral biases. For example, consider the *common ratio effect* (Kahneman and Tversky, 1979). When choosing between a sure but small amount and a larger but risky amount, most decision makers choose the safe alternative, but when all probabilities are scaled down by a common ratio, most people tend to choose the riskier option. This phenomenon has been widely replicated (Ruggeri et al., 2020; Blavatskyy et al., 2023), contradicts Expected Utility Theory, and is typically explained through a "certainty heuristic," which is a bias-based explanation. However, McGranaghan et al. (2024) have argued that standard RUMs can generate this effect simply due to strength-of-preference patterns in the noise, i.e. the fact that errors are more frequent when utility differences are smaller (see also Ballinger and Wilcox, 1997; Hey, 2005, among others). If participants prefer a sure but small amount to a larger but risky amount, there is still a (small) probability that they will choose the latter due to noise. But when the winning probabilities are scaled down, utility differences become much smaller, and the probability of an error becomes much larger. This makes reversals of the type observed in the common ratio effect more likely than the opposite ones. That is, it is unclear whether an effect due to regularities in the noise has been misinterpreted as evidence for a bias.

This work formalizes the idea that people might be following qualitatively different "rules of behavior" more or less at the same time. The intuition is that decision making can be well captured as a mixture of cognitively-imprecise (and often subjective) value maximization and computationally simpler shortcuts. The first part captures the view of decision makers acting as imperfect maximizers of some subjective value subject to cognitive constraints. That is, our model captures cognitive imprecision through well-established strength-of-preference effects.

The computationally-simpler cognitive shortcuts in the second part of the framework are akin to heuristic processes which can give rise to biases when they conflict with value maximization, and hence link to the extensive heuristics and biases literature (Kahneman, 2011). The idea is to formalize, under the same umbrella, different but conceptually similar departures from rational behavior which are computationally less involved and closer to simplified rules of behavior, ranging from be heuristics from the judgment and decision making literature (representativeness, conservatism, etc.) and the literature on decisions under risk (e.g., going for the sure outcome) and behavioral rules as often studied in game theory (myopic best reply, imitation, reinforcement, etc.). The literature has studied many such rules, but a formalization of how they interact with cognitive imprecision (strength of preference) is still missing and might bring crucial insights to economic analysis. We show that a general framework which jointly models noisy value maximization and simple rules of behavior lead to non-trivial predictions relative to choice frequency, choice inconsistency, and response times. In particular, the framework we propose provides a guideline for how to test for the presence of multiple behavioral rules contributing to choices, and show that distinguishing this case from behavior explained by a single rule behavior is far from straightforward.

Specifically, the model we propose considers *both* process multiplicity and strength of preference effects, as well as their interaction, and hence is able to identify effects of noise in the presence of bias and vice versa. The basic architecture assumes a (more) deliberative, value-maximizing process, which is affected by cognitive imprecision and hence exhibits psychometric (and chronometric) effects based on a cardinal (utility) scale. This process, however, interacts with a cognitively-simpler, heuristic process, whose prescriptions might be in conflict or in alignment with deliberation. In general, the latter process might be non-psychometric, e.g. if it is triggered by binary properties (win or lose; the presence of a riskless option; etc.), or it might exhibit a psychometric dependence on a *different* scale. For instance, an *imitative* behavioral rule might depend on a cardinal scale reflecting observed past performance.

The model predicts that strength of preference effects should be observed conditional on decisions which exhibit the same kind of interactions between processes. That is, psychometric effects (larger error frequency for smaller utility differences) should obtain when looking at all decisions where the deliberative and the heuristic process are in alignment (i.e., tend to make the same prescription), or when looking at all decisions where they are in conflict (i.e., tend to make the same prescriptions), as well as for decisions where the heuristic process is not triggered (which we call "neutral"). However, if the analyst were to neglect classifying decisions in this way, the effects of strength of preference might be difficult to observe. This is because, in addition to those effects, whether decisions are in conflict or alignment results in shifts in the frequency of errors, with alignment and conflict leading to less and more errors, respectively, for a given strength of preference.

In many cases of economic interest, heuristic processes are triggered by specific features of available alternatives, e.g. a certainty heuristic triggered by the presence of a risk-free lottery. In those cases, an appropriate renaming of alternatives allows a clear partition of decision problems, e.g. with all decisions in alignment (resp. in conflict) exhibiting positive (resp. negative) utility differences. Whenever this classification is possible, the model delivers a strong prediction which allows to detect process multiplicity even in the presence of strength-of-preference effects, hence providing empirical evidence for multiple processes while avoiding possible confounds. The prediction takes the form of a discontinuity at zero utility difference, with error rates jumping up as one moves from negative to positive differences. This discontinuity, which can be detected with well-established regression discontinuity estimation techniques (Calonico et al., 2014), is predicted to be absent for neutral decision problems, where the heuristic is not triggered.

To test these predictions, we conducted two different experiments. The first experiment focused on a well-known bias from the literature on belief updating and decision making: conservatism, i.e. the tendency to overweight prior probabilities when confronted with new information, or even to completely ignore the new information (Edwards, 1968; Grether, 1980; Fischhoff and Beyth-Marom, 1983; El-Gamal and Grether, 1995). In the terms of our model, the deliberative process is given by Bayesian updating, which prescribes to base behavior on the correctly updated posterior belief given the sample information. Strength of preference effects are operationalized on the basis of the cardinal payoff difference between the options. Hence the experiment belongs to the *objective domain*, in the sense that the assumed cardinal scale is objective and fixed before data is collected. The heuristic process models conservatism, and is active whenever there is an option with a higher prior probability (that is, it is inactive if the prior is 50-50). This heuristic ignores the new information and simply prescribes to choose the option with the higher prior probability.

The second experiment focused on the certainty effect in lottery choice, using lottery pairs corresponding both to common ratio anomalies (mentioned above; Kahneman and Tversky, 1979) and a common consequence construction (as in the Allais paradox; Allais, 1953). The certainty effect refers to the observation that many participants choose a sure option if available but a risky one in lottery pairs with transformed probabilities which should represent equivalent decisions in terms of Expected Utility Theory. As pointed out by McGranaghan et al. (2024), recent doubts have been raised on whether certainty effects uncover a bias or are merely due to strength-of-preference effects from the noise. This is important, because the certainty effect is one of the most prominent behavioral anomalies violating Expected Utility Theory, and was a crucial motivation in the development of Prospect Theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). Accordingly, this anomaly has received extensive attention in economics, and has sparked sparked the development of diverse models of non-expected utility theory beyond Prospect Theory, some of which were proposed with the specific aim to incorporate the certainty effect (e.g., Schmidt, 1998; Dillenberger, 2010; Cerreia-Vioglio et al., 2015).

Our second experiment shows how our model can accommodate the certainty effect. The deliberative process is given by expected utility maximization, with strength-of-preference operationalized on the basis of utility differences. Hence the experiment belongs to the *subjective domain*, in the sense that the assumed cardinal scale is subjective, subject-specific, and needs to be estimated. For this reason, the experiment included a block of lottery pairs designed and used for individual-level utility estimation. This allows an out-of-sample approach for the main analysis. That is, the predictions for the deliberative process are not based on any estimation using the data for which the predictions are made. The heuristic process models the certainty effect as a bias, and is active whenever there is a risk-free lottery. This is hence an example of an application where an appropriate renaming of alternatives results in a clear partition partition of decision problems, with all decisions in alignment (resp. in conflict) exhibiting positive (resp. negative) expected utility differences.

We confirm the predictions of our model in both experiments. Strength-of-preference effects can be observed once one classifies decision problems in alignment, conflict, or neutral, with larger error rates for smaller cardinal differences between the options (payoffs in the first experiment and expected utilities in the second). However, error rates are shifted when distinguishing decision problems according to whether they are in alignment or conflict, showing that the effects of noise and bias are observable separately and when controlling for each other. Further, in the second experiment, we clearly observe the predicted discontinuities which reveal process multiplicity while controlling for strength-of-preference effects.

The model also accounts for effects on response times. This is important, because the predictions of noise and bias are quite different and can easily result in confounds (Krajbich et al., 2015). In particular, a "fast and slow" intuition can often be misleading. This is because the response times of *processes* are not observable, and simply classifying decisions in fast and slow runs into a reverse inference fallacy (not all fast decisions are automatic; Evans et al., 2015; Myrseth and Wollbrant, 2016). The model presented here avoids those problems and makes testable predictions on response times which are diagnostic of process multiplicity. First, correct (utility-maximizing) responses are less frequent and slower under conflict than under alignment (generalizing the Stroop effect in psychology, Stroop, 1935). Second, under conflict errors are on average faster than correct responses, while under alignment the opposite is true. We find evidence for these effects in both experiments, hence confirming the presence of multiple decision processes.

We remark that closely-related response-time effects have previously been documented in a variety of settings including such processes as reinforcement (Achtziger and Alós-Ferrer, 2014; Spiliopoulos, 2018), brand recognition (Alós-Ferrer, 2018b), and imitation vs. best-reply (Alós-Ferrer and Ritschel, 2021) in a Cournot oligopoly. Alós-Ferrer and Garagnani (2024a) has recently reanalyzed data from 31 recent psychology studies, and shown the predictions to be robust across paradigms in cognitive control, attention, social cognition, memory, and decision making. The difference with the response-times predictions in the present work is that, in the model considered here, the predictions arise even when one controls for strength of preference effects, hence avoiding any possible confound between the effects of process multiplicity and strength-of-preference (chronometric) regularities (Krajbich et al., 2015).

Our results are also robust to some natural extensions. For instance, the basic model considers an exogenous probability of inhibition of heuristic processes in favor of more deliberative ones. In an extension, we show that the predictions survive if the probability of inhibition increases when strength of preference is larger, i.e. when opportunity costs increase (Benhabib and Bisin, 2005). This implies that biases resulting from heuristic processes should become more prominent when people are more cognitively uncertain about their decisions, a phenomenon which has been recently documented by Enke and Graeber (2023).

The remainder of the paper is structured as follows. Section 2 presents the basic model. Section 3 focuses on interaction of strength of preference (noise) and process multiplicity (bias). Subsection 3.1 derives the predictions on error rates arising from this interaction (Theorem 1). Subsection 3.2 presents our first experiment (on conservatism) and the results, which confirm the predictions of Theorem 1. Section 4 focuses the model's prediction of a discontinuity on error rates in the presence of multiple processes. The formal result (Theorem 2) is presented in Subsection 4.1. Subsection 4.2 presents our second experiment (on the certainty effect) and the results, which confirm the predictions of Theorems 1 and 2. Section 5 presents our analysis of response times. Section 5.1 derives the generalized Stroop effect (Theorem 3) and displays the evidence for this prediction in both experiments. Section 5.2 derives the prediction on the relative speed of errors (Theorem 4), and displays the evidence for the predicted effects in both experiments. Sections 5.3 and 5.4 report regression analyses on response times for the two experiments. Section 6 reports on three extensions and additional analyses. Section 7 concludes.

2 A Model of Noise and Bias

Decision makers face binary decision problems among alternatives in a set $\mathcal{A}$. Each decision problem is modeled as a pair of the form $\omega = \{A_\omega, B_\omega\}$, i.e., describes a binary decision among two options A_ω and B_ω. The labels A and B might have particular meanings for particular applications. For example, options labeled B might be riskier than options labeled A in a well-defined sense, or might refer to an urn presented on the left rather than on the right. The restriction to binary decisions is made mainly for tractability, since the experiments we focus on entail binary decisions. However, the model can be extended to decision problems with any finite number of alternatives.

Let the set of all decision problems faced by decision makers be denoted by Ω. A *decision process* is a map $b : \Omega_b \mapsto \mathcal{A}$ with $b(\omega) \in \omega$ for each $\omega \in \Omega_b$, where $\Omega_b \subseteq \Omega$. That is, a decision process singles out a particular alternative $b(\omega)$ for each decision problem ω in a particular subset Ω_b. The set Ω_b captures the subset of decision problems where a given process is triggered. The interpretation is that some decision processes might be triggered only by specific features of ω. For example, a certainty heuristic might be triggered by the presence of a riskless option. We say that a decision process b is *active* in decision problem ω if $\omega \in \Omega_b$.

We refer to $b(\omega)$ as the *prescription* or *favored response* of process b in decision problem ω. The interpretation is that the actual implementation of any process is noisy, and $b(\omega)$ is merely the most-frequent (modal) answer. That is, for a given decision problem ω a process b selects its favored response $b(\omega)$ only with a certain probability

$1 > P_b(\omega) \geq \frac{1}{2}$, given by a function $P_b : \Omega_b \mapsto [\frac{1}{2}, 1)$, whereas with the remaining probability it selects the other alternative.[2]

We assume that there is a distinguished process d with $\Omega_d = \Omega$, which we call the *deliberative* process. We refer to any other process h with $\Omega_h \subseteq \Omega$ that is not the deliberative process as a *heuristic process* or simply a *heuristic*. The interpretation is that the deliberative process is always active, whereas each heuristic is triggered by certain cues that are present in some and absent in other decision problems.

2.1 The Internal Consistency of a Decision Process

Since the favored response of a decision process is simply its modal option, and not necessarily a normatively desirable option, the probability $P_b(\omega)$ does not capture the frequency of "correct" choices, but rather the level of *internal consistency* of process b in decision problem ω. The concept of internal consistency will be helpful to model both dual-process considerations and strength-of-preference effects. The first relates to consistency across processes for a given decision problem, and the second to consistency within a process across different decision problems.

Consider consistency across processes. Dual-process theories consider heuristics as decision processes capturing more intuitive or impulsive behavioral tendencies than deliberation. In psychology, this corresponds to the concept of *automaticity* (Allport, 1954; Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977; Cohen et al., 1990; Kahneman, 2003). More automatic processes are closer to stimulus-response reactions than deliberation, and in particular they are faster and more internally consistent. In other words, deliberation is generally less internally consistent than the more automatic reactions associated with heuristics (e.g., Redish, 2016). In our terms, this will correspond to the assumption that $P_h(\omega) > P_d(\omega)$ for the appropriate set of decision problems. To this end, we introduce the following notation.

Definition 1. Process b_1 is *more consistent* than process b_2 on a set of decision problems $\Omega' \subseteq \Omega_{b_1} \cap \Omega_{b_2}$ if $P_{b_1}(\omega) > P_{b_2}(\omega)$ for all $\omega \in \Omega'$.

We now turn to consistency within a process. Psychometric or strength-of-preference effects state that, especially for deliberative processes, consistency varies systematically with choice difficulty. In particular, consistency is systematically lower for options that are more similar (Dashiell, 1937; Moyer and Landauer, 1967; Laming, 1985; Klein, 2001; Wichmann and Hill, 2001), a regularity that has also been established for economic decision making (Mosteller and Nogee, 1951; Alós-Ferrer and Garagnani, 2022a,b). To capture this effect, each decision problem $\omega = \{A_\omega, B_\omega\}$ is endowed with a cardinal difference Δ_ω between the options A_ω and B_ω. The nature of Δ_ω will depend on the specific application. For example, in the context of lottery choice, a natural notion of choice difficulty might be the difference in certainty equivalents between options. We say

[2]This allows for cases where the probability is exactly $1/2$ and hence there are two modal options. The choice of which alternative is considered favored in this knife-edge case is inconsequential.

that a process is psychometric if the probability to select A_ω from the decision problem ω is an increasing function of Δ_ω. To be able to formally state this, we first define

$$\beta_b^A(\omega) = \begin{cases} P_b(\omega) & \text{if } b(\omega) = A_\omega \\ 1 - P_b(\omega) & \text{if } b(\omega) = B_\omega, \end{cases}$$

i.e., the probability that process b selects A_ω for decision problem ω. Strength of preference for a given process is then captured by the following definition.

Definition 2. The process b is *psychometric* if β_b^A is a continuous, strictly increasing function of Δ_ω with $\beta_b^A(\omega) = \frac{1}{2}$ for $\Delta_\omega = 0$.

In particular, since $b(\omega)$ singles out the modal option if there is one, for a psychometric process b it follows that $d(\omega) = A_\omega$ for every $\Delta_\omega > 0$ and $d(\omega) = B_\omega$ for every $\Delta_\omega < 0$.

2.2 Conflict, Alignment, and Neutral Decision Problems

We will consider applications where the researcher has two candidates as the likely decision processes influencing actual decisions, a deliberative process d and an alternative, heuristic process h. For example, in the objective domain, the researcher might be contrasting normatively optimal decisions in a belief-updating task against a particular bias, e.g. conservativeness, the representativeness heuristic, etc. In the subjective domain, the researcher might be comparing preference/based decisions (e.g., on the basis of expected utility) and a heuristic hypothesized to underlie a particular behavioral anomaly.

For a given decision problem $\omega \in \Omega_h$, we say that a heuristic process h is in *conflict* with the deliberative process d if $h(\omega) \neq d(\omega)$, and we say that it is in *alignment* (with the deliberative process) if $h(\omega) = d(\omega)$. The set of decision problems Ω_h where a heuristic h is active can thus be partitioned into the a set of alignment situations $\Omega_h^A = \{\omega \in \Omega_h \mid h(\omega) = d(\omega)\}$ and a set of conflict situations $\Omega_h^C = \{\omega \in \Omega_h \mid h(\omega) \neq d(\omega)\}$.

Further, recall that the deliberative process is always active ($\Omega_d = \Omega$), whereas the heuristic process is triggered by certain cues of the decision problem ω, and is hence only active in a certain subset Ω_h of decision problems. Thus, $\Omega_h^N = \Omega \setminus \Omega_h$ is the set of all decision problems where the deliberative process d is the only active process and, thus, determines behavior. We refer to these as *neutral* decision problems, which should be identifiable for specific applications.

Let $\rho(\omega)$ denote the observable choice probability of A_ω in decision problem ω. If both processes are active, we assume that which process determines behavior is a stochastic event. Specifically, with a probability $(1 - \alpha)$ the actual response is determined by process d and with probability α by process h. In dual-process terms, $1 - \alpha$ could be conceived of as the probability of successful inhibition of the impulse represented by h. This is assumed to be an exogenous constant for simplicity, but the results can be extended to the case where α is a function of strength of preference (see Section 6.1).

The observed stochastic choice function of a decision maker is thus given by

$$\rho(\omega) = \begin{cases} (1-\alpha)\beta_d^A(\omega) + \alpha\beta_h^A(\omega) & \text{if } \omega \notin \Omega_h^N \\ \beta_d^A(\omega) & \text{if } \omega \in \Omega_h^N, \end{cases}$$

which takes into account that, for neutral decision problems, only the deliberative process is active.

3 The Interaction of Noise and Bias

We consider a decision maker whose behavior is co-determined by a deliberative process d and some heuristic process h. The deliberative process exhibits strength-of-preference effects as described above, with $d(\omega) = A_\omega$ if $\Delta_\omega > 0$ and $d(\omega) = B_\omega$ otherwise. Generally speaking, the heuristic process might also exhibit strength-of-preference effects, possibly depending on a different attribute or characteristic of the decision problem than that captured by Δ_ω. However, in many applications heuristics depend on the presence or absence of specific triggers, and hence it is reasonable to treat them as having constant consistency, i.e., $P_h(\omega) = P_h$ for some fixed $P_h > 1/2$. For example, the certainty heuristic might be triggered by the presence of a riskless option, or conservativeness might simply focus on the event with the highest prior probability.

For simplicity, in the following we refer to the alternative prescribed by the deliberative process as the *correct* choice and to the other alternative as an *error*. In the objective domain, where the deliberative process will typically capture normative behavior, this will correspond to actual, objectively/defined correct answers and errors. In the subjective domain, if the deliberative process corresponds to preference maximization, an error is simply a decision that contradicts the individual preferences, but not necessarily a normatively incorrect choice. In particular, one decision maker's errors might be another decision maker's correct choices. In any case, this means that Δ_ω captures a value difference between options such that A is objectively or subjectively correct if $\Delta_\omega > 0$ and B is correct otherwise.

3.1 Error Rates and Choice Consistency

Theorem 1 below describes the stochastic choice function of a decision maker who displays both process multiplicity and strength-of-preference effects. As discussed above, following the interpretation of a heuristic as a more automatic process than the deliberative one, we assume the former process to be more internally consistent than the latter. The results are stated in terms of the probability of a correct answer, that is, the probability of a choice that is in line with the prescription of the deliberative process d, $\text{Prob}(\rho(\omega) = d(\omega))$.

Theorem 1. *Suppose the deliberative process d is psychometric and the heuristic h has constant consistency $P_h > 1/2$. Then the following statements hold.*

(a1) For neutral decision problems, $\omega \in \Omega_h^N$, the probability of a correct answer is increasing in Δ_ω for $\Delta_\omega > 0$, and decreasing in Δ_ω for $\Delta_\omega < 0$. Further, the probability is 1/2 for $\Delta_\omega = 0$

(a2) For decision problems in alignment, $\omega \in \Omega_h^A$, the probability of a correct answer is increasing in Δ_ω for $\Delta_\omega > 0$, and decreasing in Δ_ω for $\Delta_\omega < 0$.

(a3) For decision problems in conflict, $\omega \in \Omega_h^C$, the probability of a correct answer is increasing in Δ_ω for $\Delta_\omega > 0$, and decreasing in Δ_ω for $\Delta_\omega < 0$.

(b1) For any given Δ, the probability of a correct response is larger for decisions under alignment compared to conflict decision problems. That is, for any fixed Δ and decision problems $\omega^C \in \Omega_h^C, \omega^A \in \Omega_h^A$ with $\Delta_{\omega^C} = \Delta_{\omega^A} = \Delta$, it follows that

$$\text{Prob}(\rho(\omega^A) = d(\omega^A)) > \text{Prob}(\rho(\omega^C) = d(\omega^C)).$$

(b2) For any given Δ, the probability of a correct response is smaller for decisions under conflict compared to neutral decision problems. That is, for any fixed Δ and decision problems $\omega^C \in \Omega_h^C, \omega_h \in \Omega_h^N$ with $\Delta_{\omega^C} = \Delta_{\omega^N} = \Delta$, it follows that

$$\text{Prob}(\rho(\omega^N) = d(\omega^N)) > \text{Prob}(\rho(\omega^C) = d(\omega^C)).$$

(b3) Suppose the heuristic h is more consistent than d on Ω_h. Then, for any given Δ, the probability of a correct response is larger for decisions under alignment compared to neutral decision problems. That is, for any fixed Δ and decision problems $\omega^A \in \Omega_h^A, \omega^N \in \Omega_h^N$ with $\Delta_{\omega^A} = \Delta_{\omega^N} = \Delta$, it follows that

$$\text{Prob}(\rho(\omega^A) = d(\omega^A)) > \text{Prob}(\rho(\omega^N) = d(\omega^N)).$$

The proof of this theorem is in the Appendix. Figure 1 illustrates the result. In words, if we restrict to a given class of decision problems (neutral, conflict, or alignment), the probability of a correct response increases as the difference between the alternatives, as captured by Δ_ω, becomes larger. However, that probability is uniformly shifted down for conflict, compared to neutral decision problems. For alignment, the probability of a correct response is shifted up for decisions closer to indifference, but it might also be shifted down (compared to neutral) for decisions very far away from indifference (Figure 1, right), but it will remain above the corresponding probability for conflict.

The intuition for Theorem 1 is as follows. First, focusing on neutral decision problems, where only the deliberative process is active, strength-of-preference effects result in a pattern where errors are more frequent when the cardinal difference between the

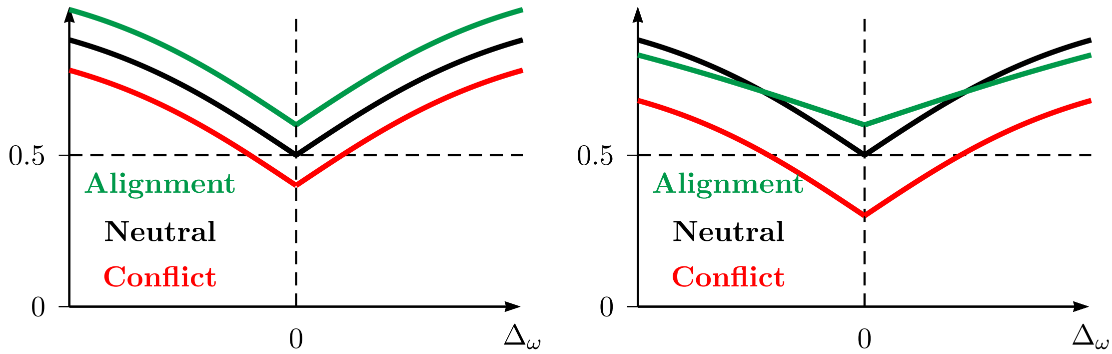


Figure 1: Illustration of Theorem 1, depicting the probability of a correct answer as a function of the cardinal difference between the alternatives, Δ_ω.

alternatives is smaller, because the decision is then more difficult, in the sense that alternatives are harder to tell apart. Second, for decisions under alignment or conflict, the pattern remains, but there is an additional probability that decisions are made by the heuristic process instead. For conflict problems, the heuristic very often leads to an error, which shifts the probability of a correct decision downwards. For alignment problems, the heuristic is very often correct, which in principle should shift the probability of a correct decision upwards. However, there is an additional effect which is reflected in part (b3) and the right-hand side of Figure 1.

This additional effect is related to the fact that none of the results of Theorem 1 depend on the additional assumption that the heuristic process is more consistent than the deliberative one, except for part (b3). This is important, because strength of preference effects imply that the deliberative process might be very inconsistent when close to indifference, but it must also become very consistent when Δ is large (easy choices). Hence, if the heuristic process has constant or near-to-constant consistency, the assumption that the deliberative process is more inconsistent will typically hold for most decisions, but be violated if the design includes decision problems which are far away from indifference. That is, one would in general expect $P_d(\omega) < P_h$ in most of the relevant domain, but $P_d(\omega) > P_h$ as Δ_ω grows large. As Theorem 1 shows, this possibility does not affect any of the results above, except possibly for the comparison of error rates between decisions in alignment and neutral decision problems. In other words, while we expect the probability of correct responses in experiments to roughly reflect the left-hand side of Figure 1, in designs including large values of Δ, they could reflect the picture in the right-hand side. This is because, for such large values, even though the heuristic process favors the correct response with a large probability, the deliberative process would deliver the correct response with an even larger probability.

The interest of the results above lie on the fact that they deliver predictions that are testable in any application where the researcher has good candidates for alternative processes influencing behavior, but strength-of-preference effects and process multiplicity might interact and confound each other. The next subsection shows the implications of the result for testability in a specific application.

In particular, as Figure 1 illustrates, an important implication of Theorem 1 is that the monotonic relation between error rates and cardinal differences between alternatives in a specific application might only become clear if one distinguishes decision problems according to process interaction, i.e. conflict or alignment (or neutral). Conversely, the effects of the presence of a specific heuristic might only become clear if one considers the psychometric relation between error rates and cardinal differences between alternatives. In other words, if the data is analyzed without taking into account whether decision problems involve conflict or alignment (or neither), the researcher might fail to realize that decisions reflect strength-of-preference effects. Conversely, if the data is analyzed without taking into account possible strength-of-preference effects, the researcher might fail to identify process multiplicity and its consequences, for example if decisions under conflict or alignment are not equally distributed along the Δ axis.

3.2 Application: Conservatism

In this section, we apply the model from Section 2 to a new experiment which examines a standard, well-known bias from the literature on belief updating and decision making: conservatism. This bias (Edwards, 1968; Fischhoff and Beyth-Marom, 1983; El-Gamal and Grether, 1995; Enke and Graeber, 2023) has been consistently observed in decision making under uncertainty and corresponds to overweighting prior probabilities which would need to be updated, using Bayes' rule, when new information on an uncertain event becomes available (or even fully ignoring the new information). This and other biases in belief updating are frequently studied using experimental designs with different urns containing balls of different colors in known proportions (e.g., Grether, 1980, 1992; El-Gamal and Grether, 1995; Alós-Ferrer and Garagnani, 2023). Achtziger et al. (2014) used electroencephalography to show that decision makers with a large proportion of conservative errors started motor action preparation to give the conservative answer *before* the new information was even presented, suggesting that conservativeness relies on a highly automatic process which often ignores new sample information.

Thus, to model conservatism in our (first) experiment, we consider two processes: Bayesian updating and Conservatism. The deliberative process d is given by Bayesian updating, which prescribes to base behavior on the correctly updated posterior belief given the sample information. That is, Bayesian updating relies on a cardinal difference between the options. Strength of preference effects are hence operationalized based on this cardinal difference. The heuristic process h is Conservatism, which is active whenever there is an option with a higher prior probability. This heuristic, however, ignores the sample information and prescribes to choose the option with the higher prior probability. Thus, this process should exhibit constant consistency as assumed in Theorem 1.

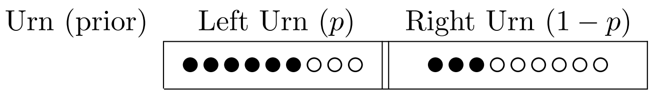


Figure 2: Schematic representation of the task used in the Conservatism Experiment. The figure shows an example with the urn compositions for $S = 9$, $n_L = 6$, and $n_R = 3$.

3.2.1 Experimental Design

For our conservatism experiment, we recruited a total of 166 participants (87 females) from the student population of the University of Zurich, with a median age of 23 ranging from 18 to 49. Participants took part in 5 experimental sessions, with between 30 and 36 individuals per session. The experiment was run at the Laboratory for Experimental and Behavioral Economics of the University of Zurich and was programmed in PsychoPy (Peirce, 2007).

The experiment used a variant of a well-established belief-updating task (Grether, 1980, 1992; El-Gamal and Grether, 1995; Achtziger et al., 2014). There are two urns (left and right) containing S colored balls each (black or white). The left urn contains n_L black and $S - n_L$ white balls, and the right urn contains n_R black and $S - n_R$ white balls (see Figure 2 for an exemplary schematic representation). We implemented a symmetric urn design, i.e., $n_R = S - n_L$. In each trial, the left urn was randomly selected with a given, known probability p (the prior), and the right urn with a the complementary probability $1 - p$. Then, a sample of four balls was extracted from the selected urn and shown to the participant, who was then asked to guess whether the sample was drawn from the left or the right urn.

There were a total of 300 trials divided into four blocks of 75 trials each, with a 30 seconds break between blocks. The order of blocks was as given in Table 1 for half the subjects and counterbalanced by using the reverse order for the other half of subjects. Each block used a different urn composition (fixed for all trials within a block) and three different priors $p_1 < 1/2$, $p_2 = 1/2$, $p_3 > 1/2$. See Table 1 for an overview of the urn compositions and priors. In each block there were 30 trials with prior p_1, 15 trials with prior $p_2 = 1/2$, and 30 trials with prior p_3. The participant was informed about the urn composition at the beginning of the block. At the beginning of each trial the participant was informed about the prior p for the current trial. Then one of the two urns was randomly selected according to the prior p, but the participant was not informed of which urn was selected. Subsequently, a sample of 4 balls was randomly extracted with replacement from the selected urn and shown to the participant on screen. Participants then guessed from which urn the sample was drawn and the trial ended. Participants received no feedback before all 300 trials were completed. At the end of the experiment, participants received 15 Rappen (0.15 Swiss Francs, CHF) for each correct guess. Sessions lasted about 60 minutes, including instructions and payment. Average payment was 50 CHF.

Block	S	n_L	n_R	p_1	p_2	p_3
1	10	7	3	$\frac{3}{10}$	$\frac{1}{2}$	$\frac{7}{10}$
2	9	6	3	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$
3	8	5	3	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{5}{8}$
4	10	6	4	$\frac{2}{5}$	$\frac{1}{2}$	$\frac{3}{5}$

Table 1: Conservatism Experiment. Urn compositions and priors for the four blocks. Each block comprised 75 trials. Order of blocks was counterbalanced.

We consider two processes: Bayesian updating and Conservatism. Bayesian updating prescribes to guess the most likely urn given the prior and the sample (except in case of exact ties). Conservatism ignores the sample information and prescribes the urn with the higher prior probability (if there is such an urn). For a given decision problem, the two processes are in *alignment* if they prescribe the same choice. Otherwise, they are in *conflict*. Whenever the prior is $p_2 = 1/2$, Conservatism is not active, hence, those decision problems are *neutral*.

For a given block, each decision problem ω is characterized by (p, m) where $p \in \{p_1, p_2, p_3\}$ is the prior probability that the sample is drawn from the left urn and $m \in \{0, 1, 2, 3, 4\}$ is the number of black balls in the sample. The design hence generates 15 different possible decision problems per block. The prescriptions of Bayesian updating and Conservatism for all decision problems are summarized in Table 2. For each $\omega = (p, m)$, the table reports the difference between the expected payoffs (in Rappen) of the left urn (L) and the right urn (R), denoted by Δ_ω. Note that, since correct guesses were paid at a constant rate and incorrect guesses were not paid, expected payoffs are just a linear transformation of the probabilities. The absolute value $|\Delta_\omega|$ provides an inverse measure of the difficulty for each (p, m)-problem. There are 24 alignment situations with 12 different levels of difficulty ($|\Delta_\omega|$), and 16 conflict situations with 8 different levels of difficulty. The remaining 20 situations are neutral ($p = 1/2$), involving 9 different levels of difficulty.[3]

3.2.2 Establishing the Presence of Conservatism: Grether Regression

Before analyzing the results in the light of Theorem 1, we briefly establish the presence of conservatism bias in our data using a standard approach from the literature. Following Grether (1980), we consider the following probit regression (see also Enke and Graeber, 2023)

$$Y_{it} \sim \alpha + \beta_1 \ln \frac{P(m|L)}{P(m|R)} + \beta_2 \ln \left(\frac{p}{1-p} \right) + \varepsilon_{it}, \tag{1}$$

[3]In the knife-edge cases where the prior was 1/2 and exactly two balls of each type were extracted, there is no "correct" answer and Bayesian updating should prescribe $P_d = 1/2$.

Table 2: Expected Payoff Differences (Left minus Right urn) for a sample with m black balls and prior p. Each cell shows the expected payoff differences (top), the prescription of Conservatism (bottom left) and Bayesian updating (bottom right). Dark gray, light gray, and white cells refer to conflict, alignment and neutral situations, respectively.

Block 1: Urns with 10 balls. Left urn 7 black, 3 white. Right urn 3 black, 7 white.

$p \setminus m$	0		1		2		3		4	
3/10	-14.57		-12.81		-6.00		6.00		12.81	
	R	R	R	R	R	R	R	L	R	L
1/2	-14.02		-10.34		0.00		10.34		14.02	
	-	R	-	R	-	-	-	L	-	L
7/10	-12.81		-6.00		6.00		12.81		14.57	
	L	R	L	R	L	L	L	L	L	L

Block 2: Urns with 9 balls. Left urn 6 black, 3 white. Right urn 3 black, 6 white.

$p \setminus m$	0		1		2		3		4	
1/3	-14.09		-11.67		-5.00		5.00		11.67	
	R	R	R	R	R	R	R	L	R	L
1/2	-13.24		-9.00		0.00		9.00		13.24	
	-	R	-	R	-	-	-	L	-	L
2/3	-11.67		-5.00		5.00		11.67		14.09	
	L	R	L	R	L	L	L	L	L	L

Block 3: Urns with 8 balls. Left urn 5 black, 3 white. Right urn 3 black, 5 white.

$p \setminus m$	0		1		2		3		4	
3/8	-12.84		-9.67		-3.75		3.75		9.67	
	R	R	R	R	R	R	R	L	R	L
1/2	-11.56		-7.06		0.00		7.06		11.56	
	-	R	-	R	-	-	-	L	-	L
5/8	-9.67		-3.75		3.75		9.67		12.84	
	L	R	L	R	L	L	L	L	L	L

Block 4: Urn with 10 balls. Left urn 6 black, 4 white. Right urn 4 black, 6 white.

$p \setminus m$	0		1		2		3		4	
2/5	-11.51		-8.14		-3.00		3.00		8.14	
	R	R	R	R	R	R	R	L	R	L
1/2	-10.05		-5.77		0.00		5.77		10.05	
	-	R	-	R	-	-	-	L	-	L
3/5	-8.14		-3.00		3.00		8.14		11.51	
	L	R	L	R	L	L	L	L	L	L

where $Y_{it} = 1$ if the Left urn was chosen and 0 otherwise, $\frac{P(m|L)}{P(m|R)}$ is the likelihood ratio for L given the sample, $\frac{p}{1-p}$ is the prior odds in favor of L, and ε_{it} is an error term with mean zero and finite variance. Bayes' rule prescribes equal weight on the evidence and the prior ($\beta_1 = \beta_2 > 0$) while conservatism prescribes greater weight on the prior

Table 3: Random effects probit regression on the probability of an L choice, excluding neutral situations.

L chosen	
Likelihood ratio for L	0.825***
	(0.008)
Prior odds for L	1.532***
	(0.016)
Constant	-0.033***
	(0.012)
Linear Combination test	
Likelihood ratio for L - Prior odds for L	0.707***
	(0.014)
Observations	39840
Subjects	166

Notes: Robust standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

($\beta_2 > \beta_1 \geq 0$). The difference $\beta_2 - \beta_1$ can be seen as continuous measure of the bias. We expected that actual behavior would exhibit conservatism, that is, $\beta_2 - \beta_1 > 0$ in non-neutral situations.

Table 3 reports the results of this regression for non-neutral situations. Both coefficients are positive and significant. Importantly, however, a linear combination test reveals that the weight on the prior is significantly larger than the weight on the likelihood ratio, that is, $\beta_2 - \beta_1 > 0$. Thus, we conclude that conservatism is present in our data.

3.2.3 Main Results (Conservatism)

We now test the predictions of the model outlined in Section 2 to the data from the conservatism experiment described above. That is, we aim to distinguish noise, as reflected by strength of preference effects, from the conservatism bias.

For the purposes of the application, a guess is *correct* if it is in line with the prescription of Bayesian updating and an *error* otherwise. Theorem 1 provides predictions on the relation between the cardinal difference Δ, which in this case captures differences in expected payoffs, and the proportion of correct choices, assuming that Bayesian updating is psychometric and that the conservatism bias reflects a heuristic process with has constant consistency. The latter assumption is reasonable since conservatism essentially focuses on the prior and ignores the sample information which underlies both the cardinal difference Δ and strength-of-preference effects.

In view of the literature described above, it is also reasonable to assume that, for non-neutral decision problems, the heuristic process driving the conservatism bias is more consistent than Bayesian updating, except for large values of Δ. However, recall that this assumption is not necessary for the results, except for Theorem 1(b3).

Figure 3 illustrates the results. Each dot in this figure depicts the proportion of correct responses for all decision problems sharing the same value of (m, p) for a given block (recall Table 1). In particular, each dot reflects a proportion of correct responses for a given value of Δ, computed as the difference in expected values between Left and Right. The figure distinguishes decision problems under alignment (green) or conflict (red), and neutral decision problems (black). For better visibility, it also includes simple regression lines for each type of decision problem, conditional on the sign of Δ.

As the figure shows, the probability of a correct answer increases as decisions are further away from indifference, i.e., when Δ is larger in absolute value. This holds separately for conflict, alignment, and neutral decisions, as predicted by Theorem 1(a1–a3). Hence, we directly observe the effects of strength of preference in our data. At the same time, there is a very clear difference between decisions under conflict and those under alignment, with the former having a much lower probability of a correct response than the latter. This corresponds to Theorem 1(b1). The probability of a correct response is also clearly shifted down for conflict compared to neutral decision problems, as predicted by Theorem 1(b2). That is, we clearly observe the effects of process multiplicity (bias) even after disentangling them from strength of preference (noise).

The only prediction of Theorem 1 relying on an additional assumption (that the heuristic process is more internally consistent) is (b3), which states that the probability of a correct response should be larger for decisions under alignment compared to neutral decision problems. This is not contradicted by the figure, even though the probabilities seem to be similar, possibly due to a ceiling effect.[4]

For a formal test of the model's predictions, we turn to random-effects probit regression analyses on correct guesses. Strength of preference effects (noise) are captured by the independent variable $|\Delta|$ (difficulty), while the effects of process multiplicity (bias) will be reflected by dummy variables indicating *Conflict* and *Alignment*. To test for the effects of strength of preference for different kinds of decision problems (as derived from process multiplicity), we also include the interactions $|\Delta| \times$ *Conflict* and $|\Delta| \times$ *Alignment*.

Table 4 reports the results of these regressions. We exclude observations with $\Delta = 0$ ($p_2 = 1/2$ and $m = 2$) as correct responses are not well defined. Models 1 to 3 are restricted to the subsamples of neutral decision problems, and decision problems under alignment and conflict, respectively. For all three cases, the probability of a correct answer is increasing in the difference in expected payoffs, $|\Delta_\omega|$, reflecting strength of preference effects and as predicted by Theorem 1(a1–a3). Model 4 shows that the probability of a correct answer is larger in decision problems under alignment and smaller for those under conflict, compared to neutral ones. This is as predicted by Theorem 1(b1–b3).

[4]Note that the four data points at exactly $\Delta = 0$ for neutral decision problems corresponds to problems with $p_2 = 1/2$ where, by chance, exactly two balls of each color were sampled. In this case, both answers are actually correct and the data points depict the proportion of "Left" answers, which are of course close to 0.5.

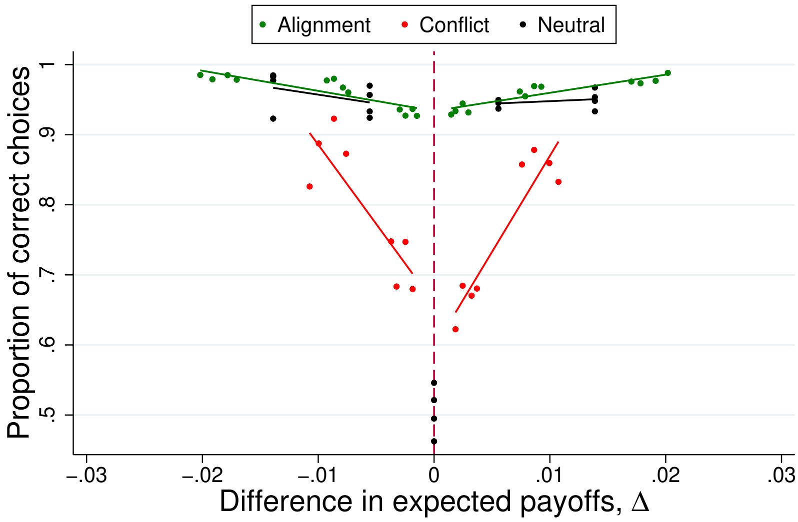


Figure 3: Proportion of correct choices as a function of $\Delta = EV(L) - EV(R)$ separately for alignment (green), conflict (red), and neutral (black) situations.

In model 5, we simultaneously examine the effects of noise and bias, and show that their effects can be disentangled and observed simultaneously. That is, even in the complete sample, which includes neutral decisions and decisions under conflict and alignment, the effects of strength of preference can be seen for all three types of decisions (significantly positive coefficient of Δ for neutral decisions, and significantly positive interactions for decisions under alignment and conflict). Further, the dummy for conflict is significant and negative, showing that process conflict reduces the probability of a correct response. The dummy for alignment, however, is not significant in the joint regression model.

In conclusion, our analysis confirms the predictions of Theorem 1 and show that both noise (strength of preference) and bias (process multiplicity) play a role in a classical economic task. This supports our main message, i.e., that focusing on either cognitive imprecision while ignoring process multiplicity or on dual-process effects while ignoring strength-of-preference regularities is likely to result in an incomplete picture of the determinants of errors in economic decision making.

4 Detecting Multiple Processes

The previous section has shown that errors in economic decisions arise both from cognitive imprecision, in the form of strength-of-preference effects, and process multiplicity, which generate dual-process effects. In the analyses we carried out in that section, we show that each kind of effect can be detected in the data even when controlling for the other kind of effect. In this section, we adopt a more diagnostic approach, showing that

Table 4: Random effects Probit regression on probability of a correct answer.

Correct	Neutral (1)	Alignment (2)	Conflict (3)	All (4)	All (5)
$\|\Delta\|$	31.638***	59.229***	120.008***		28.620***
	(9.570)	(3.567)	(6.101)		(8.714)
Alignment				0.072**	-0.069
				(0.033)	(0.075)
Conflict				-1.261***	-1.510***
				(0.034)	(0.076)
$\|\Delta\| \times$ Alignment					28.861***
					(9.322)
$\|\Delta\| \times$ Conflict					78.083***
					(10.396)
Constant	2.111***	1.863***	0.437***	2.072***	1.878***
	(0.130)	(0.071)	(0.093)	(0.069)	(0.095)
Observations	6799	28915	10925	46639	46639
Subjects	166	166	166	166	166

Notes: Robust standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

certain *discontinuities* in the data (which might be lost or overlooked in standard regression analyses) can be used to establish the presence of a heuristic (i.e., detect process multiplicity) even in the presence of strength of preference effects.

To understand the approach, recall that we model decision problems as pairs $\omega = \{A_\omega, B_\omega\}$, with the understanding that the labels A and B might have particular meanings for a specific application. This convention can be used to simplify notation by using the labels to reflect particular cues or triggers on which the heuristic processes act. Formally, a heuristic process h is *triggered by* A (resp. B) if, for every $\omega \in \Omega_h$, $h(\omega) = A_\omega$ (resp. $h(\omega) = B_\omega$).

For example, consider a dataset involving lottery choices in the gain domain. Each lottery pair is of the form $\omega = \{S_\omega, R_\omega\}$, where the R lottery is riskier than the S lottery. Consider a *certainty heuristic* h, which always favors riskless lotteries (i.e., those with a sure outcome), but is inactive if both lotteries in a pair involve some risk. The domain of the heuristic, Ω_h, contains all pairs where S_ω is riskless, and only those. The certainty heuristic prescribes $h(\omega) = S_\omega$ for all $\omega \in \Omega_h$ and is silent for $\omega \in \Omega_h^N = \Omega \setminus \Omega_h$. That is, the certainty heuristic is triggered by S.

4.1 Choice Bias and Discontinuities in Error Rates

If decisions are affected by a heuristic which is triggered by a specific action, our model predicts a clear discontinuity in the data. This is because, for this kind of heuristic, it is possible the sharply separate decision problems under conflict and under alignment. Recall that, if the deliberative process d is psychometric, then $d(\omega) = A_\omega$ for every $\Delta_\omega > 0$ and $d(\omega) = B_\omega$ for every $\Delta_\omega < 0$. If the heuristic process h is triggered by A,

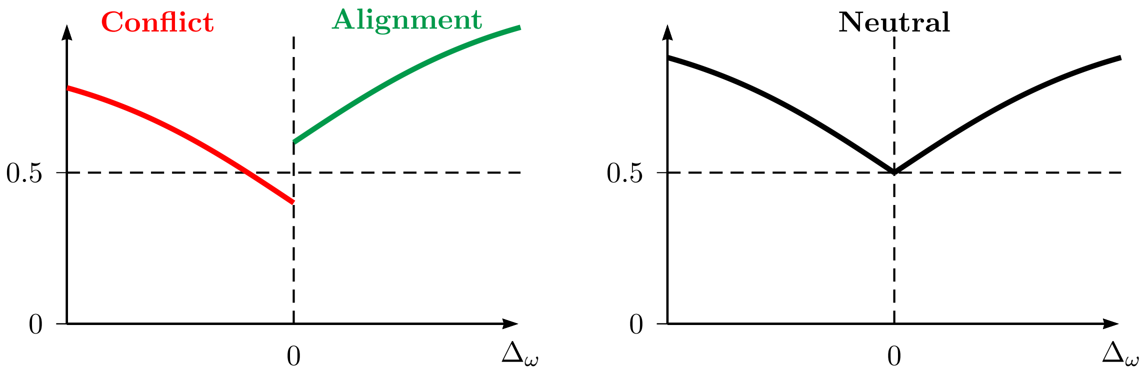


Figure 4: Illustration of Theorem 2, depicting the probability of a correct answer as a function of the cardinal difference between the alternatives, Δ_ω, when the heuristic is triggered by an action. Left: Decision problems in conflict and alignment are separated by the sign of Δ_ω, creating a discontinuity. Right: There is no discontinuity for neutral decision problems.

then every decision problem $\omega \in \Omega_h$ with $\Delta_\omega < 0$ is in conflict, because $d(\omega) = B_\omega$ but $h(\omega) = A_\omega$. Analogously, every decision problem $\omega \in \Omega_h$ with $\Delta_\omega > 0$ is in alignment, because $d(\omega) = A_\omega = h(\omega)$. Hence, for decision problems where the heuristic is actually triggered, conflict and alignment are sharply (and even graphically) distinguished.

This intuition is illustrated by the comparison between Figures 1 and 4. By Theorem 1, the probability of a correct response is shifted up for decision problems in alignment and shifted down for decision problems in conflict. If the heuristic is triggered by A, decision problems in alignment are those with positive Δ, and decision problems in conflict are those with negative Δ. Hence, the effects described in Theorem 1 create a discontinuity (an upwards jump) at $\Delta = 0$ (Figure 4, left). In contrast, this discontinuity should be absent for neutral decision problems (Figure 4, right). The following result collects these observations. The proof is a straightforward application of Theorem 1 and hence omitted.

Theorem 2. *Suppose the deliberative process d is psychometric and the heuristic h has constant consistency $P_h > 1/2$ and is triggered by A. Then the following statements hold.*

(a) The probability of a correct answer is increasing in Δ_ω for $\Delta_\omega > 0$ (alignment) *and decreasing in Δ_ω for $\Delta_\omega < 0$* (conflict).

(b) For decision problems where h is active, $\omega \in \Omega_h$, there is a discontinuity at $\Delta_\omega = 0$ with more correct answers in alignment ($\Delta_\omega > 0$) than in conflict ($\Delta_\omega < 0$).

(c) For neutral decision problems, there is no discontinuity and the probability of a correct answer at $\Delta_\omega = 0$ is $1/2$.

Regression analyses, however, are not well-suited to detect the effects predicted in Theorem 2. Analyses such as those illustrated in Figure 3 and reported in Table 4 essentially estimate a linear (decreasing) trend for $\Delta < 0$ and another linear (increasing) trend for $\Delta > 0$. While this is enough to test for the predictions of Theorem 1, if the relationship between Δ and error rates is nonlinear (as will generally be the case), one will typically observe some jump between the intersects at $\Delta = 0$, and a dummy capturing the sign of Δ will typically be significant, but these results are mechanical and do not reflect actual discontinuities.

To formally test and quantify the jump between conflict and alignment for a given choice dataset, one can rely on a local polynomial regression discontinuity estimation with robust bias-corrected confidence intervals as detailed in Calonico et al. (2014). The idea is not to rely on bins (as in Figure 3) or on the difference between intersects (as in regression tables), but rather use the observations around the predicted discontinuity threshold to investigate statistical differences. Intuitively, this is analogous and closely-related to the procedure used to test for the effectiveness of policies in regression discontinuity designs (e.g., He et al., 2020). The procedure constructs robust, non-parametric confidence intervals for the average "treatment effect" (the jump at zero) at the specified cutoff conditional on mean-squared-error optimal bandwidths. In the application below, we will rely on this procedure to test the predictions of Theorem 2(b–c), and present additional graphical and regression analyses for illustrative purposes, as well as to confirm strength-of-preference effects (Theorem 2(a)).

The next corollary shows that the size of the discontinuity at $\Delta = 0$ is meaningful. First, quantifying this discontinuity allows to back up a lower bound on the probability α that the heuristic process makes the decision. Second, the jump at zero is predicted to be symmetric, which affords a further opportunity to test the model.

Corollary 1. *Suppose the deliberative process d is psychometric and the heuristic h has constant consistency $P_h > 1/2$ and is triggered by A.*

(a) For decision problems where h is active, the discontinuity at $\Delta_\omega = 0$ is equal to $\bar{D} = \alpha(2P_h - 1)$.

(b) In particular, the parameters α and P_h are partially identified by $\bar{D}$. Specifically, $\alpha > \bar{D}$ and $P_h > \frac{\bar{D}+1}{2}$.

(c) Further, the left and right parts of the discontinuity are symmetric. That is, if P_- and P_+ are the limits of the probabilities of a correct response as $\Delta_\omega \to 0-$ and $\Delta_\omega \to 0+$, respectively, then

$$\frac{1}{2} - P_- = P_+ - \frac{1}{2}.$$

4.2 Application: Certainty Effect

The *certainty effect* refers to the robust empirical observation that, when choosing between a sure but small amount and a larger but risky amount, many participants choose the safe alternative, but when the winning probabilities are reduced in ways that should not alter the decision of an expected utility maximizer, most people tend to choose the riskier option instead. This effect, which is one of the most prominent empirical violations of Expected Utility Theory, can be observed in different ways, depending on how the related choice pairs are constructed. The *Allais paradox* (Allais, 1953) reflects a *common consequence effect*, where the same winning probability is transferred to a losing outcome for both lotteries to construct the second pair. The *common ratio effect* (Kahneman and Tversky, 1979) uses a different construction where winning probabilities are scaled down by the same ratio. In both cases, the hypothesis is that choices for the second pair, where no lottery is riskless, reflect the fact that most people are risk averse, but choices for the first pair are affected by an additional decision process which is triggered by the presence of a riskless (certain) alternative: a *certainty heuristic*. In our terms, such a heuristic process would be triggered by the certain alternative.

The common ratio and common consequence effects have been widely replicated (see, e.g., Ruggeri et al., 2020; Blavatskyy et al., 2023). Specifically, preference reversals where participants choose the sure option if one is available but the risky one if not are found to be much common than the opposite ones. However, a recent contribution by McGranaghan et al. (2024) (see also Ballinger and Wilcox, 1997; Hey, 2005; Blavatskyy, 2007, 2010) has argued that previous evidence for the certainty effect might actually be inconclusive and simply reflect strength-of-preference effects. The argument is as follows. When participants choose between a small, sure amount and a larger, risky one, they might choose the risky one by mistake (noise) even if they actually prefer the sure amount. However, if the difference in utilities (strength of preference) is large, such mistakes will be rare. In the common ratio and common consequence effects, the winning probabilities are reduced (and transferred to zero outcomes) to create the second pair. This essentially reduces the difference in utilities for that second pair, and hence the strength of preference. Thus, as long as the distribution of noise remains largely unchanged, the probability of an error becomes much larger for the second pair. This means that reversals where the sure option is taken if available but the riskier one is chosen for the second pair will be more likely than the opposite ones. McGranaghan et al. (2024) argued that previous tests, which typically have used paired choices, do not convincingly demonstrate the existence of a certainty effect. Using alternative tests based on paired valuations, McGranaghan et al. (2024) failed to observe systematic evidence for the common ratio effect, and concluded that previous reports of this effect (which have motivated prominent developments as Prospect Theory) might just be due to regularities in the noise, and, in particular, strength of preference.[5]

The certainty effect is hence a particularly interesting example where our model can be applied. A candidate *certainty heuristic* would be, by definition triggered by the sure option, and hence Theorem 2 applies. Thus, we should observe a discontinuity at

[5] Alós-Ferrer et al. (2024) conducted experiments with repetitions measuring response times, which allows for preference revelation following the techniques introduced in Alós-Ferrer et al. (2021), and concluded that both the common ratio and the common consequence effect are supported by empirical preference revelation analyses.

$\Delta = 0$ when analyzing data for pairs which involve a sure option (conflict and alignment), and this discontinuity should be absent when both lotteries involve risk (neutral decision problems). Also, this effect should obtain even though the model and the analysis directly control for strength-of-preference effects, which should still be reflected in the monotonicity of error rates as a function of Δ. Hence, the application is both an illustration of the use of our model to detect the presence of a heuristic process and an answer to the criticism of McGranaghan et al. (2024) on the certainty effect.

For this purpose, we now apply the model to the certainty effect and test its predictions, in particular Theorem 2, in a new experiment. The deliberative process d is assumed to be maximization of expected utility. This process is always active and prescribes to choose the option with the larger expected utility (which we will need to estimate out of sample). Strength-of-preference effects are operationalized based on the cardinal difference between certainty equivalents of the options derived from individually-estimated expected utilities (see Section 4.2.3 for details). The heuristic process h is assumed to be the certainty heuristic, which is triggered by lotteries involving certain outcomes. That is, if h is active, it prescribes to choose the lottery that gives a certain outcome.

4.2.1 Experimental Design

For our certainty experiment, we recruited a total of 128 subjects (84 females) from the student population of the University of Cologne, excluding students majoring in psychology and economics as well as subjects who had already participated in similar experiments involving lottery choice or in more than 10 experiments. The median age was 24, ranging from 19 to 69. Participants took part in 4 experimental sessions with 32 subjects each. The experiment was run at the Cologne Laboratory for Economic Research (CLER) and was programmed in PsychoPy (Peirce, 2007).

The main part of the experiment consisted of 132 pairs of binary lotteries, presented to subjects sequentially. For each pair, subjects were asked to choose one of the two lotteries. All lotteries were presented in the form of colored pie charts. The colors in the pie charts (green and blue) were counterbalanced across subjects. To control for order effects, each subject was randomly assigned to one of four different sequences of lottery pairs. There was no feedback during the course of the experiment, that is, subjects did not get any information regarding their earnings until the very end of the experiment. All decisions were made independently and at a subject's individual pace.

Before the beginning of the experiment, subjects read written instructions and answered four control questions to ensure their understanding of the concept of a lottery and its pie chart representation. Detailed instructions for the lottery choice part were handed out only after all participants had answered the control questions correctly.[6] Lottery choice was incentivized using the random incentive mechanism. That is, to de-

[6] A complete set of exemplary instructions can be found in Appendix E.

termine a participant's payoff, one of the 132 lottery pairs was randomly selected, and the lottery actually chosen by the participant was played out. The participant received the realized outcome as payoff. In addition, subjects received a show-up fee of 4 € for an average total remuneration of 14.43 €. Sessions lasted between 45 and 60 minutes including instructions and payment.

Among the 132 lottery pairs, 96 were based on the classical examples in Kahneman and Tversky (1979) with the intention to trigger the certainty effect and related phenomena. Specifically, 32 lottery pairs were designed to target the common ratio effect, and 32 the common consequence effect. A further 32 lottery pairs targeted an "almost-certainty effect" not involving any certain outcomes, which we will discuss in the Appendix. Further, we included 4 lottery pairs where one lottery first-order strictly dominated the other. Those were intended as a basic rationality check. Last, we included a disjoint set of 32 "neutral" lottery pairs which were used only to estimate the individual risk attitude. We took special care to make sure that those lotteries were not affected by any obvious bias. Subsection 4.2.3 below provides a detailed description of the procedure used to estimate individual utility functions.

The focus of our experiment was the certainty effect. We used two types of constructions. The first relies on the common ratio effect (Kahneman and Tversky, 1979), whereas the second is inspired by the common consequence effect (Allais, 1953) but adapted to lotteries with only two outcomes.

The common-ratio lottery pairs are constructed as follows. Given X, Y, p with $X > Y > 0$ and $0 < p < 1$, we define for $k \geq 1$ the pair of lotteries $S_1(k)$ and $R_1(k)$ by

$$S_1(k) = \begin{cases} Y\text{€ with prob.} & 1/k \\ 0\text{€} & \text{otherwise} \end{cases} \qquad R_1(k) = \begin{cases} X\text{€ with prob.} & p/k \\ 0\text{€} & \text{otherwise.} \end{cases}$$

Note that $R_1(k)$ is always the riskier lottery and $S_1(k)$ is always the safer lottery, in the sense that the former has a larger probability for the zero outcome. For $k = 1$, $S_1(1)$ is a sure gamble, whereas for $k > 1$ neither $S_1(k)$ nor $R_1(k)$ involve a certain outcome. Abusing notation, we can write $S_1(k) = (1/k, Y)$ and $R_1(k) = (p/k, X)$, which just specifies the probability and magnitude of the non-zero outcome. For instance, given $S_1(1) = (1, 16)$ and $R_1(1) = (0.80, 25)$ we obtain $S_1(4) = (0.25, 16)$ and $R_1(4) = (0.20, 25)$ by dividing the probability of winning by $k = 4$.

The construction of common-ratio pairs is such that an expected utility maximizer would prefer $S_1(1)$ over $R_1(1)$ if and only if she also prefers $S_1(k)$ over $R_1(k)$. Hence, any reversal contradicts expected utility theory. The traditional common ratio effect (Kahneman and Tversky, 1979) is the empirical observation that reversals where $S_1(1)$ and $R_1(k)$ (for $k > 1$) are chosen in their respective pairs are more frequent than the opposite ones.

The common-consequence lottery pairs are as follows. Given X, Y, p with $X > Y > 0$ and $0 < p < 1$, we define the pair of lotteries $S_2(\pi)$ and $R_2(\pi)$ for $\pi < 1$ by

$$S_2(\pi) = \begin{cases} Y\text{€ with prob.} & 1-\pi \\ 0\text{€} & \text{otherwise} \end{cases} \qquad R_2(\pi) = \begin{cases} X\text{€ with prob.} & p-\pi \\ 0\text{€} & \text{otherwise.} \end{cases}$$

Again, $R_2(\pi)$ is always more risky than $S_2(\pi)$, in the sense that the former has a larger probability for the zero outcome that the latter. For $\pi = 0$, $S_2(0)$ offers a sure gain, whereas for $\pi > 0$ both prospects $S_2(\pi)$ and $R_2(\pi)$ are risky. For example, given $S_2(0) = (1, 25)$ and $R_2(0) = (0.96, 28)$ we obtain $S_2(0.44) = (0.56, 25)$ and $R_2(0.44) = (0.52, 28)$ by reducing the probability of winning by $\pi = 44$ percentage points. That is, a common probability π is shifted from the non-zero outcomes to the common consequence of a zero outcome.

The common-consequence effect reflected by the Allais' paradox Allais (1953) and related phenomena (Kahneman and Tversky, 1979) uses lotteries with three outcomes to guarantee that any reversal is incompatible with expected utility theory. Our construction simplifies the lotteries, making them comparable to the ones used for common-ratio violations, but only the analogous reversals are inconsistent with expected utility theory. Specifically, if an expected utility maximizer prefers $S_2(0)$ over $R_2(0)$, then this agent must also prefer $S_2(\pi)$ over $R_2(\pi)$ for any $\pi \in [0, 1)$, as long as the Bernoulli utility of money is increasing. The argument is as follows. Let u be the Bernoulli utility of money, and assume w.l.o.g. that $u(0) = 0$. Then,

$$EU(S_2(\pi)) - EU(R_2(\pi)) = EU(S_2(0)) - EU(R_2(0)) + \pi\,(u(X) - u(Y))\,.$$

Hence, $EU(S_2(0)) > EU(R_2(0))$ implies that $EU(S_2(\pi)) > EU(R_2(\pi))$ since $u(X) - u(Y) > 0$ as $X > Y$. Thus any expected utility maximizer choosing $S_2(0)$ over $R_2(0)$ must also choose $S_2(\pi)$ over $R_2(\pi)$, and reversals where $R_2(\pi)$ is chosen instead reflect a "certainty effect."

We constructed 32 *certainty pairs* involving certain outcomes. The first 16 pairs followed the common-ratio construction above with $k = 1$, while the remaining 16 pairs followed the the common-consequence construction with $\pi = 0$. For every certainty pair, we constructed a matched *neutral pair* not involving certain outcomes to elicit possible reversals reflecting the certainty effect. Specifically, for each common-ratio pair $(S_1(1), R_1(1))$, we included a neutral pair $(S_1(k), R_1(k))$ with $k > 1$, and for each common-consequence pair $(S_2(0), R_2(0))$ we included a neutral pair $(S_2(\pi), R_2(\pi))$ with $\pi > 0$. We refer to the resulting 32 double-pairs consisting of a certainty pair and the corresponding neutral pair as "Allais-double-pairs" (see Appendix D.1 for the complete list of lottery pairs). For each Allais-double pair, observing the safe choice in the certainty pair but the riskier choice in the neutral pair contradicts expected utility maximization and suggests a certainty effect.

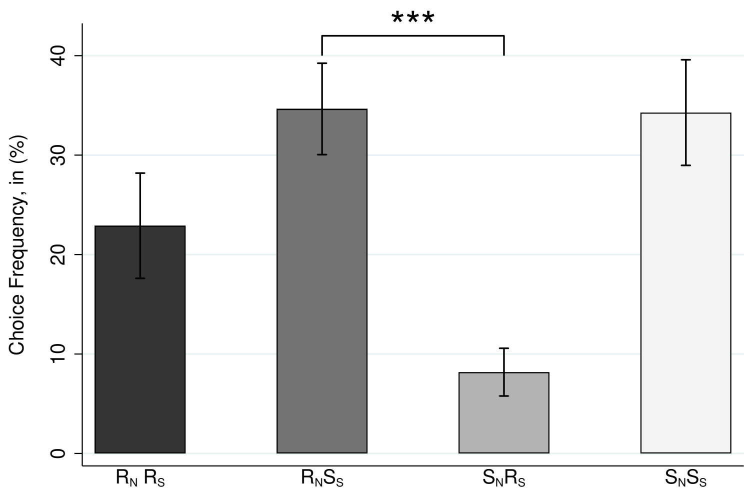


Figure 5: Choice frequencies in Allais-double pairs showing the certainty effect.

4.2.2 Establishing the Traditional Certainty Effect

For the analysis below, we will not distinguish between Allais-double-pairs obtained via the common-ratio or common-consequence constructions. Results are always similar for both types, and we present pooled analyses only for brevity.

We first explore whether our data replicates the traditional certainty effect. We remark that, following McGranaghan et al. (2024), this cannot be taken as sufficient evidence for the presence of an anomaly (recall the discussion above), and we conduct this analysis merely to show comparability with the previous literature.

For a fixed Allais-double-pair, denote the decisions in the neutral and certainty pairs by S_N vs. R_N and S_C vs. R_C, respectively. Thus, $R_N S_C$ indicates that the risky lottery was chosen in the neutral pair and the safe lottery was chosen in the certainty pair. The traditional certainty effect refers to the empirical observation that reversals of the form $R_N S_C$ (which are always violations of expected utility in our design) are frequent, and in particular more frequent than reversals of the form $S_N R_C$. Figure 5 illustrates the distribution of choices for all 128 subjects for the 32 Allais-double pairs. Violations of expected utility suggesting a certainty effect, i.e. $R_N S_C$ choice patterns, are frequent, occurring in 34.64% of all Allais-double pairs. Importantly, these violations are systematic, and in particular significantly more frequent than the opposite choice pattern $S_N R_C$ (8.18%; Wilcoxon Signed-Rank (WSR) test, $N = 128$, $z = 7.569$, $p < 0.001$). This shows that our design indeed elicited choice patterns in agreement with the well-documented certainty effect.

4.2.3 Utility estimation

The previous subsection has documented choice patterns suggesting a certainty effect in our data. Our aim, however, is to apply our model to separate cognitive imprecision (noise) as reflected by strength-of-preference effects from an actual heuristic process capturing a bias towards certainty. Data from previous experiments is ill-suited for this purpose, because, in order to account for strength-of-preference effects, we need an estimation of individual-level utility differences for each lottery pair.

For this purpose, and unlike in comparable experiments on the certainty effect, we included an additional, disjoint set of 32 lottery pairs used exclusively to estimate the risk attitude of each participant out-of-sample. That is, the experiment started with a separate phase where participants made binary choices for 32 lottery pairs different from the ones used in the Allais-double-pairs which came later. The set of lotteries used in this first phase was constructed with the aim to maximize the precision of the estimated risk attitudes. To achieve this, we relied on optimal design theory (Silvey, 1980) in the context of non linear (binary) models (Ford et al., 1992), in agreement with the recommendations of Moffatt (2015).

For utility estimation, we assume that the structure of errors follows an additive random utility model (e.g., Thurstone, 1927; Luce, 1959; McFadden, 2001). Estimation of individual risk attitudes relies on a maximum likelihood procedure (e.g., see Bellemare et al., 2008; Train, 2009) as standard in the literature (Von Gaudecker et al., 2011; Conte et al., 2011; Moffatt, 2015; Moffatt et al., 2015). See Appendix B for more details on the estimation procedure.

All $T = 32$ trials used for utility estimation involved binary choices between lotteries of the form $A = (p, x)$ and $B = (q, y)$. That is, A and B pay x with probability p and y with probability q, respectively, and 0 otherwise. We index the trials in the experiment by $t = 1, \ldots, 32$. That is, at trial t subjects faced the choice between $A_t = (p_t, x_t)$ and $B_t = (q_t, y_t)$. Further, we index the $N = 190$ subjects by $i = 1, \ldots, N$. We assume a normalized power-utility function with constant relative risk aversion (CRRA) as in Conte et al. (2011), which is given by $u(x) = \frac{x^r}{r}$. We further restrict r to be strictly positive, which is standard when lotteries include a zero outcome.

4.2.4 Main Results (Certainty Effect)

We now turn to a direct empirical test of the model. We operationalize strength-of-preference via the difference in (estimated) certainty equivalents (CE) between the safe option and the risky option. Specifically, for a given lottery pair (S, R) and decision maker i with estimated utility function u_i, we consider

$$\Delta_\omega^i = CE_i(S) - CE_i(R) = u_i^{-1}(EU_i(S)) - u_i^{-1}(EU_i(R)).$$

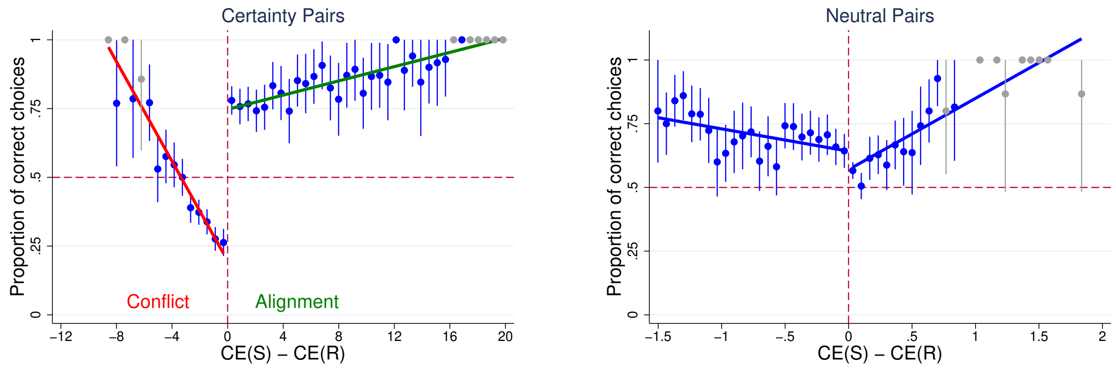


Figure 6: Proportion of correct choices as a function of $\Delta^i_\omega = CE_i(S) - CE_i(R)$ separately for certainty/DP pairs (left) and normal/neutral pairs (right).

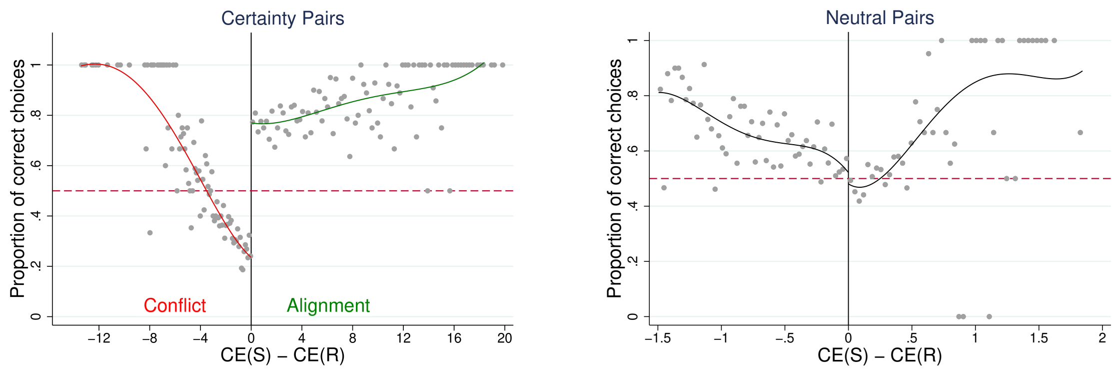


Figure 7: Proportion of correct choices as a function of $\Delta^i_\omega = CE_i(S) - CE_i(R)$ separately for certainty/DP pairs (left) and normal/neutral pairs (right).

Recall that we assume the deliberative process to be maximization of expected utility and, thus, we refer to a choice as *correct* if it maximizes (estimated) individual expected utility. More precisely, choosing S is correct for subject i if $\Delta^i_\omega \geq 0$ whereas choosing R is correct for subject i if $\Delta^i_\omega \leq 0$. Theorem 1 provides predictions about the relationship between differences in CEs, $\Delta^i_\omega = CE_i(S) - CE_i(R)$ and the proportion of correct choices assuming that maximization of expected utility is psychometric and the certainty heuristic has constant consistency.

For illustration purposes, Figure 6 depicts choice data using a binning procedure for Δ, creating bins of observations that potentially contain different trials and different subjects.[7] We then calculate the proportion of correct choices for each bin. Note that this binning procedure is used only for a graphical illustration, and neither the formal regression analyses nor the discontinuity tests below rely on it.

[7]To construct the bins, we first order all observations by the Δ_i. We then divide the range of values in 50 bins of equal width and calculate the mean, standard deviation, and 95% confidence interval for each of them. We graphically highlight bins containing less than 10 observations in gray and interpolate the plotted means using two separate linear regressions for positive and negative values of Δ.

Figure 6 clearly shows the monotonicities and the discontinuity predicted by Theorem 2. That is, as in Theorem 1, psychometric effects can be seen in the fact that the probability of a correct response increases as Δ^i_ω moves away from zero, and this regularity is observed separately for neutral decision problems and for those in either conflict or alignment. Crucially, for certainty pairs there is a clear upward jump at $\Delta^i_\omega = 0$, which separates conflict ($\Delta^i_\omega \leq 0$) and alignment ($\Delta^i_\omega \geq 0$) decision problems, with the first clearly below 50% and the second clearly above that reference quantity. This is diagnostic for the presence of a separate decision process reflecting the certainty heuristic, which lowers the proportion of correct answers in conflict and raises it in alignment. In contrast, this jump is absent for neutral pairs, where the heuristic is not triggered. These results are in agreement with the predictions of Theorem 2.

To formally test and quantify the discontinuity for certainty pairs, as commented above we rely on a local polynomial regression discontinuity estimation with robust bias-corrected confidence intervals as detailed in Calonico et al. (2014). This does not use the binning procedure and linear trends depicted in Figure 6, but rather tests for statistical differences around the predicted threshold ($\Delta = 0$).[8] The test, which is often used in regression discontinuity designs to test for the effectiveness of treatments and policies, is based on robust, non-parametric confidence intervals for the average "treatment effect" at the specified threshold conditional on mean squared error optimal bandwidths.

The procedure is illustrated in Figure 7 (left). The test detects a statistically significant upward jump of 55.53% ($z = 12.30$, $p < 0.0001$) in the proportion of correct choices at zero, with more correct choices in alignment than conflict, in line with Theorem 2(b). To ensure that this result is not a mere mechanical effect driven by an unintended aspect of the experimental design, we ran the analogous estimation for neutral pairs. In agreement with Theorem 2(c), we find no evidence for a similar jump (Figure 7, right; the difference in intercepts is only -5.818% and not significantly different from zero, $z = -1.630$, $p = 0.103$).

The magnitude of the discontinuity is computed as the difference between the intercepts as Δ approaches zero from the right and from the left. Corollary 1(a) states that this upward jump should be of size $\alpha(2P_h - 1)$. By Corollary 1(b) this constrains the parameters of the model, yielding lower bounds of $\alpha > 0.56$ and $P_h > 0.78$. To illustrate, suppose P_h was close to one, i.e. the heuristic process has large internal consistency. This implies that α is close to 0.56, i.e., in the presence of a riskless option, the certainty heuristic drives individual choices slightly more than half of the time. Analogously, suppose α was close to 1, i.e., the certainty heuristics almost always drives choices whenever a riskless option is available. This implies that P_h is close to 0.78. While we cannot identify the parameters α, P_h beyond these bounds, Corollary 1 does imply that

[8] In particular, the procedure uses a 4^{th}-degree polynomial fit to approximate the population-conditional-mean functions left and right of the 0. It employs a kernel function to construct the global polynomial estimators and a data-driven procedure to select the number of bins which are created to compare observations on the left and on the right of the critical threshold. We use the default options of this procedure, implemented by the command *rdrobust* in STATA.

$\alpha(2P_h - 1) \simeq 0.56$. That is, there is a tradeoff in the sense that to explain the data with a larger α, a smaller P_h is required and vice versa.

Corollary 1(c) offers another opportunity to validate the model. Specifically, it predicts that the jump at $\Delta = 0$ must be symmetrically distributed between the left- and the right-hand parts. Following the procedure described above (Calonico et al., 2014) allows to derive the intercepts (Figure 7, left), i.e. the limits of the probabilities of correct responses as Δ approaches zero, conditional on conflict ($\Delta < 0$) or alignment ($\Delta > 0$). Those are $P_- = 0.2354$ and $P_+ = 0.7700$, respectively. As predicted by Corollary 1(c), $(1/2) - P_- = 0.2646 \simeq 0.23 = P_+ - (1/2)$.

Last, we conduct random effects panel Probit regressions on the probability of a correct answer, which are reported in Table 5. Models 1-3 consider only alignment, conflict, or neutral decision problems, respectively, with the absolute difference in certainty equivalents ($|\Delta^i_\omega|$) as an independent variable. For all models we find highly significant strength-of-preference effects, that is, the probability of a correct answer is increasing in the absolute difference in certainty equivalents, separately for neutral decisions and those in either conflict or alignment. This shows that the effects of cognitive imprecision, as predicted in Theorems 1 and 2, are readily observable.

Model 4 considers all decision problems with $\Delta^i_\omega > 0$, which includes both decision problems in alignment and neutral decision problems where the deliberative response should be S. The dummy for alignment decision problems is significantly positive, showing that the certainty heuristic shifts up the probability of a correct response in this case. Conversely, Model 5 considers all decision problems with $\Delta^i_\omega < 0$, which includes both decision problems in conflict and neutral decision problems where the deliberative response should be R. The dummy for conflict decision problems is significantly negative, showing that the certainty heuristic shifts down the probability of a correct response in this case.

Model 6 considers all certainty pairs, i.e. all decision problems in either conflict or alignment. The regression controls for a dummy for $\Delta^i_\omega > 0$, which in this case corresponds to decision problems in alignment. The coefficient for $|\Delta^i_\omega|$ is significantly positive, which again shows the strength-of-preference effect for decisions under conflict. The linear combination test including the interaction between the dummy and $|\Delta^i_\omega|$ is significant and positive, showing the strength-of-preference effect for decision problems under alignment. The dummy itself is significantly positive, reflecting the upper jump at zero. However, we remark that this is not the correct way of testing for a discontinuity, as it merely reflects the difference in intersect points of linear trends estimated above and below zero, and might be misleading if the data displays a nonlinear relation to Δ. The correct way to test for a discontinuity is the procedure we described above (Calonico et al., 2014).

Model 7 is the analogue of Model 6 but considering only neutral decision problems. That is, there is neither conflict nor alignment. The coefficient for $|\Delta^i_\omega|$ is significantly positive, showing strength-of-preference effect for decisions where the deliberative pro-

Table 5: Random effects Probit regressions on correct answers.

Correct	Alignment (1)	Conflict (2)	Neutral (3)	$\Delta^i_\omega > 0$ (4)	$\Delta^i_\omega < 0$ (5)	Certainty (6)	Neutral (7)
$\lvert\Delta^i_\omega\rvert$	0.083***	0.210***	0.423***			0.171***	0.250***
	(0.019)	(0.029)	(0.051)			(0.025)	(0.058)
Alignment				0.818***			
				(0.053)			
Conflict					−1.147***		
					(0.047)		
$\Delta^i_\omega > 0$ (dummy)						1.539***	−0.359***
						(0.079)	(0.066)
$\lvert\Delta^i_\omega\rvert \times (\Delta^i_\omega > 0)$						−0.115***	0.586***
						(0.001)	(0.154)
Constant	1.537***	−1.106***	0.283***	0.105	0.648***	−0.690***	0.492***
	(0.241)	(0.170)	(0.056)	(0.087)	(0.104)	(0.096)	(0.070)
Linear Combination							
$\lvert\Delta^i_\omega\rvert + \lvert\Delta^i_\omega\rvert \times (\Delta^i_\omega > 0)$						0.056***	0.837***
						(0.013)	(0.142)
Observations	1614	2482	4096	3565	4627	4096	4096

Notes: Robust standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

cess favors S. The linear combination test including the interaction between the sign dummy and $\left|\Delta^i_\omega\right|$ is significant and positive, showing the strength-of-preference effect for decisions where the deliberative process favors S. The dummy itself is also significantly positive, but, as commented above, this just reflects a difference in intersect points of linear trends estimated above and below zero, and not an actual discontinuity (compare the right-hand sides of Figures 6 and 7).

5 Response Times

The recent literature in economics has recognized the value of response times as an additional source of data which can be used to reveal latent variables and processes in economic decision making. On the one hand, it is a well-established fact that strength-of-preference effects also affect response times. Specifically, response times are longer when the alternatives are more similar, i.e. more difficult to discriminate (Cattell, 1893; Dashiell, 1937; Moyer and Landauer, 1967; Laming, 1985; Klein, 2001; Wichmann and Hill, 2001; Palmer et al., 2005). This *chronometric effect* has also been shown for preference-based economic tasks, i.e. decisions are slower when decision makers are closer to indifference (Chabris et al., 2009; Alós-Ferrer and Garagnani, 2022a,b). Alós-Ferrer et al. (2021) showed that this property can be used to reveal preferences without distributional assumptions on noise terms, and Alós-Ferrer and Garagnani (2024b) demonstrated that preferences revealed through response times improve out-of-sample prediction compared to standard econometric estimations.

On the other hand, choice models based on the interaction of different processes in the human brain are built upon the assumption that more heuristic or intuitive processes are faster than more deliberative ones (Sloman, 1996; Kahneman, 2003; Evans, 2008;

Weber and Johnson, 2009). A number of works in economics have built upon this extensive literature from psychology to understand the role of intuition and impulses in decision making (e.g., Rubinstein, 2007, 2016; Gennaioli and Shleifer, 2010; Cappelen et al., 2016; Heller et al., 2017; Spiliopoulos and Ortmann, 2018). For example, Achtziger and Alós-Ferrer (2014) derived testable predictions for a dual-process model applied to belief-updating tasks, and Alós-Ferrer and Ritschel (2021) used a similar approach to analyze the interplay of imitation and best-reply in Cournot oligopolies.

Since we consider both cognitive imprecision and process multiplicity, response times in our model and applications should reflect both a chronometric effect linked to strength of preference and speed differences across processes. However, these effects can interact and create confounds. As Krajbich et al. (2015) pointed out, studies comparing response times across individuals might mistakenly identify dual-process effects even in the absence of process multiplicity. For example, a group of people might make fast decisions in, say, a binary dictator game, because they have strongly prosocial preferences and they do not struggle when evaluating a fair allocation compared to a selfish decision. A different group of people might make slow decisions in the same situation because their social preferences are weaker, hence the decision is a closer one for them. This just describes a chronometric effect, but an analyst might be led to believe that the first group is faster because prosociality is intuitive, while the second group is slower because selfishness is more deliberative. This is a confound (see also Myrseth and Wollbrant, 2016), which highlights the need for well-specified models distinguishing both effects.

Therefore, we now add response times to the model, which, as choices, are assumed to be stochastic. Let $T_b(\omega) = E[RT \mid b, \omega]$ denote the *expected response time* in decision problem $\omega \in \Omega_b$ conditional on the response being selected by a given process b. To keep the model tractable, we assume that, for a given decision problem, expected response times do not depend on the actually-selected response, but only on which process determined the response.

First, to incorporate the possible chronometric effects of cognitive imprecision, we consider the analogous concept to psychometric processes (Definition 2). In words, a process is chronometric if it reflects the standard relation between choice difficulty and response times, with more difficult decisions being slower.

Definition 3. A process b is *chronometric* if $T_b(\omega)$ is a strictly decreasing function of $|\Delta_\omega|$.

Second, to capture the distinguishing features of deliberative and heuristic processes, it is useful to have a notion of when a process is "faster" than another one.

Definition 4. Process b_1 is *faster* than process b_2 on a set of decision problems $\Omega_{b_1 b_2} \subseteq \Omega_{b_1} \cap \Omega_{b_2}$ if $T_{b_1}(\omega) < T_{b_2}(\omega)$ for all $\omega \in \Omega_{b_1 b_2}$.

For instance, if d is a deliberative process with $\Omega_d = \Omega$ and h is a heuristic process with $\Omega_h \subsetneq \Omega$, the standard assumption in a dual-process model is that h is faster than

d on Ω_h, i.e. whenever h is active. It is, however, important to note that this standard assumption concerns the *unobservable* response times of cognitive processes. What the analyst actually observes are the response times conditional on either decision problems or responses. Since processes are stochastic, the analyst does not know which process generated which response and cannot infer whether a decision is, say, "more intuitive" just from its response time. As we will see below, however, within the context of our model, assumptions on the relative speed of processes do yield testable predictions on observable response times.

For all results below, $E[RT(\omega)]$ denotes the expected *observable* response time for a decision problem ω, i.e. the response time of the actual decision, and in particular without conditioning on a process. That is,

$$E[RT(\omega)] = \begin{cases} (1-\alpha)T_d(\omega) + \alpha T_h(\omega) & \text{if } \omega \notin \Omega_h^N \\ T_d(\omega) & \text{if } \omega \in \Omega_h^N. \end{cases}$$

5.1 Slower Decisions Due to Process Conflict

Our first testable prediction parallels a well-known effect from psychology known as the *Stroop Effect* (Stroop, 1935). This can be illustrated with the Stroop task, a widely-used paradigm where participants are asked to verbally report the color in which certain words are printed. Some of those words, however, name colors themselves. For instance, a stimulus might be the word "blue" printed in blue, while another stimulus might be the same word "blue" but printed in red. This classical task illustrates the conflict between the process to read a word, which is rather automatic and fast, and the slower, more deliberative process to identify the font color, recover it from memory, and name it. The Stroop effect, which is one of the most robust effects in the psychological literature, refers to the fact that correct responses in conflict trials (e.g., "blue" printed in red) are systematically slower than those in alignment trials (e.g., "blue" printed in blue). This is typically interpreted as resulting from the need to inhibit an automatic response (just reading the word) in case of conflict.

The following result demonstrates that this effect generalizes to any situation where a heuristic process interacts with a deliberative one, and shows how to detect it even when response times are affected by chronometric effects. The result only requires that the heuristic process is faster than the deliberative one. In particular, it does not require an assumption on the relative consistency of processes or an additional "slowdown" due to the resolution of a conflict between processes, as typically assumed in psychology.

Theorem 3 (Generalized Stroop effect)**.** *Consider a set of decision problems Ω. Suppose T_d and T_h are functions of $|\Delta_\omega|$, d is psychometric, h is faster than d on Ω_h, and $P_h(\omega) > 1/2$ on Ω_h. Let ω^A in Ω_h^A (alignment) and ω^C in Ω_h^C (conflict). Then:*

(a) *If* $\Delta_{\omega^A} = \Delta_{\omega^C}$, *then*

$$E[RT(\omega^A) \mid \rho(\omega^A) = d(\omega^A)] < E[RT(\omega^C) \mid \rho(\omega^C) = d(\omega^C)],$$

that is, correct answers in alignment are faster than correct answers in comparable conflict situations.

(b) *Additionally, suppose* d *is symmetric in the sense that* $P_d(\omega) = P_d(\omega')$ *if* $|\Delta_{\omega'}| = |\Delta_\omega|$*. Then, if* $|\Delta_{\omega^A}| = |\Delta_{\omega^C}|$*, the property in (a) also holds.*

(c) *If* d *is symmetric and* $\omega, \omega' \in \Omega_h^N$ *are neutral decision problems with* $|\Delta_{\omega'}| = |\Delta_\omega|$*, then* $E[RT(\omega) \mid \rho(\omega) = d(\omega)] = E[RT(\omega') \mid \rho(\omega') = d(\omega')]$.

Note that the theorem does not actually assume that d is chronometric, because the results are for fixed values of Δ. Parts (a) and (b) are the Stroop effect. Part (b) is needed to spell out this effect when the heuristic is triggered by an action, and hence conflict and alignment decision problems have different signs of Δ. Part (c) merely states that, even in that case, the analogous statement of the Stroop effect should be absent for neutral decision problems.

Note that dual-process models from psychology typically impose stronger assumptions (and considerably more parametric structure) than Theorem 3. For instance, to capture conflict detection and resolution, one might assume an additional non-decision time for decision problems under conflict (MacCleod, 1991). Since this imposes an additional slowdown for conflict compared to alignment, it would just make the effect in the result above stronger. Additionally, note that we do not impose that h is more consistent than d, even though this is a standard property of heuristic processes (Redish, 2016). Last, note that the result does not need that h exhibits chronometric effects on Δ.

Theorem 3 provides a simple indicator for the presence of dual processes. Sections 5.3 and 5.4 below provide statistical tests for the predictions in this theorem in our two experiments, which will take the form of regression analyses as this allows us to show that the regularities identified in the theorem are not confounded by strength of preference effects. However, we can already provide a simple graphical illustration of both Theorem 3 and its empirical support in the two applications (conservatism and certainty experiments). This is shown in Figure 8. The top-left panel depicts the area where data should be expected to be if one plots the differences in response times (conflict minus alignment) for correct responses as a function of Δ. For this purpose, one must identify decision problems with the same value of Δ. For the conservatism experiment, Δ is the difference in expected payoffs between the urns. Recall that, in that experiment, there were 8 different values of Δ for conflict decision problems, and 12 for alignment decision problems. Of those, there are 6 values of Δ where we have decision problems both in conflict and in alignment, and hence we can compute $E[RT(\omega^C) \mid \rho(\omega^C) = d(\omega^C)] - E[RT(\omega^A) \mid \rho(\omega^A) = d(\omega^A)]$ with $\Delta_{\omega^A} = \Delta_{\omega^C}$. The

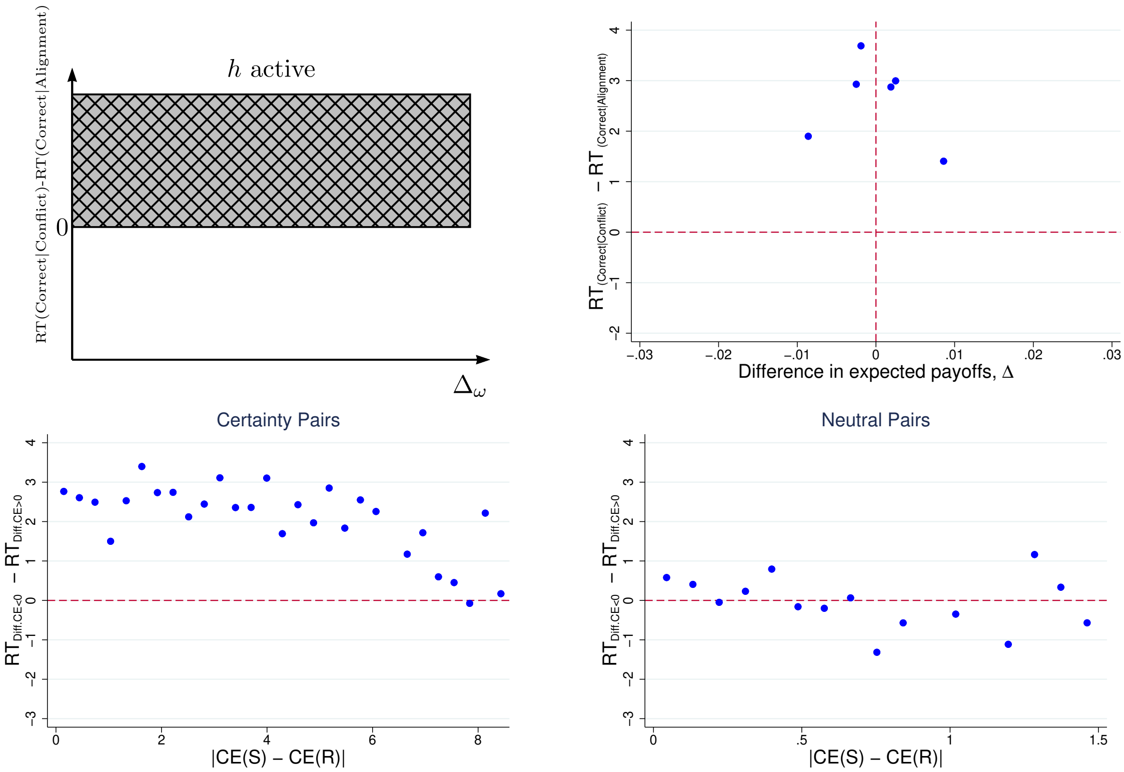


Figure 8: Top-left: Illustration of Theorem 3. $RT^C(\text{correct}) - RT^A(\text{correct})$ should be positive when comparing decision problems in conflict and alignment while keeping Δ constant, for any value of Δ. Top-right: the predicted effect is confirmed in the conservatism experiment. Bottom-left: the predicted effect is also confirmed in the certainty experiment. Here the comparison is for certainty pairs between negative CE differences (conflict; $\Delta^i_\omega < 0$) and positive CE differences (alignment; $\Delta^i_\omega < 0$), as a function of the binned distance in CEs ($|\Delta^i_\omega|$). Bottom-right: the effect is absent when conducting the analogous comparison for neutral pairs, where the sign of Δ^i_ω does not indicate conflict or alignment.

top-right panel of Figure 8 depicts those differences in the data, which are all positive as predicted by Theorem 3(a).

The bottom-left panel of Figure 8 shows the empirical support for Theorem 3 in the certainty experiment. In this experiment, recall that Δ is the difference in certainty equivalents between the lotteries, and that positive values (resp. negative) correspond to choice pairs in alignment (resp. conflict). Thus, to illustrate the result, we identify certainty pairs in conflict and alignment with similar values of Δ but opposite sign (part (b) or the result). Specifically, we follow a binning procedure analogous to the one described for Figure 6 above, but now applied to $|\Delta^i_\omega|$. The panel then plots the difference between the average response time of correct answers in alignment ($\Delta^i_\omega > 0$) and conflict ($\Delta^i_\omega < 0$) for each bin. As can be seen in the figure, the differences in response times $E[RT(\omega^C) \mid \rho(\omega^C) = d(\omega^C)] - E[RT(\omega^A) \mid \rho(\omega^A) = d(\omega^A)]$ are positive for all but one bin, confirming the prediction of Theorem 3(b).

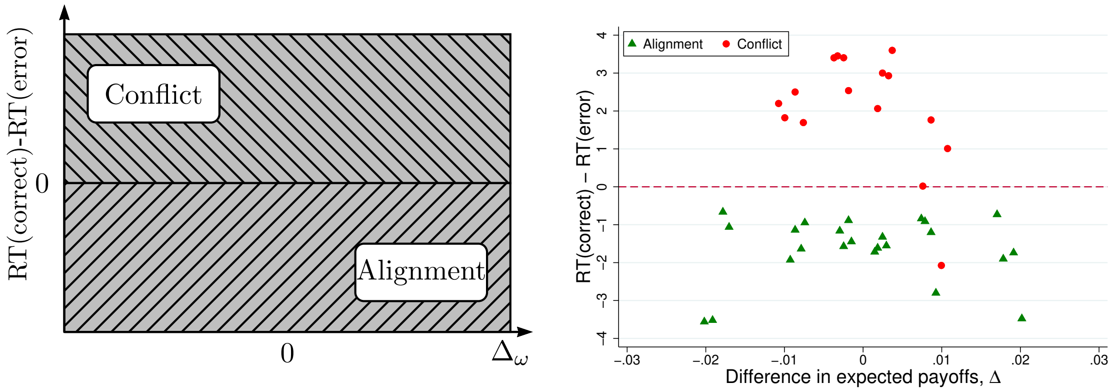


Figure 9: Left: Illustration of Theorem 4. $RT(\text{correct}) - RT(\text{error})$ should be positive for decision problems in conflict and negative for those in alignment, for any value of Δ. Right: the predicted effect is confirmed in the conservatism experiment.

To ensure that these results are not a mere mechanical effect driven by some other aspect of the experimental design, the bottom-right panel of Figure 8 plots the analogous graph for neutral pairs, for which Theorem 3(a,b) does not apply and the sign of Δ_ω does not identify conflict or alignment. Here, the difference has different sign for different bins, but is generally close to zero, reflecting Theorem 3(c), in stark contrast to the observed result for certainty pairs, for which part (b) of the theorem applies.

5.2 The Relative Speed of Errors

Our second testable prediction concerns the relative speed of different responses, i.e. whether errors are faster or slower than correct responses for a fixed decision problem. There is a long discussion about the relative speed of errors in psychology (e.g., Laming, 1985; Luce, 1986; Alós-Ferrer and Garagnani, 2024a). The following result shows that the answer depends on whether decision problems are in conflict or in alignment.

Theorem 4. *Consider a set of decision problems Ω. If h is faster than d and $P_h(\omega) > \frac{1}{2}$ on Ω_h, then the following statements hold.*

(a) For any $\omega \in \Omega_h^C$ (conflict), errors are faster than correct answers.

(b) Suppose h is more consistent than d. For any $\omega \in \Omega_h^A$ (alignment), errors are slower than correct answers.

Intuitively, for decision problems in alignment, errors are slower than correct answers because h is more consistent than d, implying that relatively more errors result from d than from h. Since d is slower than h, many errors are relatively slow and many correct responses are relatively fast, and the result follows. Analogously, for decision problems in conflict, errors are faster than correct answers because relatively more errors result from the faster process h.

Theorem 4 provides a second simple indicator for the presence of dual processes. Again, Sections 5.3 and 5.4 below provide statistical tests for the predictions in this

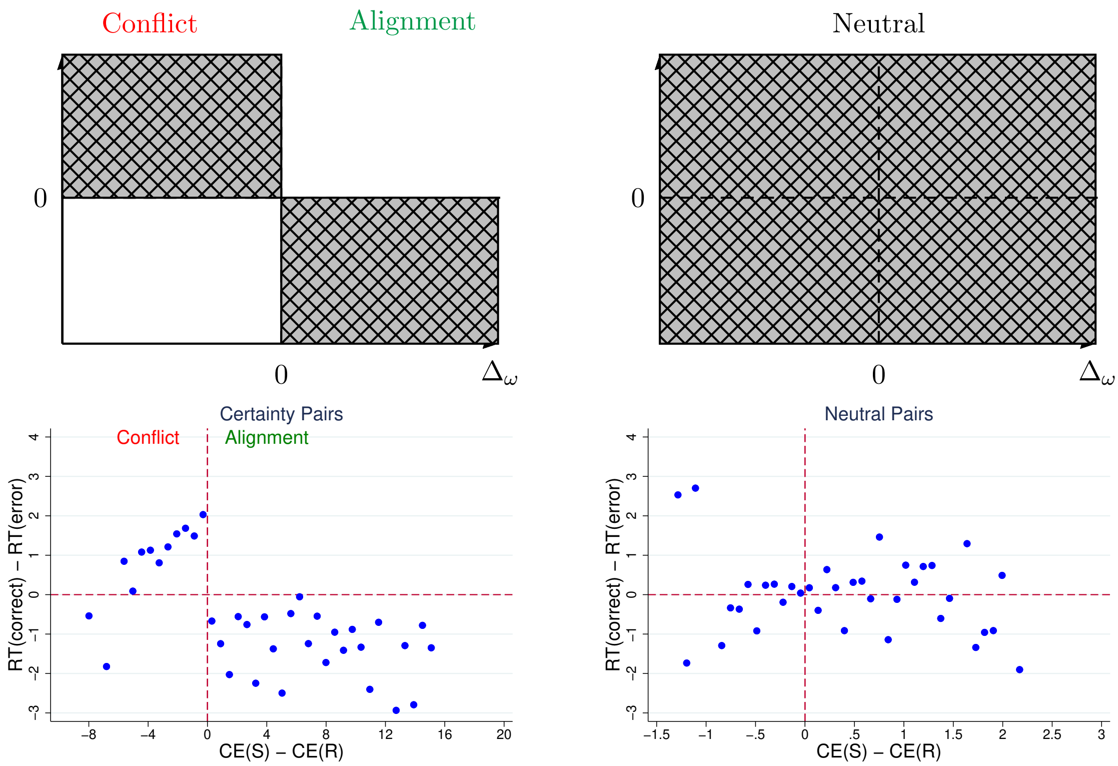


Figure 10: Top: Illustration of Theorem 4 for heuristics triggered by an alternative. $RT(\text{correct}) - RT(\text{error})$ as a function of Δ_ω have different signs depending on conflict and alignment (top left) but are unrestricted for neutral pairs (top right). Bottom: Average difference in response times between correct choices and errors in the certainty experiment, as a function of $\Delta^i_\omega = CE_i(S) - CE_i(R)$ (binned), separately for certainty pairs (left) and neutral pairs (right). The predicted effect is confirmed for certainty pairs, and is absent for neutral pairs, for which the result does not apply.

theorem in our two experiments, using regressions to ensure that the results are not confounded by strength of preference effects. Before turning to the statistical analyses, however, we illustrate the predictions and the empirical support in our two experiments in Figures 9 and 10. Figure 9(left) illustrates that, when a heuristic is active, the difference in average response times between correct choices and errors should be positive in conflict and negative in alignment. The right panel shows that this effect is overwhelmingly present for decision problems in our conservatism experiment. That is, for each decision problem in that experiment, we compute the difference in response times between correct responses and errors and plot it as a function of Δ. We indeed find a sharp separation, with all but one of the differences in case of conflict in the upper half and all differences in case of alignment in the lower half of the figure.

Figure 10 illustrates Theorem 4 and its empirical support when the heuristic is triggered by an alternative. In this case, all decision problems in conflict (resp. alignment) have negative (resp. positive) values of Δ. Thus, plotting the data should result in a pattern as in the top-left panel of the figure, while for neutral decision problems the pattern is unrestricted (top-right panel). The bottom panels of the figure show the empirical support for these predictions in our certainty experiment. Using the same binning procedure as in Figure 6, the panels plot the difference between the average response time of correct answers and errors against the average difference in certainty equivalents for each bin.[9] In case of conflict, which corresponds to bins with $\Delta_\omega^i < 0$, this difference is positive for almost all bins, indicating that indeed errors are faster than correct answers. In stark contrast, for alignment situations, that is, bins with $\Delta_\omega^i > 0$, the difference is negative for all bins, showing that errors are slower than correct answers. To ensure that this asymmetry in response time differences is evidence for process multiplicity and cannot be explained by other elements of the experimental design, the bottom-right panel of Figure 10 shows the analogous graph for neutral pairs. For those pairs, the certainty heuristic is absent, hence no effects of process multiplicity should be present (and the sign of Δ_ω does not indicate conflict or alignment). Indeed, as expected, we no longer observe an asymmetric pattern of response time differences in this case. Instead, the differences are spread across all four quadrants.

A version of the response-time predictions in Theorems 3 and 4 was provided in Achtziger and Alós-Ferrer (2014), which discussed a dual-process model and applied it to an experiment on belief-updating where Bayesian updating potentially conflicted with reinforcement learning. The model presented there, called the Dual-Process Diffusion Model (DPDM), was given a psychological microfoundation in Alós-Ferrer (2018b), where each of the decision processes was assumed to follow a Drift-Diffusion model (Ratcliff, 1978, 1981; Fudenberg et al., 2018). The DPDM was also applied to experimental data on Cournot oligopolies in Alós-Ferrer and Ritschel (2021), where the potentially-conflicting processes were myopic best reply and imitation, and to repeated constant-sum games in Spiliopoulos (2018), where the conflict was between simple reinforcement and a more complex version of reinforcement involving detecting patterns in the opponent's play. Recently, Alós-Ferrer and Garagnani (2024a) have reanalyzed 31 experiments from publications in cognitive psychology and demonstrated that the predictions of the DPDM are supported in a large variety of paradigms involving potential conflict between processes. The difference with Theorems 3 and 4 as presented here is that those works concentrated on the interaction between processes but ignored possible strength-of-preference effects. In other words, the results presented here show that the response-times predictions of the DPDM extend to our model and could not be explained by potential confounds with strength of preference.

5.3 Response Time Analysis (Conservatism)

Figures 8, 9, and 10 have illustrated the empirical support for Theorems 3 and 4 in our two experiments. Those figures already avoid possible confounds with strength of preference (Krajbich et al., 2015), since they control for the magnitude of Δ. To provide statistical tests of the predictions while taking into account both strength-of-preference

[9] Bins which did not contain both errors and correct choices are excluded from the graph.

Table 6: Random effects regression on logRTs for the conservatism experiment, for decision problems where h is active.

logRT	Non-Neutral problems			Neutral problems
	(1)	(2)	(3)	(4)
$\|\Delta\|$	-18.246***		-17.997***	-33.319***
	(0.599)		(0.593)	(1.785)
Conflict	0.648***	0.831***	0.831***	
	(0.012)	(0.008)	(0.013)	
$\|\Delta\| \times$ Conflict	1.665		-10.491***	
	(2.114)		(2.114)	
Error		0.153***	0.100***	0.062*
		(0.019)	(0.019)	(0.034)
Error $\times$ Conflict		-0.582***	-0.564***	
		(0.023)	(0.023)	
Constant	0.459***	0.315***	0.453***	0.647***
	(0.024)	(0.023)	(0.024)	(0.028)
$\|\Delta\| + \|\Delta\| \times$ Conflict	-16.581***		-28.488***	
	(2.027)		(2.030)	
Error + Error $\times$ Conflict		-0.429***	-0.464***	
		(0.014)	(0.014)	
Observations	39840	39840	39840	6799

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

and dual-process effects, we now report linear regressions for the response times in the experiments.

We start with the conservatism experiment. We ran linear random-effects panel regressions on log-transformed response times with independent variables $|\Delta|$ (difficulty), and dummy variables for Conflict and Errors. Table 6 reports the results. Regression models 1 to 3 consider all decision problems where the heuristic is active, i.e. those under conflict or alignment. Model 1 focuses on strength of preference, i.e., the effect of $|\Delta|$, Model 2 concentrates on dual-process effects, and Model 3 considers both. In Model 1, the coefficient for $|\Delta|$ is significantly negative, demonstrating the chronometric effect, that is, response times are longer for larger values of $|\Delta|$. Since the regression includes a conflict dummy, the effect captured by the coefficient for $|\Delta|$ corresponds to decisions in alignment. To show that the chronometric effect is also present for decisions in conflict, we examine the linear combination test $|\Delta| +$ Conflict $\times$ $|\Delta|$, which is also significantly negative. These effects also hold in Model 3, which controls for possible dual-process effects.

Models 2 and 3 include dummies for conflict and errors, hence the reference category are correct decisions under alignment. In Model 2, the coefficient for the conflict dummy is significantly positive, showing that correct decisions under conflict are slower than correct decisions under alignment, as predicted by Theorem 3 (the generalized Stroop effect). This effect is robust to controlling for strength-of-preference effects (Model 3). The coefficient for the error dummy is also significantly positive in both models, showing that, in alignment, errors are slower than correct responses, as predicted by Theorem 4(b), and this effect persists when controlling for chronometric effects. The interaction of conflict and error is significantly negative, showing that this effect changes when we switch attention to decisions under conflict, as expected from Theorem 4(a). The linear combination test Error + Conflict × Error is significantly negative in Model 2, in alignment with both Theorem 4(a) and Figure 9. This effect persists in Model 3, where we control for strength of preference.

Model 4 considers neutral decisions (uniform prior), where the heuristic is not active and all decisions should arise from the deliberative process (that is, there is neither conflict nor alignment). Again, and as expected, the coefficient for $|\Delta|$ is significantly negative, demonstrating the chronometric effect also for neutral decision problems. The coefficient for the dummy error is marginally positively significant (although also smaller in magnitude compared to problems in conflict or alignment), suggesting that errors arising from the deliberative process might be slightly slower than correct responses.

5.4 Response Time Analysis (Certainty Effect)

We now turn to the certainty experiment. We ran the analogous linear random-effects regressions on log-transformed response times. We again include the independent variable $|\Delta|$ to capture strength of preference and a dummy variable for errors. Since in this experiment decision problems in alignment and conflict correspond to positive and negative values of Δ, respectively, we include a dummy for $\Delta > 0$.

Table 7 reports the results. Models 1-3 analyze response times for certainty pairs, which are either in conflict or alignment. Models 4-6 are the corresponding analyses for neutral pairs, where the sign of Δ does not correspond to conflict or alignment.

Models 1 and 4 concentrate on possible strength-of-preference (chronometric) effects. These are clear for neutral pairs (Model 4). Both the coefficient for $|\Delta|$ and the linear combination test $|\Delta| + |\Delta| \times (\Delta > 0)$ are significantly negative, indicating that decisions further away from indifference (on either side) were made faster. However, in Model 1, neither the coefficient for $|\Delta|$ nor the linear combination test are significant, meaning that the chronometric effect is too weak to be detected in pairs where the certainty heuristic is active, both for conflict and alignment pairs. A possible interpretation is that, for neutral pairs, all decisions are made by the (chronometric) deliberative process, hence chronometric effects dominate. In contrast, for certainty pairs, many decisions are made by the heuristic process, which is triggered by the presence of a riskless option and hence is not chronometric on Δ.

Models 2 and 5 concentrate on possible dual-process effects. For certainty pairs, in Model 2 the dummy for $\Delta > 0$ (alignment) is significant and negative, confirming the prediction of Theorem 3 (generalized Stroop effect). That is, correct decisions in conflict are slower than correct decisions in alignment. The dummy for errors is significantly negative, showing that errors in conflict are faster than correct responses, as predicted by Theorem 4(a). In contrast, the linear combination test Error+Error $\times(\Delta > 0)$ is significantly positive, showing that errors in alignment are slower than correct responses, as predicted by Theorem 4(b). Both the Stroop effect and the asymmetry in the relative speed of errors are absent for neutral pairs (Model 5), which should not involve a heuristic. We observe that the dummy for errors and the corresponding linear combination test are both significantly positive. This suggests that the errors of the deliberative process are in general slower than correct responses.

Models 3 and 6 put together chronometric and dual-process effects. For certainty pairs, Model 3 shows that all effects detected in Model 2 are robust to controlling for strength of preference, i.e. the predictions of Theorems 3 and 4 hold even when controlling for possible chronometric regularities. For neutral pairs, Model 6 shows that the chronometric effects detected in Model 4 are robust to controlling for asymmetries arising for the sign of Δ. However, in this model errors are not significantly slower than correct responses for positive Δ.

Overall, the analyses in this and the previous subsection show that our experiments support the predictions of theorems 3 and 4, which in turn suggest that in both cases decisions are the result of the interaction of two cognitively-different processes. Our analyses also show that these effects subsist when controlling for possible strength-of-preference effects, i.e., the results suggest multiple processes while avoiding possible confounds with strength of preference.

6 Extensions and Additional Analyses

6.1 Generalization I: Cognitive Control Costs

For simplicity, the model presented above considers a single parameter $\alpha \in (0, 1)$ to capture the probability of process selection whenever both processes are active. Adopting a dual-process view, $1 - \alpha$ is the *inhibition probability*, that is, the probability that the decision maker will be able to inhibit an automatic or intuitive reaction (following process h) and make a decision following process d instead. The view in psychology is that process selection is a *central executive* function, associated to cognitive functions and consuming cognitive resources as working memory (Baddeley, 1992). In particular, inhibition probability should be affected by the availability of cognitive resources (cognitive load, time pressure, etc.) and also by motivational and attentional factors.

A particularly interesting extension of the model is to generalize the parameter α to allow inhibition probability to depend on the importance and relevance of the decision. Decision makers pay more attention when the possible consequences of a decision are more substantial. Thus, it is natural to assume that the probability that the heuristic process h will take over is smaller for larger values of $|\Delta|$, which measures the cardinal

Table 7: Random effect regression on log-transformed response times.

Log.RT	Certainty pairs (1)	(2)	(3)	Neutral pairs (4)	(5)	(6)
$\|\Delta\|$	0.027		0.000	-0.766**		-0.664**
	(0.032)		(0.032)	(0.318)		(0.321)
$(\Delta > 0)$ (dummy)	-0.018	-0.757***	-0.699***	-0.243	-0.133	-0.125
	(0.113)	(0.135)	(0.158)	(0.170)	(0.192)	(0.210)
$\|\Delta\| \times (\Delta > 0)$	-0.048		-0.017	0.531		0.440
	(0.036)		(0.036)	(0.336)		(0.339)
Error		-0.723***	-0.718***		0.482***	0.433**
		(0.119)	(0.120)		(0.180)	(0.182)
Error $\times(\Delta > 0)$		1.203***	1.176***		-0.170	-0.141
		(0.225)	(0.227)		(0.270)	(0.271)
Constant	2.629***	3.148***	3.158***	4.906***	4.538***	4.703***
	(0.136)	(0.136)	(0.156)	(0.191)	(0.200)	(0.211)
Linear combination tests						
$\|\Delta\| + \|\Delta\| \times (\Delta > 0)$	-0.021		-0.017	-0.234**		-0.224**
	(0.015)		(0.015)	(0109)		(0.109)
Error+Error $\times(\Delta > 0)$		0.480***	0.458***		0.312*	0.292
		(0.171)	(0.172)		(0.189)	(0.189)
Observations	4096	4096	4096	4096	4096	4096

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

scope of the consequences of the decision (opportunity costs). Specifically, assume that $\alpha = \alpha(|\Delta|)$, where $\alpha(\cdot)$ is a continuous and strictly decreasing function of $|\Delta|$ taking values in $(0, 1)$. For instance, this would be the result of a model where cognitive control costs are stochastic and the heuristic process is inhibited only if the costs are lower than the potential gains $|\Delta|$.

This approach is related to Benhabib and Bisin (2005), who postulated a neuroeconomic model applied to intertemporal consumption decisions where a supervisory attention system overrode automatic processing whenever the utility difference between the actions chosen by deliberative and automatic processes exceeded a fixed parameter measuring attention costs. The difference is that here the utility difference is measured by Δ and not by the anticipated choices of both decision processes, and that the relation is assumed to be gradual rather than binary. Also, this extension implies that heuristic processes will be followed more often, and the biases resulting from them will occur more frequently, when decision makers are less sure of their preferences. Enke and Graeber (2023) find evidence for this implication, showing that a number of biases are more prominent when people are more cognitively uncertain about their decisions.

We refer to this model as the *model with cognitive control costs.* Almost all our results hold for this more general case (Appendix A contains the sketched proofs for each result when needed, after the corresponding proofs for the results of the main model). All predictions in Theorem 1 hold, except the prediction of psychometric effects for decision problems in alignment. The intuition is as follows. In extreme cases, the heuristic process might be very internally consistent, with P_h close to one. Under alignment, this means a very low error rate. Hence, if Δ becomes larger and the probability of inhibition increases, this might be counterproductive as the deliberative process is selected more often but has a larger error rate. This might partially counteract the positive, psychometric effect resulting from a lower probability of error conditional on the deliberative process being selected. In contrast, under conflict, an increase in Δ is never counterproductive, as it both increases the probability of a correct response for the deliberative process and the chances of the heuristic process (which under conflict typically produces errors) being selected.

All predictions in Theorem 2 also hold, with the exception of part (a) in case of alignment, for the same reason as above. In particular, the discontinuity in the presence of multiple decision processes is also a prediction of the model with cognitive control costs. Corollary 1 also holds, replacing α with $\alpha(0)$.

Theorems 3 and 4 hold without change, as they concern either comparisons of decision problems with the same $|\Delta|$ (hence the same α) or comparisons involving a fixed decision problem (hence α is just replaced by $\alpha(|\Delta|)$).

6.2 Generalization II: Slow Process Errors

A large literature in psychology has discussed the issue of whether, in simple cognitive tasks, errors are faster or slower than correct responses (e.g., Laming, 1968; Luce, 1986; Ratcliff and Rouder, 1998, 2000; Ratcliff et al., 2004). Theorem 4 provides an answer to this question, namely that errors are slower when underlying decision processes are in alignment and faster when they are in conflict. This is consistent with previous evidence from psychology. For example, White et al. (2011) and Mulder et al. (2012) showed that errors tend to be slower than correct responses in certain simple tasks (the Flanker task and a random-dot motion taks, respectively) when trials are "congruent," meaning that two underlying processes favor the same response, but the opposite is true in incongruent trials. Alós-Ferrer and Garagnani (2024a) have found evidence for this asymmetry in a large number of datasets from cognitive psychology.

However, the model is based on the simplifying assumption that, *for a given, fixed process*, there are no differences in expected response times between errors and correct responses. This is natural, as any other assumption would introduce an exogenous asymmetry which would confound the effect identified in Theorem 4. That is, this theorem predicts an asymmetry in response times even though no such asymmetry is assumed at the individual-process level.

Also, in their most basic formulation, classical models from psychology predict no asymmetry in the responses times of errors and correct responses. For example, the original, symmetric-boundaries drift-diffusion model of Ratcliff (1978, 1981) predicts identical response times distributions for either response, unless trial-by-trial variability in either drift rates or starting points is assumed. However, works as Fudenberg et al. (2018) have found evidence for slow errors in tasks which should not involve multiple processes, which can be generated in the drift-diffusion model through appropriate assumptions on the shape of the boundaries or the trial-by-trial distribution of drift rates. In terms of our model, this would replace the assumption that expected process response times are independent of the selected answer with a "slow errors" assumption. Consider a decision process b and a decision problem $\omega = (A_\omega, B_\omega) \in \Omega_b$. Let x denote an option in ω, that is, $x \in \{A_\omega, B_\omega\}$. Denote by $T_b(\omega, x)$ the expected response time of process b conditional on the response x being selected.

Call this the *model with alternative-dependent response times.* The property that the heuristic process is faster than the deliberative one needs to be reformulated as follows.

Definition 5. For the model with option-dependent response times, process h is faster than process d on Ω_h if $T_h(\omega, h(\omega)) < T_d(\omega, d(\omega))$ for all $\omega \in \Omega_h$.

The slow errors property is then captured as follows.

Definition 6. For the model with option-dependent response times, a process b exhibits slow process errors if $T_b(\omega, b(\omega)) < T_b(\omega, x)$ for $x \neq b(\omega)$.

Almost all our results hold for this more general case. First, Theorems 1 and 2 hold without change, as they are unaffected by assumptions on response times. Theorem 3, on the generalized Stroop effect, holds replacing the assumption that h is faster than d in the sense of Definition 4 with the analogous, weaker property spelled out in Definition 5. Note that the condition that T_b is a function of $|\Delta|$ needs to be weakened to the statement that $T_b(\omega, b(\omega))$ is a function of $|\Delta|$.

Theorem 4 also holds without changes except for part (a). That is, observed errors are still predicted to be slower than correct responses under alignment, but it might not be true that they are faster than correct responses under conflict. This is because slow process errors might counteract the prediction of fast observed errors in case of conflict, while they compound the prediction of slow observed errors in case of alignment. Appendix A contains the sketched proofs for Theorems 3 and 4 for this extended model.

6.3 The Almost-Certainty Heuristic

Our second experiment focused on the common ratio and common consequence effects as expressions of the *certainty heuristic*, which is assumed to be triggered when a lottery involves a sure outcome. However, it has been argued that similar behavioral anomalies might occur when lotteries involve very likely but still uncertain outcomes. Kahneman and Tversky (1979) included an example of the common ratio construction where the probabilities of non-zero outcomes were substantial (but not one) in one pair (\$6000 with probability 0.45 vs. \$3000 with probability 0.9) and were scaled down to become "minuscule" in the second, associated pair (\$6000 with probability 0.001 vs. \$3000 with probability 0.002). They argued that, in the first pair, most people chose the lottery where winning was most probable. One could hence hypothesize that the certainty heuristic could be weakened to an 'almost-certainty" heuristic when the probabilities of winning are close to but smaller than one.

To examine this hypothesis, the certainty-effect experiment described in Section 4.2 also included 32 lottery double-pairs that targeted this "almost-certainty effect," using the common-ratio construction from Kahneman and Tversky (1979). We can then apply the model presented in Sections 2 and 4 to the almost-certainty effect and test its predictions. For this purpose, the deliberative process d is again assumed to be maximization of expected utility, and strength-of-preference effects are operationalized based on the individually-estimated expected utilities already used for the rest of the experiment (recall Section 4.2.3). The "almost-certainty" heuristic h is triggered by lotteries involving outcomes with probability close to one, i.e. it is a heuristic triggered by alternative S, as described in Section 4.

Appendix C describes the design and the results. We find support for Theorems 1 and 2, and in particular both evidence for strength-of-preference effects and a discontinuity at $\Delta = 0$ as predicted due to process multiplicity (note that the almost-certainty heuristic is triggered by S). We also find support for the response-times predictions in Theorems 3 and 4, further confirming the presence of dual-process effects when controlling for possible confounds from strength of preference.

7 Discussion

Classical economic theories share the idea that *all* observed choices of an economic agent should be organized as if arising from the maximization of a single, abstract preference relation. However, there is abundant evidence that this interpretation is at odds with actual human behavior. The reasons are twofold.

First, choice arises from noisy computations in the human brain, and is hence stochastic and inherently linked to cardinal decision values. Cognitive imprecision in the perception, computation, and maximization of those values generates well-known psychometric and chronometric effects which can be broadly described as "strength of preference." That is, errors are more frequent and choices are slower when decision values are closer to each other.

Second, oftentimes choices are not determined by a single decision process, but rather are the result of the interplay and even conflict between different decision processes, behavioral rules, or heuristics that can lead to fundamental inconsistencies in observed behavior, which systematically contradict preference maximization. This creates a multiplicity of decision processes which generates error patterns that are fundamentally different from the ones arising from cognitive imprecision.

In this work, we have proposed a simple, parsimonious model where decisions arise from a mixture of cognitively-imprecise value maximization and simpler cognitive shortcuts or heuristics. Hence, economic errors arise both due to noise and as a result of bias. In two different experiments, targeting anomalies in belief updating and decisions under risk, respectively, we show that the model successfully predicts complex patterns in choice frequencies, error rates, and response times. In a belief-updating experiment, we find strength-of-preference patterns as predicted by cognitive imprecision, but only when controlling for the interaction between (noisy) expected-value maximization and a simple conservatism heuristic. In a lottery-choice experiment focused on the celebrated certainty effect, we reproduce those findings and further find clear evidence (in the form of a diagnostic discontinuity in error rates) for the existence of a heuristic favoring risk-free alternatives which conflicts with noisy expected-utility maximization.

These findings emphasize the importance of accounting for both process multiplicity (and, in particular, the presence of cognitive shortcuts) and cognitive imprecision in value maximization (and, specifically, strength-of-preference effects). Neither one nor the other suffice, in isolation, to explain observed patterns of behavior. We conclude that jointly modeling noise, bias, and their interaction provides a parsimonious, unified framework which can fully account for deviations from normative prescriptions across different economic domains.

References

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

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

Achtziger, A., C. Alós-Ferrer, S. Hügelschäfer, and M. Steinhauser (2015). Higher Incentives Can Impair Performance: Neural Evidence on Reinforcement and Rationality. *Social Cognitive and Affective Neuroscience 10*(11), 1477–1483.

Agranov, M. and P. Ortoleva (2017). Stochastic Choice and Preferences for Randomization. *Journal of Political Economy 125*(1), 40–68.

Allais, M. (1953). Le Comportement de l'Homme Rationnel Devant le Risque, Critique des Postulates et Axiomes de l'Ecole Americaine. *Econometrica 21*, 503–546.

Allport, G. (1954). *The Nature of Prejudice.* Reading, Massachusetts: Addison-Wesley.

Alós-Ferrer, C. (2018a). A Review Essay on Social Neuroscience: Can Research on the Social Brain and Economics Inform Each Other? *Journal of Economic Literature 56*(1), 1–31.

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

Alós-Ferrer, C., E. Fehr, H. Fehr-Duda, and M. Garagnani (2024). Common Ratio and Common Consequence Effects Arise from True Preferences. Working Paper (available at `alosferrer.github.io`).

Alós-Ferrer, C., E. Fehr, and N. Netzer (2021). Time Will Tell: Recovering Preferences when Choices are Noisy. *Journal of Political Economy 129*(6), 1828–1877.

Alós-Ferrer, C. and M. Garagnani (2022a). Strength of Preference and Decisions Under Risk. *Journal of Risk and Uncertainty 64*(3), 309–329.

Alós-Ferrer, C. and M. Garagnani (2022b). The Gradual Nature of Economic Errors. *Journal of Economic Behavior and Organization 200*, 55–66.

Alós-Ferrer, C. and M. Garagnani (2023). Part-Time Bayesians: Incentives and Behavioral Heterogeneity in Belief Updating. *Management Science 69*(9), 5523–5542.

Alós-Ferrer, C. and M. Garagnani (2024a). Errors, Fast and Slow. Submitted (available at `alosferrer.github.io`).

Alós-Ferrer, C. and M. Garagnani (2024b). Improving Risky-Choice Predictions Using Response Times. *Journal of Political Economy: Microeconomics 2(2)*, 335–354.

Alós-Ferrer, C., A. Jaudas, and A. Ritschel (2021). Effortful Bayesian Updating: A Pupil-dilation Study. *Journal of Risk and Uncertainty 63*(1), 81–102.

Alós-Ferrer, C. and A. Ritschel (2021). Multiple Behavioral Rules in Cournot Oligopolies. *Journal of Economic Behavior and Organization 183*, 250–267.

Alós-Ferrer, C. and F. Strack (2014). From Dual Processes to Multiple Selves: Implications for Economic Behavior. *Journal of Economic Psychology 41*, 1–11.

Alós-Ferrer, C. and S. Weidenholzer (2008). Contagion and Efficiency. *Journal of Economic Theory 143*, 251–274.

Alós-Ferrer, C. and S. Weidenholzer (2014). Imitation and the Role of Information in Overcoming Coordination Failures. *Games and Economic Behavior 87*, 397–411.

Anderson, S. P., J.-F. Thisse, and A. De Palma (1992). *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.

Atkinson, A. C. (1996). The Usefulness of Optimum Experimental Designs. *Journal of the Royal Statistical Society 51*(1), 59–76.

Baddeley, A. D. (1992). Working Memory. *Science 255*(5044), 556–559.

Ballinger, T. P. and N. T. Wilcox (1997). Decisions, Error and Heterogeneity. *Economic Journal 107*(443), 1090–1105.

Baltas, G. and P. Doyle (2001). Random Utility Models in Marketing Research: A Survey. *Journal of Business Research 51*(2), 115–125.

Bellemare, C., S. Kröger, and A. van Soest (2008). Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities. *Econometrica 76*(4), 815–839.

Benhabib, J. and A. Bisin (2005). Modeling Internal Commitment Mechanisms and Self-Control: A Neuroeconomics Approach to Consumption-Saving Decisions. *Games and Economic Behavior 52*(2), 460–492.

Blavatskyy, P., V. Panchenko, and A. Ortmann (2023). How Common is the Common-Ratio Effect? *Experimental Economics 26*(2), 253–272.

Blavatskyy, P. R. (2007). Stochastic Expected Utility Theory. *Journal of Risk and Uncertainty 34*, 259–286.

Blavatskyy, P. R. (2010). Reverse Common Ratio Effect. *Journal of Risk and Uncertainty 40*(3), 219–241.

Block, H. D. and J. Marschak (1960). Random Orderings and Stochastic Theories of Responses. In I. Olkin (Ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pp. 97–132. Stanford: Stanford University Press.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica 82*(6), 2295–2326.

Camerer, C. F. (1989). Does the Basketball Market Believe in the 'Hot Hand'? *American Economic Review 79*, 1257–1261.

Camerer, C. F., G. Loewenstein, and D. Prelec (2005). Neuroeconomics: How Neuroscience Can Inform Economics. *Journal of Economic Literature 43*(1), 9–64.

Caplin, A. (2025). *An Introduction to Cognitive Economics: The Science of Mistakes.* Cham, Switzerland: Palgrave Macmillan.

Cappelen, A. W., U. H. Nielsen, B. Tungodden, J.-R. Tyran, and E. Wengström (2016). Fairness is Intuitive. *Experimental Economics 19*(4), 727–740.

Cappellari, L. and S. P. Jenkins (2003). Multivariate Probit Regression Using Simulated Maximum Likelihood. *The Stata Journal 3*(3), 278–294.

Cattell, J. M. (1893). On Errors of Observation. *The American Journal of Psychology 5*(3), 285–293.

Cattell, J. M. (1902). The Time of Perception as a Measure of Differences in Intensity. *Philosophische Studien 19*, 63–68.

Cerigioni, F. (2021). Dual Decision Processes: Retrieving Preferences When Some Choices Are Automatic. *Journal of Political Economy 129*(6), 1667–1704.

Cerreia-Vioglio, S., D. Dillenberger, and P. Ortoleva (2015). Cautious Expected Utility and the Certainty Effect. *Econometrica 83*(2), 693–728.

Chabris, C. F., C. L. Morris, D. Taubinsky, D. Laibson, and J. P. Schuldt (2009). The Allocation of Time in Decision-Making. *Journal of the European Economic Association 7*(2-3), 628–637.

Cohen, J. D., K. Dunbar, and J. L. McClelland (1990). On the Control of Automatic Processes: A Parallel Distributed Processing Account of the Stroop Effect. *Psychological Review 97*(3), 332–361.

Conlisk, J. (1989). Three Variants on the Allais Example. *American Economic Review 79*(3), 392–407.

Conte, A., J. D. Hey, and P. G. Moffatt (2011). Mixture Models of Choice Under Risk. *Journal of Econometrics 162*(1), 79–88.

Dashiell, J. F. (1937). Affective Value-Distances as a Determinant of Aesthetic Judgment-Times. *American Journal of Psychology 50*, 57–67.

Davidson, D. and J. Marschak (1959). Experimental Tests of a Stochastic Decision Theory. In W. Churchman and P. Ratoosh (Eds.), *Measurement: Definitions and Theories*, Volume I, Part I, pp. 233–269. New York: Wiley.

Debreu, G. (1958). Stochastic Choice and Cardinal Utility. *Econometrica 26*(3), 440–444.

Dillenberger, D. (2010). Preferences for One-Shot Resolution of Uncertainty and Allais-Type Behavior. *Econometrica 78*(6), 1973–2004.

Edwards, W. (1968). Conservatism in Human Information Processing. In B. Kleinmuntz (Ed.), *Formal Representation of Human Judgment*, pp. 17–52. New York, NY: Wiley.

El-Gamal, M. A. and D. M. Grether (1995). Are People Bayesian? Uncovering Behavioral Strategies. *Journal of the American Statistical Association 90*(432), 1137–1145.

Enke, B. and T. Graeber (2023). Cognitive Uncertainty. *Quarterly Journal of Economics 138*(4), 2021–2067.

Enke, B., T. Graeber, and R. Oprea (2025). Complexity and Time. *Journal of the European Economic Association forthcoming*.

Evans, A. M., D. G. Rand, and K. D. Dillon (2015). Fast But Not Intuitive, Slow But Not Reflective: Decision Conflict Drives Reaction Times in Social Dilemmas. *Journal of Experimental Psychology: General 144*(5), 951–966.

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

Fechner, T. G. (1860). *Elemente der Psychophysik [Elements of Psychophysics]*. Leipzig: Breitkopf & Härtel.

Fischhoff, B. and R. Beyth-Marom (1983). Hypothesis Evaluation from a Bayesian Perspective. *Psychological Review 90*(3), 239–260.

Ford, I., B. Torsney, and C. F. J. Wu (1992). The Use of a Canonical Form in the Construction of Locally Optimal Designs for Non-Linear Problems. *Journal of the Royal Statistical Society 54*(2), 569–583.

Frydman, C. and L. J. Jin (2022). Efficient Coding and Risky Choice. *Quarterly Journal of Economics 137*(1), 161–213.

Fudenberg, D. and D. K. Levine (2006). A Dual-Self Model of Impulse Control. *American Economic Review 96*(5), 1449–1476.

Fudenberg, D. and D. K. Levine (2012). Timing And Self-Control. *Econometrica 80*(1), 1–42.

Fudenberg, D., P. Strack, and T. Strzalecki (2018). Speed, Accuracy, and the Optimal Timing of Choices. *American Economic Review 108*(12), 3651–3684.

Gennaioli, N. and A. Shleifer (2010). What Comes to Mind. *Quarterly Journal of Economics 125*(4), 1399–1433.

Grether, D. M. (1980). Bayes Rule as a Descriptive Model: The Representativeness Heuristic. *Quarterly Journal of Economics 95*, 537–557.

Grether, D. M. (1992). Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence. *Journal of Economic Behavior and Organization 17*, 31–57.

Halton, J. H. (1960). On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals. *Numerische Mathematik 2*(1), 84–90.

Harless, D. W. and C. F. Camerer (1994). The Predictive Utility of Generalized Expected Utility Theories. *Econometrica 62*(6), 1251–1289.

Harrison, G. W. and E. E. Rutström (2008). Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods. In C. R. Plott and V. L. Smith (Eds.), *Handbook of Experimental Economics Results*, Volume 1, Part 5, pp. 752–767. Elsevier.

He, G., S. Wang, and B. Zhang (2020). Watering Down Environmental Regulation in China. *Quarterly Journal of Economics 135*(4), 2135–2185.

Heller, S. B., A. K. Shah, J. Guryan, J. Ludwig, S. Mullainathan, and H. A. Pollack (2017). Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago. *Quarterly Journal of Economics 132*(1), 1–54.

Hey, J. D. (2005). Why We Should Not Be Silent About Noise. *Experimental Economics 8*, 325–345.

Hey, J. D. and C. Orme (1994). Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica 62*(6), 1291–1326.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review 93*(5), 1449–1475.

Kahneman, D. (2011). *Thinking, Fast and Slow*. London: Macmillan.

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

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica 47*(2), 263–291.

Kandori, M. and R. Rob (1995). Evolution of Equilibria in the Long Run: A General Theory and Applications. *Journal of Economic Theory 65*, 383–414.

Khaw, M. W., Z. Li, and M. Woodford (2021). Cognitive Imprecision and Small-Stakes Risk Aversion. *Review of Economic Studies 88*(4), 1979–2013.

Klein, A. S. (2001). Measuring, Estimating, and Understanding the Psychometric Function: A Commentary. *Attention, Perception, & Psychophysics 63*(8), 1421–1455.

Krajbich, I., B. Bartling, T. Hare, and E. Fehr (2015). Rethinking Fast and Slow Based on a Critique of Reaction-Time Reverse Inference. *Nature Communications 6*(7455), 1–9.

Laming, D. (1985). Some Principles of Sensory Analysis. *Psychological Review 92*(4), 462–485.

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

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis.* New York: Wiley.

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

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

Machina, M. J. (1982). Expected Utility Analysis without the Independence Axiom. *Econometrica 50*(2), 277–323.

McFadden, D. L. (2001). Economic Choices. *American Economic Review 91*(3), 351–378.

McGranaghan, C., K. Nielsen, T. O'Donoghue, J. Somerville, and C. D. Sprenger (2024). Distinguishing Common Ratio Preferences from Common Ratio Effects Using Paired Valuation Tasks. *American Economic Review 114*, 307–347.

Moffatt, P. G. (2015). *Experimetrics: Econometrics for Experimental Economics.* London: Palgrave Macmillan.

Moffatt, P. G., S. Sitzia, and D. J. Zizzo (2015). Heterogeneity in Preferences Towards Complexity. *Journal of Risk and Uncertainty 51*(2), 147–170.

Mosteller, F. and P. Nogee (1951). An Experimental Measurement of Utility. *Journal of Political Economy 59*, 371–404.

Moyer, R. S. and T. K. Landauer (1967). Time Required for Judgements of Numerical Inequality. *Nature 215*(5109), 1519–1520.

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

Myrseth, K. O. R. and C. E. Wollbrant (2016). Commentary: Fairness is Intuitive. *Frontiers in Psychology 7*(654), 1–2.

Nownes, A. J. (1992). Primaries, General Elections, and Voter Turnout: A Multinomial Logit Model of the Decision to Vote. *American Politics Quarterly 20*(2), 205–226.

Oechssler, J. (2002). Cooperation as a Result of Learning with Aspiration Levels. *Journal of Economic Behavior and Organization 49*, 405–409.

Oprea, R. (2024). Decisions Under Risk are Decisions Under Complexity. *American Economic Review 12*(114), 3789–3811.

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

Peirce, J. W. (2007). PsychoPy – Psychophysics Software in Python. *Journal of Neuroscience Methods 162*(1), 8–13.

Polanía, R., M. Woodford, and C. C. Ruff (2019). Efficient Coding of Subjective Value. *Nature Neuroscience 22*(1), 134–142.

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

Ratcliff, R. (1981). A Theory of Order Relations in Perceptual Matching. *Psychological Review 88*, 552–572.

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

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

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

Redish, A. D. (2016). Vicarious Trial and Error. *Nature Reviews Neuroscience 17*(3), 147–159.

Rubinstein, A. (2007). Instinctive and Cognitive Reasoning: A Study of Response Times. *Economic Journal 117*(523), 1243–1259.

Rubinstein, A. (2016). A Typology of Players: Between Instinctive and Contemplative. *Quarterly Journal of Economics 131*(2), 859–890.

Ruggeri, K., S. Alí, M. L. Berge, G. Bertoldo, L. D. Bjørndal, A. Cortijos-Bernabeu, C. Davison, E. Demić, C. Esteban-Serna, M. Friedemann, et al. (2020). Replicating Patterns of Prospect Theory for Decision Under Risk. *Nature Human Behaviour 4*(6), 622–633.

Schmidt, U. (1998). A Measurement of the Certainty Effect. *Journal of Mathematical Psychology 42*(1), 32–47.

Schneider, W. and R. M. Shiffrin (1977). Controlled and Automatic Human Information Processing: I. Detection, Search, and Attention. *Psychological Review 84*(1), 1–66.

Shadlen, M. N. and R. Kiani (2013). Decision Making as a Window on Cognition. *Neuron 80*, 791–806.

Shadlen, M. N. and D. Shohamy (2016). Decision Making and Sequential Sampling from Memory. *Neuron 90*, 927–939.

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

Silvey, S. D. (1980). *Optimal Design: An Introduction to the Theory for Parameter Estimation*, Volume 1. New York: Chapman and Hall.

Sloman, S. A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin 119*(1), 3–22.

Spiliopoulos, L. (2018). The Determinants of Response Time in a Repeated Constant-Sum Game: A Robust Bayesian Hierarchical Dual-Process Model. *Cognition 172*, 107–123.

Spiliopoulos, L. and A. Ortmann (2018). The BCD of Response Time Analysis in Experimental Economics. *Experimental Economics 21*(2), 383–433.

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

Thaler, R. H. and H. M. Shefrin (1981). An Economic Theory of Self-Control. *Journal of Political Economy 89*(2), 392–406.

Thurstone, L. L. (1927). A Law of Comparative Judgement. *Psychological Review 34*, 273–286.

Train, K. E. (2009). *Discrete Choice Methods with Simulation* (2nd ed.). New York: Cambridge University Press.

Tversky, A. (1969). Intransitivity of Preferences. *Psychological Review 76*, 31–48.

Tversky, A. and D. Kahneman (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

Vieider, F. M. (2024). Decisions Under Uncertainty as Bayesian Inference on Choice Options. *Management Science forthcoming*.

Von Gaudecker, H.-M., A. Van Soest, and E. Wengström (2011). Heterogeneity in Risky Choice Behavior in a Broad Population. *American Economic Review 101*(2), 664–694.

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

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

Wichmann, A. F. and N. J. Hill (2001). The Psychometric Function: I. Fitting, Sampling, and Goodness of Fit. *Attention, Perception, & Psychophysics 63*(8), 1293–1313.

ONLINE APPENDIX

Noise and Bias: The Cognitive Roots of Economic Errors

C. Alós-Ferrer, J. Buckenmaier, and M. Garagnani

A Proofs

Theorem 1. *Suppose the deliberative process d is psychometric and the heuristic h has constant consistency $P_h > 1/2$. Then the following statements hold.*

(a1) For neutral decision problems, $\omega \in \Omega_h^N$, the probability of a correct answer is increasing in Δ_ω for $\Delta_\omega > 0$, and decreasing in Δ_ω for $\Delta_\omega < 0$. Further, the probability is 1/2 for $\Delta_\omega = 0$

(a2) For decision problems in alignment, $\omega \in \Omega_h^A$, the probability of a correct answer is increasing in Δ_ω for $\Delta_\omega > 0$, and decreasing in Δ_ω for $\Delta_\omega < 0$.

(a3) For decision problems in conflict, $\omega \in \Omega_h^C$, the probability of a correct answer is increasing in Δ_ω for $\Delta_\omega > 0$, and decreasing in Δ_ω for $\Delta_\omega < 0$.

(b1) For any given Δ, the probability of a correct response is larger for decisions under alignment compared to conflict decision problems. That is, for any fixed Δ and decision problems $\omega^C \in \Omega_h^C, \omega^A \in \Omega_h^A$ with $\Delta_{\omega^C} = \Delta_{\omega^A} = \Delta$, it follows that

$$\text{Prob}(\rho(\omega^A) = d(\omega^A)) > \text{Prob}(\rho(\omega^C) = d(\omega^C)).$$

(b2) For any given Δ, the probability of a correct response is smaller for decisions under conflict compared to neutral decision problems. That is, for any fixed Δ and decision problems $\omega^C \in \Omega_h^C, \omega_h \in \Omega_h^N$ with $\Delta_{\omega^C} = \Delta_{\omega^N} = \Delta$, it follows that

$$\text{Prob}(\rho(\omega^N) = d(\omega^N)) > \text{Prob}(\rho(\omega^C) = d(\omega^C)).$$

(b3) Suppose the heuristic h is more consistent than d on Ω_h. Then, for any given Δ, the probability of a correct response is larger for decisions under alignment compared to neutral decision problems. That is, for any fixed Δ and decision problems $\omega^A \in \Omega_h^A, \omega^N \in \Omega_h^N$ with $\Delta_{\omega^A} = \Delta_{\omega^N} = \Delta$, it follows that

$$\text{Prob}(\rho(\omega^A) = d(\omega^A)) > \text{Prob}(\rho(\omega^N) = d(\omega^N)).$$

Proof of Theorem 1. To show (a1), consider $\omega \in \Omega_h^N$, i.e., neutral decision problems. For these problems, the probability of a correct answer is simply $P_d(\omega)$ as only d is active. For $\Delta_\omega > 0$, we have that $P_d(\omega) = \beta_d^A(\omega)$, which is increasing in Δ_ω because d is psychometric. For $\Delta_\omega < 0$ we have $P_d(\omega) = 1 - \beta_d^A(\omega)$, which is decreasing for the same reason. Last, for $\Delta_\omega = 0$ it follows that $P_d(\omega) = \beta_d^A(\omega) = \frac{1}{2}$ by Definition 2. This completes the proof of (a1).

Consider now decision problems under alignment, $\omega \in \Omega_h^A$. For $\Delta_\omega > 0$ the probability of a correct answer is $(1-\alpha)P_d(\omega) + \alpha P_h = (1-\alpha)\beta_d^A(\omega) + \alpha P_h$, which is increasing in Δ_ω. For $\Delta_\omega < 0$ the probability of a correct answer is $(1-\alpha)P_d(\omega) + \alpha P_h = (1-\alpha)(1-\beta_d^A(\omega)) + \alpha P_h$, which is decreasing in Δ_ω. This shows (a2).

The argument for decision problems under conflict, $\omega \in \Omega_h^C$, is analogous. For $\Delta_\omega > 0$ the probability of a correct answer is $(1-\alpha)P_d(\omega) + \alpha(1-P_h) = (1-\alpha)\beta_d^A(\omega) +$

$\alpha(1-P_h)$, which is increasing in Δ_ω. For $\Delta_\omega < 0$ the probability of a correct answer is $(1-\alpha)P_d(\omega) + \alpha(1-P_h) = (1-\alpha)(1-\beta_d^A(\omega)) + \alpha(1-P_h)$, which is decreasing in Δ_ω. This shows (a3).

Fix Δ. To see (b1)–(b3), let $\omega^C \in \Omega_h^C$, $\omega^A \in \Omega_h^A$, and $\omega^N \in \Omega_h^N$ with $\Delta_{\omega^C} = \Delta_{\omega^A} = \Delta_{\omega^N} = \Delta$. Since d is psychometric, it follows that $P_d(\omega^C) = P_d(\omega^A) = P_d(\omega^N) = P_d$. Thus,

$$\text{Prob}(\rho(\omega^A) = d(\omega^A)) = (1-\alpha)P_d + \alpha P_h > (1-\alpha)P_d + \alpha(1-P_h) = \text{Prob}(\rho(\omega^C) = d(\omega^C))$$

because $1 - P_h < 1/2 < P_h$. This shows (b1). Analogously,

$$\text{Prob}(\rho(\omega^N) = d(\omega^N)) = P_d > (1-\alpha)P_d + \alpha(1-P_h) = \text{Prob}(\rho(\omega^C) = d(\omega^C))$$

because $1 - P_h < 1/2 \leq P_d$. This shows (b2). Last, assume that h is more consistent than d. Then,

$$\text{Prob}(\rho(\omega^A) = d(\omega^A)) = (1-\alpha)P_d + \alpha P_h > P_d = \text{Prob}(\rho(\omega^N) = d(\omega^N))$$

because $P_h > P_d$ as h is more consistent than d. This shows (b3). □

Proof of Theorem 1 with cognitive control costs. With cognitive control costs, α is replaced with $\alpha(|\Delta_\omega|)$, for a strictly decreasing function $\alpha(\cdot)$. (a1) holds as h is not active, hence the probability α plays no role. As explained in the main text, (a2) might not hold. (b1), (b2), and (b3) hold, as they compare decision problems with a fixed Δ, hence α is simply replaced by $\alpha(\Delta)$.

It just remains to prove (a3). Let $\omega, \omega' \in \Omega_h^C$, and let $\alpha = \alpha(|\Delta_\omega|)$ and $\alpha' = \alpha(|\Delta_{\omega'}|)$.

Suppose first that $\Delta_\omega > \Delta_{\omega'} \geq 0$. The probability of a correct response is larger for ω compared to ω' if

$$(1-\alpha)\beta_d^A(\omega) + \alpha(1-P_h) > (1-\alpha')\beta_d^A(\omega') + \alpha'(1-P_h)$$

This holds if and only if

$$(1-\alpha)\left[\beta_d^A(\omega) - \beta_d^A(\omega')\right] + (\alpha' - \alpha)\beta_d^A(\omega') > (\alpha' - \alpha)(1-P_h)$$

Since $\beta_d^A(\omega) > \beta_d^A(\omega')$ (since β_d^A is increasing) and $\alpha' > \alpha$ (since $\alpha(\cdot)$ is strictly decreasing), the latter inequality holds if $\beta_d^A(\omega') > (1-P_h)$. However, $\beta_d^A(\omega') \geq 1/2$ and $1 - P_h < 1/2$. Hence, the conclusion follows.

Suppose now that $\Delta_\omega < \Delta_{\omega'} \leq 0$. The probability of a correct response is larger for ω compared to ω' if

$$(1-\alpha)(1-\beta_d^A(\omega)) + \alpha(1-P_h) > (1-\alpha')(1-\beta_d^A(\omega')) + \alpha'(1-P_h)$$

This holds if and only if

$$(1-\alpha)\left[\beta_d^A(\omega') - \beta_d^A(\omega)\right] + (\alpha' - \alpha)(1-\beta_d^A(\omega')) > (\alpha' - \alpha)(1-P_h)$$

In this case, $\beta_d^A(\omega') > \beta_d^A(\omega)$ (since β_d^A is increasing), and again $\alpha' > \alpha$ (since $\alpha(\cdot)$ is strictly decreasing in $|\Delta|$). Thus, the inequality above holds if $1 - \beta_d^A(\omega') > 1 - P_h$. However, $1 - \beta_d^A(\omega') \geq 1/2$ (as $\Delta_{\omega'} \leq 0$) and $1 - P_h < 1/2$, and the conclusion follows. □

Theorem 2. *Suppose the deliberative process d is psychometric and the heuristic h has constant consistency $P_h > 1/2$ and is triggered by A. Then the following statements hold.*

(a) The probability of a correct answer is increasing in Δ_ω for $\Delta_\omega > 0$ (alignment) *and decreasing in Δ_ω for $\Delta_\omega < 0$* (conflict).

(b) For decision problems where h is active, $\omega \in \Omega_h$, there is a discontinuity at $\Delta_\omega = 0$ with more correct answers in alignment ($\Delta_\omega > 0$) than in conflict ($\Delta_\omega < 0$).

(c) For neutral decision problems, there is no discontinuity and the probability of a correct answer at $\Delta_\omega = 0$ is $1/2$.

Proof of Theorem 2. (a) For $\Delta_\omega > 0$ (resp. $\Delta_\omega < 0$), decision problems are in alignment (resp. conflict). Hence, the probability of a correct answer is increasing (resp. decreasing) in Δ_ω by Theorem 1(a2, resp. a3).

(b) When h is active and $\Delta_\omega < 0$, the probability of a correct response is

$$\text{Prob}(\rho(\omega^C) = d(\omega^C)) = (1-\alpha)P_d(\omega^C) + \alpha(1-P_h).$$

Since d is psychometric (Definition 2), the limit of this value as $\Delta_\omega \to 0-$ is

$$P_- = (1-\alpha)\frac{1}{2} + \alpha(1-P_h) = \frac{1}{2} + \alpha\left(\frac{1}{2} - P_h\right).$$

When h is active and $\Delta_\omega > 0$, the probability of a correct response is

$$\text{Prob}(\rho(\omega^A) = d(\omega^A)) = (1-\alpha)P_d(\omega^A) + \alpha P_h.$$

Again, since d is psychometric, the limit of this value as $\Delta_\omega \to 0+$ is

$$P_+ = (1-\alpha)\frac{1}{2} + \alpha P_h = \frac{1}{2} + \alpha\left(P_h - \frac{1}{2}\right).$$

Since $P_h > \frac{1}{2}$, it follows that $P_+ - P_- = \alpha\,(2P_h - 1) > 0$, i.e. there is a discontinuity at zero, with a positive jump when moving from negative to positive values of Δ_ω.

(c) When h is inactive (neutral decision problems), the probability of a correct response is

$$\text{Prob}(\rho(\omega^N) = d(\omega^N)) = P_d(\omega^N),$$

which is continuous and takes the value $1/2$ at $\Delta_\omega = 0$ since d is psychometric. □

Proof of Theorem 2 with cognitive control costs. Let $\alpha = \alpha(|\Delta_\omega|)$ for a strictly decreasing function $\alpha(\cdot)$. Part (a) for the case of conflict follows from Theorem 1(a3), which also holds under cognitive control costs. For the case of alignment, as explained in the text, the result might not hold. For part (b), the proof is as above with the only differences that the expressions for limits P_- and P_+ involve $\alpha(0)$ instead of α. Last, part (c) holds as it only involves neutral decision problems and α plays no role. □

Corollary 1. *Suppose the deliberative process d is psychometric and the heuristic h has constant consistency $P_h > 1/2$ and is triggered by A.*

(a) For decision problems where h is active, the discontinuity at $\Delta_\omega = 0$ is equal to $\bar{D} = \alpha(2P_h - 1)$.

(b) *In particular, the parameters* α *and* P_h *are partially identified by* $\bar{D}$. *Specifically,* $\alpha > \bar{D}$ *and* $P_h > \frac{\bar{D}+1}{2}$.

(c) *Further, the left and right parts of the discontinuity are symmetric. That is, if* P_- *and* P_+ *are the limits of the probabilities of a correct response as* $\Delta_\omega \to 0-$ *and* $\Delta_\omega \to 0+$, *respectively, then*

$$\frac{1}{2} - P_- = P_+ - \frac{1}{2}.$$

Proof of Corollary 1. (a) and (c) follow from the expressions for P_+ and P_- computed in the proof of Theorem 2. To see (b), note that $P_h < 1$ implies that $\bar{D} = \alpha(2P_h - 1) < \alpha$, and $\alpha < 1$ implies that $\bar{D} < 2P_h - 1$, hence $\frac{1}{2}\left(\bar{D} + 1\right) < P_h$. For the model with cognitive control costs, the results follow replacing α with $\alpha(0)$. □

Theorem 3 (Generalized Stroop effect)**.** *Consider a set of decision problems* Ω. *Suppose* T_d *and* T_h *are functions of* $|\Delta_\omega|$, d *is psychometric,* h *is faster than* d *on* Ω_h, *and* $P_h(\omega) > 1/2$ *on* Ω_h. *Let* ω^A *in* Ω_h^A *(alignment) and* ω^C *in* Ω_h^C *(conflict). Then:*

(a) *If* $\Delta_{\omega^A} = \Delta_{\omega^C}$, *then*

$$E[RT(\omega^A) \mid \rho(\omega^A) = d(\omega^A)] < E[RT(\omega^C) \mid \rho(\omega^C) = d(\omega^C)],$$

that is, correct answers in alignment are faster than correct answers in comparable conflict situations.

(b) *Additionally, suppose* d *is symmetric in the sense that* $P_d(\omega) = P_d(\omega')$ *if* $|\Delta_{\omega'}| = |\Delta_\omega|$. *Then, if* $|\Delta_{\omega^A}| = |\Delta_{\omega^C}|$, *the property in (a) also holds.*

(c) *If* d *is symmetric and* $\omega, \omega' \in \Omega_h^N$ *are neutral decision problems with* $|\Delta_{\omega'}| = |\Delta_\omega|$, *then* $E[RT(\omega) \mid \rho(\omega) = d(\omega)] = E[RT(\omega') \mid \rho(\omega') = d(\omega')]$.

Proof. The response time of a correct choice in alignment situation ω^A is

$$E[RT(\omega^A) \mid \rho(\omega^A) = d(\omega^A)] = \frac{(1-\alpha)P_d(\omega^A)T_d(\omega^A) + \alpha P_h(\omega^A)T_h(\omega^A)}{(1-\alpha)P_d(\omega^A) + \alpha P_h(\omega^A)}$$

and the response time of a correct answer in conflict situation ω^C is

$$E[RT(\omega^C) \mid \rho(\omega^C) = d(\omega^C)] = \frac{(1-\alpha)P_d(\omega^C)T_d(\omega^C) + \alpha(1 - P_h(\omega^C))T_h(\omega^C)}{(1-\alpha)P_d(\omega^C) + \alpha(1 - P_h(\omega^C))}.$$

Thus, $E[RT(\omega^C) \mid \rho(\omega^C) = d(\omega^C)] > E[RT(\omega^A) \mid \rho(\omega^A) = d(\omega^A)]$ if and only if

$$\begin{aligned}
&(1-\alpha)^2 P_d(\omega^A)P_d(\omega^C)(T_d(\omega^C) - T_d(\omega^A)) + \\
&\qquad \alpha(1-\alpha)P_d(\omega^A)(1 - P_h(\omega^C))(T_h(\omega^C) - T_d(\omega^A)) \\
&\qquad + \alpha(1-\alpha)P_d(\omega^C)P_h(\omega^A)(T_d(\omega^C) - T_h(\omega^A)) \\
&\qquad\qquad + \alpha^2 P_h(\omega^A)(1 - P_h(\omega^C))(T_h(\omega^C) - T_h(\omega^A)) > 0
\end{aligned}$$

Since T_d is a function of $|\Delta|$ and $\Delta_{\omega^A} = \Delta_{\omega^C}$, we have $T_d(\omega^A) = T_d(\omega^C) = T_d$. Further, since T_h is also a function of $|\Delta|$, we have $T_h(\omega^A) = T_h(\omega^C) = T_h$. Hence, the above expression simplifies to

$$(T_d - T_h)\left[P_d(\omega^C)P_h(\omega^A) - P_d(\omega^A)(1 - P_h(\omega^C))\right] > 0. \tag{A.1}$$

Since d is psychometric and $\Delta_{\omega^A} = \Delta_{\omega^C}$, we have $P_d(\omega^A) = P_d(\omega^C) = P_d > 0$ and $T_d > T_h$ by assumption, it follows that inequality (A.1) holds if and only if $P_h(\omega^A) > 1 - P_h(\omega^C)$. The latter follows from $P_h(\omega^A) > \frac{1}{2} > 1 - P_h(\omega^C)$, which completes the proof of part (a). To see part (b), note that, if $|\Delta_{\omega^A}| = |\Delta_{\omega^C}|$ and d is symmetric, then $P_d(\omega^A) = P_d(\omega^C)$ and the argument to show (A.1) follows as above. Part (c) is immediate by symmetry. □

Proof of Theorem 3 with option-dependent response times. Note that part (c) does not depend on the response times of process errors, and hence holds without change. For parts (a) and (b), let $P_d = P_d(\omega^A) = P_d(\omega^C)$. The response time of a correct choice in ω^A and is ω^C are as given in the proof of Theorem 3 replacing $T_b(\omega)$ with $T_b(\omega, d(\omega))$ for $b = h, d$ and $\omega = \omega^A, \omega^C$. Note that, since $T_d(\omega, d(\omega))$ and $T_h(\omega, h(\omega))$ are functions of $|\Delta|$ and $|\Delta_{\omega^A}| = |\Delta_{\omega^C}|$, we have that $T_d(\omega^A, d(\omega^A)) = T_d(\omega^C, d(\omega^C)) = T_d$ and $T_h(\omega^A, h(\omega^A)) = T_h(\omega^C, h(\omega^C)) = T_h$. Further, $T_h(\omega^A, d(\omega^A)) = T_h(\omega^A, h(\omega^A)) = T_h$ since ω^A is in alignment. Thus, $E[RT(\omega^C) \mid \rho(\omega^C) = d(\omega^C)] > E[RT(\omega^A) \mid \rho(\omega^A) = d(\omega^A)]$ if and only if

$$\alpha(1-\alpha)P_d(1 - P_h(\omega^C))(T_h(\omega^C, d(\omega^C)) - T_d) + \alpha(1-\alpha)P_d P_h(\omega^A)(T_d - T_h) \\ + \alpha^2 P_h(\omega^A)(1 - P_h(\omega^C))(T_h(\omega^C, d(\omega^C)) - T_h) > 0$$

Further, applying the slow process errors property for h, $T_h(\omega^A, d(\omega^A)) = T_h(\omega^A, h(\omega^A)) = T_h(\omega^C, h(\omega^C)) < T_h(\omega^C, d(\omega^C))$. Thus, the third summand in the expression above is strictly positive as $T_h(\omega^C, d(\omega^C)) > T_h$. Hence, a sufficient condition for the above property to hold is

$$(1 - P_h(\omega^C))(T_h(\omega^C, d(\omega^C)) - T_d) + P_h(\omega^A)(T_d - T_h) > 0$$

Since $T_h(\omega^C, d(\omega^C)) > T_h(\omega^C, h(\omega^C)) = T_h$ by slow errors, the expression in the last inequality is strictly larger than

$$(1 - P_h(\omega^C))(T_h - T_d) + P_h(\omega^A)(T_d - T_h) = (T_d - T_h)\left[P_h(\omega^A) - (1 - P_h(\omega^C))\right]$$

which is strictly positive as $P_h(\omega^A) > \frac{1}{2} > 1 - P_h(\omega^C)$. □

Theorem 4. *Consider a set of decision problems Ω. If h is faster than d and $P_h(\omega) > \frac{1}{2}$ on Ω_h, then the following statements hold.*

(a) For any $\omega \in \Omega_h^C$ (conflict), errors are faster than correct answers.

(b) Suppose h is more consistent than d. For any $\omega \in \Omega_h^A$ (alignment), errors are slower than correct answers.

Proof. To see (a), consider a decision problem $\omega \in \Omega_h^C$ (conflict). The response time of a correct answer in conflict is

$$E[RT \mid \rho(\omega) = d(\omega)] = \frac{(1-\alpha)P_d(\omega)T_d(\omega) + \alpha(1 - P_h(\omega))T_h(\omega)}{(1-\alpha)P_d(\omega) + \alpha(1 - P_h(\omega))}$$

and the response time of an error in conflict is

$$E[RT \mid \rho(\omega) \neq d(\omega)] = \frac{(1-\alpha)(1 - P_d(\omega))T_d(\omega) + \alpha P_h(\omega)T_h(\omega)}{(1-\alpha)(1 - P_d(\omega)) + \alpha P_h(\omega)}.$$

A straightforward computation shows that $E[RT \mid \rho(\omega) = d(\omega)] > E[RT \mid \rho(\omega) \neq d(\omega)]$ if and only if

$$\alpha(1-\alpha)\left[P_d(\omega)P_h(\omega) - (1-P_d(\omega))(1-P_h(\omega))\right](T_d(\omega) - T_h(\omega)) > 0. \tag{A.2}$$

Since h is faster than d, we have $T_d(\omega) > T_h(\omega)$. Since $P_h(\omega) > \frac{1}{2}$ and $P_d(\omega) \geq \frac{1}{2}$, we have $P_d(\omega)P_h(\omega) - (1-P_d(\omega))(1-P_h(\omega)) > 0$, which shows (a).

To see (b), consider a decision problem $\omega \in \Omega_h^A$. The response time of a correct choice in alignment is

$$E[RT \mid \rho(\omega) = d(\omega)] = \frac{(1-\alpha)P_d(\omega)T_d(\omega) + \alpha P_h(\omega)T_h(\omega)}{(1-\alpha)P_d(\omega) + \alpha P_h(\omega)}$$

and the response time of an error in alignment is

$$E[RT \mid \rho(\omega) \neq d(\omega)] = \frac{(1-\alpha)(1-P_d(\omega))T_d(\omega) + \alpha(1-P_h(\omega))T_h(\omega)}{(1-\alpha)(1-P_d(\omega)) + \alpha(1-P_h(\omega))}.$$

A computation shows that $E[RT \mid \rho(\omega) \neq d(\omega)] > E[RT \mid \rho(\omega) = d(\omega)]$ if and only if

$$\alpha(1-\alpha)(P_h(\omega) - P_d(\omega))(T_d(\omega) - T_h(\omega)) > 0. \tag{A.3}$$

Since h is faster than d, we have $T_d(\omega) > T_h(\omega)$. Since h is also more consistent, we also have $P_h(\omega) > P_d(\omega)$. Thus (A.3) holds, which shows (b). □

Proof of Theorem 4 with option-dependent response times. As explained in the main text, part (a) might not hold. To see part (b), let $T_d = T_d(\omega, d(\omega))$ and $T_h = T_h(\omega, h(\omega))$. Since ω is in alignment, $d(\omega) = h(\omega)$. Let $x \neq d(\omega)$ denote the other option in ω (the error).

The expected response time of a correct answer is as given in the proof of Theorem 4(b) replacing $T_d(\omega)$ and $T_h(\omega)$ by T_d and T_h, respectively. The expected response time of an error is as given in the proof of Theorem 4(b) replacing $T_d(\omega)$ and $T_h(\omega)$ by $T_d(\omega, x)$ and $T_h(\omega, x)$, respectively. By the slow errors property, $T_d(\omega, x) > T_d$ and $T_h(\omega, x) > T_h$. Hence, the expected response time of an error is bounded below by the same expression replacing $T_d(\omega, x)$ by T_d and $T_h(\omega, x)$ by $> T_h$.

Then, a computation analogous to the one in the proof of Theorem 4(b) shows that a sufficient condition for $E[RT \mid \rho(\omega) \neq d(\omega)] > E[RT \mid \rho(\omega) = d(\omega)]$ to hold is

$$\alpha(1-\alpha)(P_h(\omega) - P_d(\omega))(T_d(\omega) - T_h(\omega)) > 0.$$

Since h is faster than d (in the sense of Definition 5), we have $T_d > T_h$. Since h is also more consistent, we also have $P_h(\omega) > P_d(\omega)$. Thus part (b) holds. □

B Utility Estimation

To be able to define what is a correct answer and to control for strength of preference effects in our second experiment (certainty effect), we included 32 "neutral" lottery pairs to estimate the risk attitude of each participant out-of-sample. To guide the choice of lotteries for the utility estimation, we follow optimal design theory (Silvey, 1980) in the context of non linear (binary) models (Ford et al., 1992; Atkinson, 1996). Given that the experiment verged on the identification of the certainty bias, we also took care to estimate the utility function from lotteries not affected by this bias.

All $T = 32$ trials used for the utility estimation involved binary choices between lotteries of the form $A = (p, x)$ and $B = (q, y)$, where is A and B pay x with probability p and y with probability q, respectively, and 0 otherwise. We index the trials in the experiment by $t = 1, ..., 32$, that is, at trial t subjects face the choice between $A_t = (p_t, x_t)$ and $B_t = (q_t, y_t)$. Further, we index the $N = 128$ subjects by $i = 1, ..., N$. In the main analysis we assume a normalized constant relative risk aversion (CRRA), which is given by

$$u(x \mid r) = \begin{cases} \frac{x^{1-r}}{1-r}, & \text{if } r \neq 1 \\ \ln(x), & \text{if } r = 1 \end{cases}$$

However, the results are qualitatively unchanged when we assume utility function with constant relative risk aversion (CARA) instead. Under the assumption of Expected Utility maximization, subject i with utility function $u(x \mid r_i)$ chooses A_t over B_t if the difference in expected utilities is positive, that is,

$$\nabla_t(r_i) := p_t u(x_t \mid r_i) - q_t u(y_t \mid r_i) = p_t \frac{x_t^{1-r_i}}{1-r_i} - q_t \frac{y_t^{1-r_i}}{1-r_i} > 0. \tag{B.4}$$

In order to be able to estimate the parameters of the model, we now add noise to the model. We follow a standard approach based on a Random Utility Model (RUM). We add an error term $\epsilon_{it} \sim N(0, \sigma^2)$ with $\sigma^2 > 0$ to (B.4). That is, A_t is chosen if and only if

$$\nabla_t(r_i) + \epsilon_{it} > 0. \tag{B.5}$$

Define the binary choice indicator for trial t by

$$\gamma_{it} = \begin{cases} 1 & \text{if } A_t \text{ chosen by subject } i \\ -1 & \text{if } B \text{ chosen by subject } i. \end{cases}$$

The probability that subject i chooses A_t in trial t is

$$P\left(\gamma_{it} = 1\right) = P\left(\nabla_t(r_i) > -\epsilon_{it}\right) = \Phi\left(\frac{\nabla_t(r_i)}{\sigma}\right) \tag{B.6}$$

where Φ is the standard normal cumulative distribution function. Conversely, the probability of B is

$$P\left(\gamma_{it} = -1\right) = P\left(\epsilon_{it} < -\nabla_t(r_i)\right) = \Phi\left(-\frac{\nabla_t(r_i)}{\sigma}\right). \tag{B.7}$$

The conditional probabilities described in Equations B.6 and B.7 above were derived conditional on a subject's risk parameter r_i. To account for individual heterogeneity, we assume that the risk parameter is distributed over the population and we estimate the parameters of this distribution (e.g., see Harless and Camerer, 1994; Harrison and Rutström, 2008; Bellemare et al., 2008; Von Gaudecker et al., 2011; Conte et al., 2011; Moffatt, 2015). This approach is standard by now, and is generally preferred to estimating risk parameters separately for each individual (e.g., see Hey and Orme, 1994). The main reason is that the former greatly reduces the degrees of freedom compared to individual-level estimates, avoiding possible overfitting problems (see Conte et al., 2011, for a more detailed discussion). Further, the population estimation allows to obtain estimates even for extremely risk-averse or risk-seeking subjects, which is sometimes impossible at the individual level due to identification issues (Moffatt, 2015). Last, population-level estimates are generally preferable when predicting the behavior of the same or a similar group of individuals (Conte et al., 2011).

Assume that the individual risk attitudes are distributed normally in the population (from which our subjects were drawn) according to

$$r \sim N(\mu, \eta^2).$$

Hence, the log-likelihood of a sample given by the matrix $\Gamma = (\gamma_{it})$ consisting of T trials and N subjects is

$$\log L = \sum_{i=1}^{N} \ln \int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left(\gamma_{it} \frac{\nabla_t(r)}{\sigma}\right) f(r \mid \mu, \eta) dr \tag{B.8}$$

where $f(r \mid \mu, \eta) = \frac{1}{\sqrt{2\pi\eta^2}} e^{-\frac{1}{2}\left(\frac{r-\mu}{\eta}\right)^2}$ is the density function of the risk parameter r.

The log likelihood can be maximized by standard methods to obtain the maximum likelihood estimates. Since the integral in the likelihood function does not have a closed-form solution, we approximate it using standard simulation techniques (see Train, 2009, for details). Specifically, we approximate this integral by the average

$$\frac{1}{H} \sum_{h=1}^{H} \left(\prod_{t=1}^{T} \Phi\left(\gamma_{it} \frac{\nabla_t(r_{ih}(\mu, \eta))}{\sigma}\right)\right) \tag{B.9}$$

using a sequence of H (transformed) Halton draws $(r_{i1}, \ldots, r_{iH})$ from $N(\mu, \eta^2)$ for each subject i (fixed across trials t). Halton draws, a by-now-standard procedure, are pseudo-random draws that ensure even coverage of the parameter space (e.g. avoiding clustering) using Halton sequences (Halton, 1960; Moffatt, 2015). For the estimation, we use the Stata implementation "mdraws" of this procedure (Cappellari and Jenkins, 2003) for a uniform distribution $U(0, 1)$ and transform the resulting sequence $(g_1, g_2, \ldots, g_H)$ to obtain draws $r_{ih}(\mu, \eta) = \mu + \eta\Phi^{-1}(g_h)$ from $N(\mu, \eta^2)$.

The maximum simulated likelihood (MSL) approach amounts to replacing the integral in (B.8) by (B.9) and then maximize the resulting function $\log \hat{L}$. Maximization of $\log \hat{L}$ is carried out using standard MLE routines to obtain the estimates $(\hat{\mu}, \hat{\eta}, \hat{\sigma})$. Given those estimates we obtain the posterior expectation of each subject's risk attitude $\hat{r}_i$ conditional on their T choices approximating the conditional expectation by

$$\hat{r}_i = E(r_i | \gamma_{i1}, ..., \gamma_{iT}) \approx \frac{\frac{1}{H}\sum_{h=1}^{H} r_{ih} \left(\prod_{t=1}^{T} \Phi\left(\gamma_{it}\frac{\nabla_t(r_{ih})}{\hat{\sigma}}\right)\right)}{\frac{1}{H}\sum_{h=1}^{H} \left(\prod_{t=1}^{T} \Phi\left(\gamma_{it}\frac{\nabla_t(r_{ih})}{\hat{\sigma}}\right)\right)} \tag{B.10}$$

where $r_{ih} = r_{ih}(\hat{\mu}, \hat{\eta})$. Given the estimated individual risk parameter $\hat{r}_i$, we obtain

$$\hat{u}_i(x) = \frac{1 - e^{-\hat{r}_i x}}{1 - e^{-\hat{r}_i x_{\max}}} \text{ for } \hat{r}_i \neq 0$$

as the estimated utility function of subject i.

C The Almost-Certainty Effect

There is mixed evidence regarding whether violations of expected utility based on the presence of a sure option depend on the special nature of certainty, or whether similar behavioral anomalies can be observed even in the absence of a completely risk-free option (Machina, 1982; Conlisk, 1989). Consequently, the certainty-effect experiment described in Section 4.2 also included 16 lottery pairs that targeted this "almost-certainty effect."

C.1 Design

This additional set of lotteries followed an example in Kahneman and Tversky (1979, Problems 7–8) aimed to trigger the common ratio effect without relying on certain outcomes. Instead, the lotteries involved almost-certain outcomes, i.e. outcomes with a probability smaller than but close to one. In contrast with the 32 Allais-double-pairs used in Section 4.2.4, pairs with almost-certain outcomes were not paired with a corresponding "neutral pair", but instead with a pair that offered a long-shot, that is, a lottery with a low probability of winning a relatively high amount of money.

In total, this part of the experiment involved 32 lottery pairs forming 16 Allais-double-pairs. The construction is analogous to the one underlying the common-ratio effect (Kahneman and Tversky, 1979). Specifically, given an outcome $X > 0$, a probability $0 < q < 1/2$, and an integer $k \geq 1$, we define for the pair of lotteries $R_3(k)$ and $S_3(k)$ by

$$S_3(k) = \begin{cases} X/2 \text{ €} \text{ with prob.} & (2q)/k \\ 0 \text{ €} & \text{otherwise} \end{cases} \qquad R_3(k) = \begin{cases} X \text{ €} \text{ with prob.} & q/k \\ 0 \text{ €} & \text{otherwise} \end{cases}$$

That is, $S_3(k)$ offers half the outcome with double the probability compared to $R_3(k)$, and is hence safer. For $k = 1$ and q close to but smaller than 0.5, the probabilities of winning in $S_3(k)$ are close to one, whereas for a large $k > 1$ the winning probabilities are minuscule in both lotteries. For instance (abusing notation as in the main text), given $S_3(1) = (0.9, 15)$ and $R_3(1) = (0.45, 30)$, we obtain $S_3(15) = (0.06, 15)$ and $R_3(15) = (0.03, 30)$. This reflects the argument in Kahneman and Tversky (1979) that, in the first pair, "the probabilities of winning are substantial [...], and most people choose the prospect where winning is more probable," while, in the second pair, "there is a *possibility* of winning, although the probabilities of winning are minuscule [...] in both prospects. In this situation where winning is possible but not probable, most people choose the prospect that offers the larger gain." Of course, such behavior contradicts expected utility theory, which implies that a decision maker should choose $S_3(1)$ over $R_3(1)$ if and only if he or she also chooses $S_3(k)$ over $R_3(k)$.

We constructed 16 *almost certainty* pairs from the family above, with $k = 1$ and q close to 0.5. For every pair of lotteries $(R_3(1), S_3(1))$ we included a *neutral* pair $(R_3(k), S_3(k))$ with k large enough to ensure that the probability of winning was below 10% for both lotteries. Note that the neutral pair could also be seen as a *long-shot* pair because the involved lotteries have very small probabilities of relatively large outcomes.

We apply the model presented in Section 2 to the almost-certainty effect and test its predictions. The deliberative process d is again assumed to be maximization of expected utility. This process is always active and prescribes to choose the option with the larger expected utility. The heuristic process is assumed to be an "almost-certainty" heuristic h, which is triggered by lotteries involving outcomes with probability close to one. If h is active, it prescribes to choose the safer (almost-certain) lottery. Hence, this is a heuristic triggered by alternative S, as described in Section 4. Strength-of-preference effects are operationalized based on the cardinal difference between certainty equivalents of the options derived from the individually-estimated expected utilities (recall Section 4.2.3).

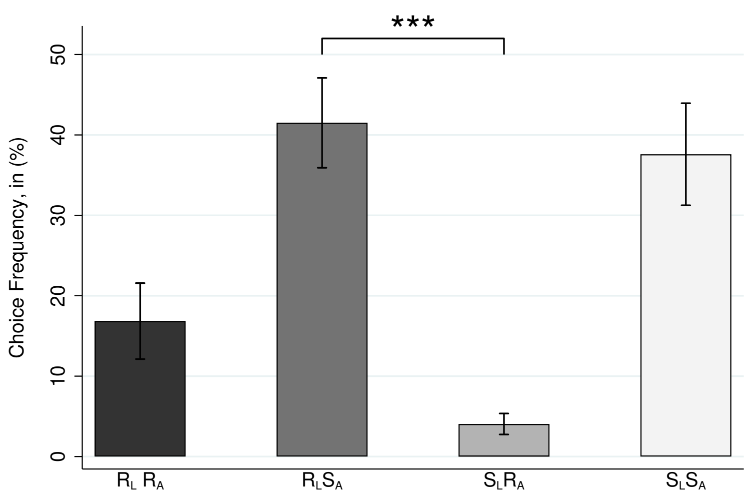


Figure C.1: Choice frequencies in double pairs showing the almost-certainty effect.

C.2 Results

We first check whether our data replicates the almost-certainty effect. For an almost-certainty/long-shot double pair we denote by $D_L \in \{R_L, S_A\}$ and $D_A \in \{R_A, L_A\}$ the decision in the long-shot and almost-certainty pair, respectively. Thus, $R_L S_A$ indicates that the risky lottery was chosen in the long-shot pair and the safe lottery was chosen in the almost-certainty pair. The almost-certainty effect obtains when reversals of the form $R_L S_A$ are more frequent than reversals of the opposite type, $S_L R_A$. Figure C.1 illustrates the distribution of choices for all 128 subjects for the 16 double pairs. Choices were inconsistent for 45.56% of all double pair choices. These violations are systematic with inconsistencies of the form $R_L S_A$ (41.50%) occurring significantly more than the opposite pattern $S_L R_A$ (4.05%; WSR test, $N = 128$, $z = 8.628$, $p < 0.001$). This indicates that our subjects exhibit the typical behavioral asymmetry associated with the almost-certainty effect.

We now turn to a direct empirical test of our model, which allows to separate noise (strength-of-preference) from an actual bias towards almost-certain outcomes (almost-certainty heuristic). Recall that a choice is called "correct" if it maximizes (estimated) expected utility. Theorem 1 makes predictions about the relationship between $\Delta^i_\omega = CE_i(S) - CE_i(R)$ and the proportion of correct choices. The illustrations below again rely on a binning procedure similar to the one used in Section 4.2.4.

As Figure C.2 shows, we observe the monotonicities predicted by Theorem 1 for alignment, conflict, and neutral situations. Reflecting strength of preference effects, the proportion of correct choices increases with distance to indifference, both for alignment decision problems ($\Delta^i_\omega > 0$) and for conflict ones ($\Delta^i_\omega < 0$). Reflecting process multiplicity, and as predicted by Theorem 1, there is a clear jump at $\Delta^i_\omega = 0$. For neutral pairs, the monotonicity due to strength of preference can also be observed, but, as expected, there is

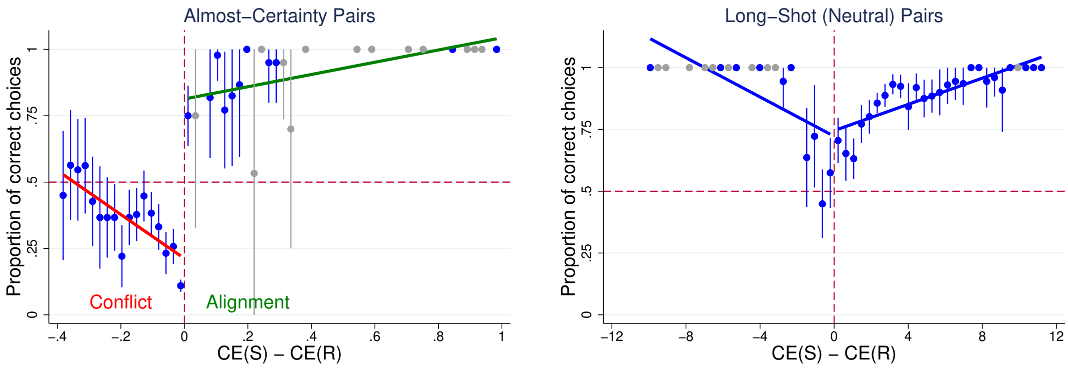


Figure C.2: Proportion of correct choices as a function of $\Delta_\omega^i = CE_i(S) - CE_i(R)$ separately for certainty/DP pairs (left) and normal/neutral pairs (right).

no jump at $\Delta_\omega^i = 0$. To quantify the prediction, we test for an upward jump at zero using a local polynomial regression discontinuity estimation (Calonico et al., 2014) as in Section 4.2.4. The results show a statistically-significant jump at zero for almost-certainty pairs (39.73%, $z = 6.301$, $p < 0.001$) but not for neutral pairs (10.43%, $z = 1.135$, $p = 0.256$). Further, as predicted by Corollary 1(c), the jump at zero is symmetrically distributed between the left- and the right-hand parts. The limits of the probabilities of correct responses as Δ approaches zero, conditional on $\Delta < 0$ or $\Delta > 0$ are $P_- = 0.258$ and $P_+ = 0.750$, respectively. Thus $(1/2) - P_- = 0.242 \simeq 0.250 = P_+ - (1/2)$.

This test is also interesting for a different reason. Following Kahneman and Tversky (1979), one could conclude that, in the case of long-shot pairs, decision makers might be following a "long-shot heuristic" where minuscule probabilities are ignored and the lottery with the largest non-zero outcome, i.e., lottery R, is picked. If this were the case, neutral pairs would actually reflect dual-process effects with respect to this heuristic. Specifically, in Figure C.2, decision problems with $\Delta_\omega^i = CE_i(S) - CE_i(R) > 0$ would be in conflict, as the hypothesized long-shot heuristic always favors R. Analogously, decision problems with $\Delta_\omega^i = CE_i(S) - CE_i(R) < 0$ would be in alignment. Hence, Theorem 2 would predict a discontinuity at zero, specifically a jump down in the frequency of correct answers. This is not observed in the data, and hence we conclude that there is no evidence for a long-shot heuristic in the data. However, an examination of Figure C.2 shows an asymmetry between negative and positive values of Δ, with bins far away from zero on the negative side exhibiting almost no errors. This might be consistent with a long-shot heuristic playing a role for a subset of decision makers.

Table C.1 reports the results of random effects Probit regressions on the probability of a correct answer, which is the equivalent of Table 5 in the main text for almost-certainty pairs instead of certainty pairs. Models 1-3 consider only alignment, conflict, or neutral decision problems, respectively, with the absolute difference in certainty equivalents ($|\Delta_\omega^i|$) as an independent variable. For alignment and conflict we find significant strength-of-preference effects, that is, the probability of a correct answer is increasing in the absolute difference in certainty equivalents. However, we do not find psychometric effects for neutral situations (model 3).

Model 4 considers all decision problems with $\Delta_\omega^i > 0$, which includes both decision problems in alignment and neutral decision problems where the deliberative response should be S. The dummy for alignment decision problems is significantly positive, showing that the almost-certainty heuristic shifts up the probability of a correct response

Table C.1: Random effects Probit regressions on correct answers with a dummy for positive differences in certainty equivalents ($\Delta^i_\omega > 0$).

Correct	Alignment (1)	Conflict (2)	Neutral (3)	$\Delta^i_\omega > 0$ (4)	$\Delta^i_\omega < 0$ (5)	Certainty (6)	Neutral (7)
$\lvert\Delta\rvert$	6.218**	3.226***	0.962			6.251**	−2.315
	(2.741)	(0.451)	(2.980)			(2.662)	(3.288)
Alignment				0.771***			
				(0.183)			
Conflict					−0.556***		
					(0.046)		
$\Delta > 0$ (dummy)						1.250*	−11.143
						(0.646)	(6.933)
$\lvert\Delta\rvert \times (\Delta > 0)$						−3.026	71.969
						(2.699)	(51.542)
Constant	−0.758	0.483***	0.262	0.714*	0.722***	−0.767	2.109**
	(0.668)	(0.097)	(0.445)	(0.389)	(0.071)	(0.641)	(0.824)
Linear combination tests							
$\lvert\Delta\rvert + \lvert\Delta\rvert \times (\Delta > 0)$						3.225***	69.654
						(0.451)	(51.320)
Observations	208	1840	2048	416	3680	2048	2048

Notes: Robust standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

in this case (Theorem 1(b1)). Conversely, Model 5 considers all decision problems with $\Delta^i_\omega < 0$, which includes both decision problems in conflict and neutral decision problems where the deliberative response should be R. The dummy for conflict decision problems is significantly negative, showing that the almost-certainty heuristic shifts down the probability of a correct response in this case (Theorem 1(b2)).

Model 6 considers all (almost-)certainty pairs, i.e. all decision problems in either conflict or alignment. The regression controls for a dummy for $\Delta^i_\omega > 0$, which in this case corresponds to decision problems in alignment. The coefficient for $\left|\Delta^i_\omega\right|$ is significantly positive, which again shows the strength-of-preference effect for decisions under conflict. The linear combination test including the interaction between the dummy and $\left|\Delta^i_\omega\right|$ is significant and positive, showing the strength-of-preference effect for decision problems under alignment. The dummy itself is positive (albeit marginally significant, $p = 0.055$), reflecting the upper jump at zero. However, we remark again that this is not the correct way of testing for a discontinuity, as it merely reflects the difference in intersect points of linear trends estimated above and below zero, and might be misleading if the data displays a nonlinear relation to Δ. The correct way to test for a discontinuity is the procedure we described above (Calonico et al., 2014).

Model 7 is the analogue of Model 6 but considering only neutral decision problems. That is, there is neither conflict nor alignment. In contrast to Figure C.2, we do not find evidence for strength-of-preference effects for these neutral decisions. As expected, the dummy ($\Delta^i_\omega > 0$) for the discontinuity between positive and negative CEs values is not significant.

We now turn to the analysis of response times. Figure C.3 provides a simple graphical illustration of the empirical support for Theorem 3 for the almost-certainty effect.

The left panel plots the difference between the average response time of correct answers in alignment ($\Delta^i_\omega < 0$) and conflict ($\Delta^i_\omega > 0$) following a binning procedure, as in Figure 8. As predicted by Theorem 3(a), most points are above the 0 horizontal line, indicating that correct answers are faster in alignment than in comparable conflict situations (i.e., controlling for strength of preference). The right panel is the equivalent graphical representation for neutral (long-shot) pairs, where the sign of Δ does not

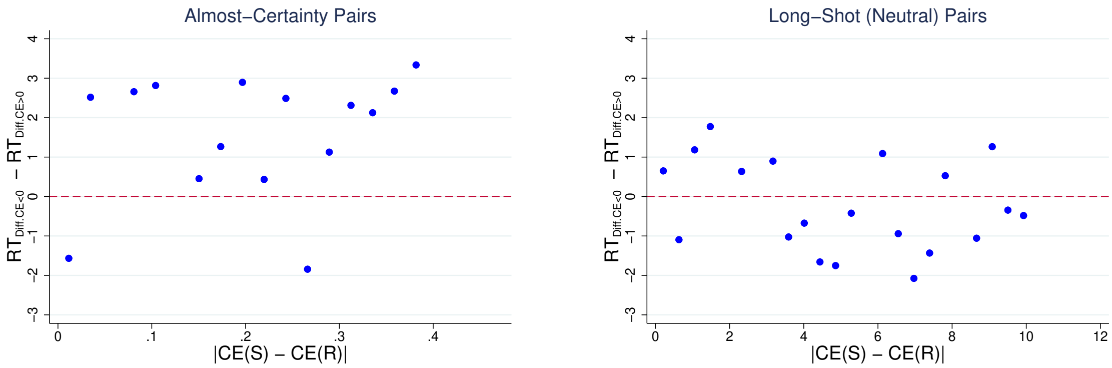


Figure C.3: Difference in response times for correct answers between negative CE differences ($\Delta^i_\omega < 0$) and postive CE differences ($\Delta^i_\omega < 0$) as a function of the binned distance in CEs ($|\Delta^i_\omega|$), separately for certainty pairs (left) and neutral pairs (right).

correspond to conflict or alignment. We see that the distribution of observations is more dispersed around the zero line, with data points both below and above it.

Figure C.4 illustrates Theorem 4 and its empirical support when the almost-certainty heuristic is triggered. As in the case of the certainty effect discussed in the main text, all decision problems in conflict (resp. alignment) have negative (resp. positive) values of Δ. Using the same binning procedure as in Figure 6, the left-hand panel of the figure plots the difference between the average response time of correct answers and errors against the average difference in certainty equivalents for each bin. In case of conflict, which corresponds to $\Delta^i_\omega < 0$, this difference is positive for all bins, indicating that indeed errors are faster than correct answers as predicted (Theorem 4(a)). In stark contrast, for alignment situations, that is, bins with $\Delta^i_\omega > 0$, the difference is negative for almost all bins, showing that errors are slower than correct answers (Theorem 4(b)). To ensure that this asymmetry in response time differences is evidence for process multiplicity and cannot be explained by other elements of the experimental design, the right panel of Figure C.4 shows the analogous graph for neutral (long-shot) pairs. For those pairs, no effects of process multiplicity should be present (and the sign of Δ_ω does not indicate conflict or alignment). Indeed, as expected, we no longer observe an asymmetric pattern of response time differences in this case (although there are few observations with $\Delta < 0$).

We then turn to linear random-effects regressions on log-transformed response times. Table C.2 has the same structure as Table 7. Models 1-3 analyze response times for certainty pairs, which are either in conflict or alignment. Models 4-6 are the corresponding analyses for neutral pairs, where the sign of Δ does not correspond to conflict or alignment.

Models 1 and 4 concentrate on possible strength-of-preference (chronometric) effects. These are clear for neutral pairs (Model 4), but not for almost-certainty pairs (Model 1). Model 2 concentrates on possible dual-process effects. For certainty pairs, in Model 2 the dummy for $\Delta > 0$ (alignment) is significant and negative, confirming the prediction of Theorem 3 (generalized Stroop effect). That is, correct decisions in conflict are slower than correct decisions in alignment. The dummy for errors is significantly negative, showing that errors in conflict are faster than correct responses, as predicted by Theorem 4(a). In contrast, the linear combination test Error+Error $\times(\Delta > 0)$ is significantly positive, showing that errors in alignment are slower than correct responses, as predicted by Theorem 4(b). Crucially, Model 3 shows that all these results hold when we control

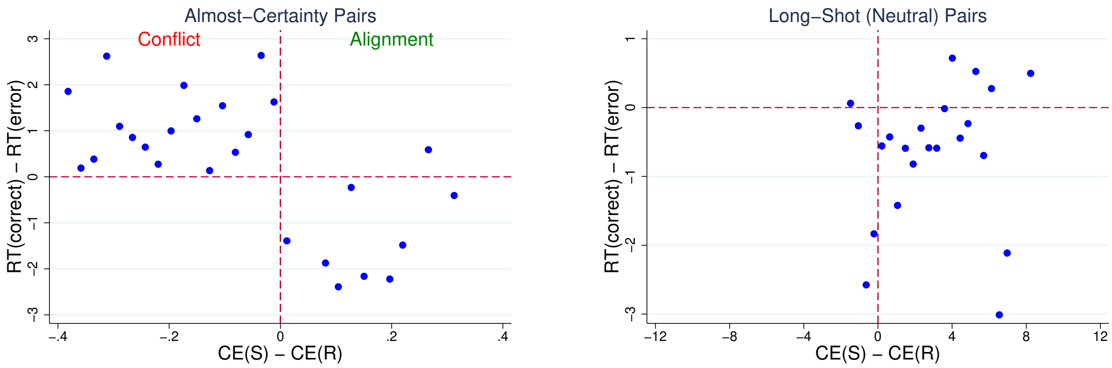


Figure C.4: Average difference in response times between correct choices and errors as a function of $\Delta^i_\omega = CE_i(S) - CE_i(R)$ (binned), separately for certainty pairs (left) and neutral pairs (right).

for strength of preference, that is, the predictions of Theorems 3 and 4, i.e. the effects arising from process multiplicity, are not confounded with chronometric effects.

Models 5 and 6 are the analogues of Models 2 and 3 for neutral pairs, where there should be no dual-process effects. Indeed, there is no Stroop-analogue effect, and the dummy for $\Delta > 0$ is instead significantly positive, and non-significant once one controls for strength of preference. We observe generally slower errors in neutral situations, independently of the sign of Δ. Model 6 further shows that the chronometric effects detected in Model 4 are robust to controlling for asymmetries arising from the sign of Δ.

Table C.2: Random effect regression on log-transformed response times.

Log.RT	Certainty pairs			Neutral pairs		
	(1)	(2)	(3)	(4)	(5)	(6)
$\|\Delta\|$	1.597		3.131	-0.095***		-0.091***
	(5.143)		(5.169)	(0.028)		(0.028)
$\Delta > 0$ (dummy)	-0.661	-1.578***	-1.192**	1.078**	1.058**	0.788
	(0.464)	(0.440)	(0.492)	(0.496)	(0.470)	(0.523)
$\|\Delta\| \times (\Delta > 0)$	-1.983		-3.493	0.055		0.077
	(5.148)		(5.173)	(0.091)		(0.091)
Error		-0.528***	-0.540***		0.627***	0.608***
		(0.192)	(0.193)		(0.176)	(0.176)
Error $\times(\Delta > 0)$		2.195***	2.141***		0.499	0.516
		(0.812)	(0.813)		(0.478)	(0.479)
Constant	3.565***	3.884***	3.837***	3.780***	3.366***	3.673***
	(0.158)	(0.168)	(0.185)	(0.167)	(0.146)	(0.170)
Linear combination tests						
$\|\Delta\| + \|\Delta\| \times (\Delta > 0)$	-0.386*		-0.363	-0.040		-0.014
	0.221		(0.222)	(0.086)		(0.087)
Err. + Err. $\times(\Delta > 0)$		1.667**	1.601**		1.126**	1.124**
		(0.789)	(0.790)		(0.444)	(0.445)
Observations	2048	2048	2048	2048	2048	2048

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

D Lottery pairs (Certainty Experiment)

The tables below list the actual lottery double-pairs (and their expected values) used in our second experiment. Table D.1 lists the common-ratio lotteries (S_1 and R_1 in the main text), Table D.2 give the common-consequence lotteries (S_2 and R_2 in the main text), and Table D.3 details the almost-certainty lotteries (S_3 and R_3) described in Section 6.3 and Appendix C. Last, Table D.4 lists the lotteries used for the individual estimation of utilities described in Section 4.2.3 and Appendix B.

Table D.1: Common-ratio double pairs used in Experiment 2.

Pair Nr	p	S	$EV(R)$	q	Y	$EV(S)$
33	0.85	20	17	1	15	15
34	0.28	20	5.67	0.33	15	5
35	0.85	19	16.15	1	13	13
36	0.21	19	4.04	0.25	13	3.25
37	0.85	18	15.3	1	12	12
38	0.17	18	3.06	0.2	12	2.40
39	0.85	17	14.45	1	14	14
40	0.14	17	2.41	0.17	14	2.33
41	0.8	26	20.8	1	18	18
42	0.27	26	6.93	0.33	18	6
43	0.8	25	20	1	16	16
44	0.2	25	5	0.25	16	4
45	0.8	24	19.2	1	15	15
46	0.16	24	3.84	0.2	15	3
47	0.8	23	18.4	1	17	17
48	0.13	23	3.07	0.17	17	2.83
49	0.75	27	20.25	1	18	18
50	0.25	27	6.75	0.33	18	6
51	0.75	26	19.5	1	16	16
52	0.19	26	4.88	0.25	16	4
53	0.75	25	18.75	1	15	15
54	0.15	25	3.75	0.2	15	3
55	0.75	24	18	1	17	17
56	0.13	24	3	0.17	17	2.83
57	0.7	35	24.5	1	20	20
58	0.23	35	8.17	0.33	20	6.67
59	0.7	34	23.8	1	18	18
60	0.18	34	5.95	0.25	18	4.50
61	0.7	33	23.1	1	17	17
62	0.14	33	4.62	0.2	17	3.40
63	0.7	32	22.4	1	19	19
64	0.12	32	3.73	0.17	19	3.17

Table D.2: Common-consequence double pairs.

Pair Nr	p	X	$EV(R)$	q	Y	$EV(S)$
1	0.52	28	14.56	0.56	25	14
2	0.96	28	26.88	1	25	25
3	0.58	26	15.08	0.61	24	14.64
4	0.97	26	25.22	1	24	24
5	0.62	24	14.88	0.66	22	14.52
6	0.96	24	23.04	1	22	22
7	0.66	23	15.18	0.7	21	14.7
8	0.96	23	22.08	1	21	21
9	0.74	22	16.28	0.77	19	14.63
10	0.97	22	21.34	1	19	19
11	0.76	21	15.96	0.8	18	14.4
12	0.96	21	20.16	1	18	18
13	0.79	20	15.8	0.85	17	14.45
14	0.94	20	18.8	1	17	17
15	0.85	19	16.15	0.9	16	14.4
16	0.95	19	18.05	1	16	16
17	0.42	34	14.28	0.47	28	13.16
18	0.95	34	32.3	1	28	28
19	0.47	32	15.04	0.52	25	13
20	0.95	32	30.4	1	25	25
21	0.52	33	17.16	0.56	27	15.12
22	0.96	33	31.68	1	27	27
23	0.57	28	15.96	0.59	24	14.16
24	0.98	28	27.44	1	24	24
25	0.62	29	17.98	0.65	25	16.25
26	0.97	29	28.13	1	25	25
27	0.64	28	17.92	0.68	24	16.32
28	0.96	28	26.88	1	24	24
29	0.68	27	18.36	0.73	23	16.79
30	0.95	27	25.65	1	23	23
31	0.76	26	19.76	0.78	23	17.94
32	0.98	26	25.48	1	23	23

Table D.3: Long-shot and almost-certainty pairs.

Pair Nr	p	X	$EV(R)$	q	Y	$EV(S)$
65	0.49	26	12.74	0.98	13	12.74
66	0.03	26	0.78	0.06	13	0.78
67	0.47	28	13.16	0.94	14	13.16
68	0.03	28	0.84	0.06	14	0.84
69	0.45	30	13.5	0.9	15	13.5
70	0.03	30	0.9	0.06	15	0.9
71	0.42	32	13.44	0.84	16	13.44
72	0.03	32	0.96	0.06	16	0.96
73	0.4	34	13.6	0.8	17	13.6
74	0.01	34	0.34	0.02	17	0.34
75	0.38	36	13.68	0.76	18	13.68
76	0.01	36	0.36	0.02	18	0.36
77	0.36	38	13.68	0.72	19	13.68
78	0.01	38	0.38	0.02	19	0.38
79	0.33	40	13.2	0.66	20	13.2
80	0.01	40	0.4	0.02	20	0.4
81	0.48	10	4.8	0.96	5	4.8
82	0.04	10	0.4	0.08	5	0.4
83	0.46	12	5.52	0.92	6	5.52
84	0.04	12	0.48	0.08	6	0.48
85	0.44	14	6.16	0.88	7	6.16
86	0.04	14	0.56	0.08	7	0.56
87	0.41	16	6.56	0.82	8	6.56
88	0.04	16	0.64	0.08	8	0.64
89	0.39	18	7.02	0.78	9	7.02
90	0.02	18	0.36	0.04	9	0.36
91	0.37	20	7.4	0.74	10	7.4
92	0.02	20	0.4	0.04	10	0.4
93	0.35	22	7.7	0.7	11	7.7
94	0.02	22	0.44	0.04	11	0.44
95	0.32	24	7.68	0.64	12	7.68
96	0.02	24	0.48	0.04	12	0.48

Table D.4: List of lotteries used for the utility estimation.

Pair Nr	p	X	EV	q	Y	EV
1	0.05	20	1	0.8	10	8
2	0.1	20	2	0.7	10	7
3	0.2	20	4	0.8	10	8
4	0.3	20	6	0.8	10	8
5	0.1	30	3	0.8	5	4
6	0.4	20	8	0.8	10	8
7	0.3	30	9	0.8	10	8
8	0.2	25	5	0.8	5	4
9	0.4	20	8	0.7	10	7
10	0.2	20	4	0.6	5	3
11	0.4	30	12	0.8	10	8
12	0.2	25	5	0.6	5	3
13	0.4	20	8	0.6	10	6
14	0.5	30	15	0.8	10	8
15	0.3	20	6	0.65	5	3.25
16	0.3	25	7.5	0.6	5	3
17	0.4	30	12	0.8	5	4
18	0.5	20	10	0.6	10	6
19	0.4	25	10	0.7	5	3.5
20	0.6	20	12	0.7	10	7
21	0.7	20	14	0.8	10	8
22	0.7	30	21	0.8	10	8
23	0.5	25	12.5	0.7	5	3.5
24	0.6	25	15	0.8	5	4
25	0.5	20	10	0.65	5	3.25
26	0.5	25	12.5	0.6	5	3
27	0.6	25	15	0.7	5	3.5
28	0.7	25	17.5	0.8	5	4
29	0.5	25	12.5	0.55	5	2.75
30	0.65	30	19.5	0.7	5	3.5
31	0.75	25	18.75	0.8	5	4
32	0.77	25	19.25	0.8	5	4

E Experimental Instructions

E.1 Instructions for Experiment 1: Conservatism

[*These are the on-screen instructions for Experiment 1 given to subjects during the experiment. The original instructions were in English. Text in brackets [...] was not displayed to subjects.*]

Welcome

Please read the instructions carefully.

When you have completely read a page, you can use the arrow-keys on the keyboard to go to the next page. The end of the instructions are indicated by the "Done" button.

Please note the following rules:

1. Please do not talk to other participants.
2. If you have a question, please raise your hand. We will answer your question privately.
3. Please do not use any feature of the computer that is not part of the experiment.

General Instructions

The experiment consists of multiple parts:

1. The instructions and some understanding questions about the instructions (you are currently reading them).
2. A decision task which consists of **300 rounds divided in 4 blocks of 75 rounds**. There will be a short break of 30 seconds after each block.
3. A demographic questionnaire.
4. Information about your performance and your final payment depending on your decisions.

The following pages will explain the decision task and how you can earn money in this experiments.

One Round of the Decision Task (1/4)

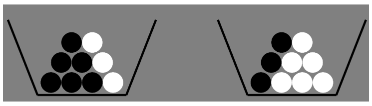
The general procedure of a round is that the computer randomly selects one of two urns and it is **your task to guess which urn the computer selected**. **For each correct guess you will earn money**, for each incorrect guess you do not earn any money. All details will be explained on this and the following pages.

The Urns:
There are always **two urns**, a left urn and a right urn. Both urns contain **either 8, 9, or 10 colored balls**. Each urn will have **some black balls and some white balls**.

At the beginning of each block, both urns will be displayed revealing the number of balls in each urn and the composition of black and white balls.

Below is an example of the urn composition with 9 balls (see Figure below). The left urn consists of **6 black and 4 white** balls and the right urn consists of **3 black and 6 white** balls.

The computer will randomly **select one of the urns** (how the computer selects the urn will be explained later).

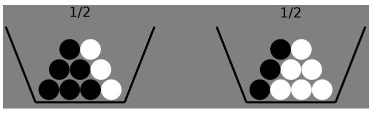


One Round of the Decision Task (2/4)

Urn Selection:
A the beginning of each round the computer randomly selects one of the urns. Above each urn you will see the probability (in fractions) that each urn is selected. Below is an example of "1/2-1/2" but **other probabilities (fractions) may occur during the experiment**.

The fraction "1/2" means "1 out of 2" and you can visualize the likelihood of an urn being selected by imagining a box of raffle tickets. For this example, there are 2 raffle tickets in the box, on 1 there is written "left" and on the other one is written "right".

Then all raffle tickets are shuffled around and one ticket is randomly picked from the box. If the randomly drawn ticket has "left" written on it, then the computer selects the left urn. If the ticket has "right" written on it, then the computer selects the right urn.

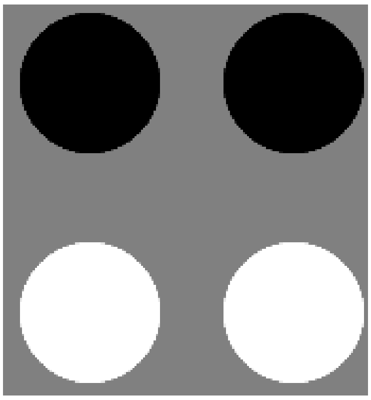


One Round of the Decision Task (3/4)

The Sample:
After the computer selected one of the urns, you have to press the "Space bar." Then the computer will **randomly draw balls with replacement** from the selected urn.

"**With replacement**" means that the computer draws a ball, writes down the color, puts the ball back in the urn, and then draws again from the same urn while the full number of colored balls (either 8, 9, 10 balls). That means the computer always draws from the same urn in the same round with the same number of balls.

The computer always **draws a sample of 4 balls** from the urn. For sake of clarity we show the number of black balls in one row and the number of white balls in a second row. The example draw below shows there were two black and two white balls drawn from the urn.



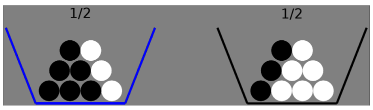
One Round of the Decision Task (4/4)

Your Decision:

After you have seen the sample you have to **guess from which urn the sample was drawn**. You indicate your choice by pressing the “F” key for choosing the left urn or pressing the “J” key for choosing the right urn. The computer will briefly confirm your choice by highlighting the urn you chose in blue (see example below where the left urn was chosen).

The computer will check whether you correctly guessed from which urn the sample was drawn. But the computer will not inform you whether or not you selected the correct urn.

After you made your choice, a new round will start directly. That means the computer will select one of the two urns and wait for you to press the “Spacebar” to draw the sample.



Your Payment

The computer keeps track after each round whether you **guessed correctly** the urn the sample was drawn from.

You will be paid for the total number of correct guesses. For each correct guess, you will earn **0.15 CHF**. At the end of the experiment, we will show you an overview of the number of correct guesses and how much you earned from your decisions.

E.2 Instructions for Experiment 2: The Certainty Effect

[*These are the written instructions for Experiment 2 given to subjects during the experiment. The original instructions were in German. Text in brackets [...] was not displayed to subjects.*]

General Instructions

Welcome! In this experiment you will be asked to make a series of decisions that will determine your earnings at the end of the experiment. The total duration of the experiment is about 1 hour.

If you have a question, please raise your hand and remain seated. An experimenter will come and answer your question.

It is important, that you read the instructions carefully before you make your decisions.

During the experiment you are not allowed to talk or communicate in any other way with the other participants. If you violate this rule, you might be excluded from the experiment.

We now explain the general course of the experiment: The experiment consists of a decision part and a questionnaire. In the decision part you have to make multiple decisions.

In the decision part, you can earn money. How much money you earn will depend on your decisions in that part and chance. Your earnings in the decision part will be paid to you anonymously and in cash at the end of the experiment. In addition to this amount you will receive €4 for your participation in the experiment.

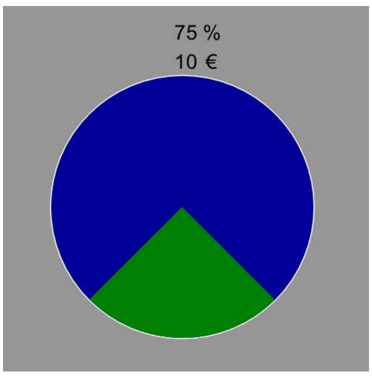
Below you will find further general information for the experiment. Once you have read them, please proceed to answer the comprehension questions on page three.

You will receive further specific instructions for the decision part directly before the beginning of the decision part.

Instructions: Lotteries In the decision part of the experiment you will be asked to make decisions about lotteries. Hence, we will now explain in detail what a lottery is:

A lottery consists of two potential outcomes, each of which will occur with a given probability. One of the two outcomes is always €0 (zero). The other outcome will differ from lottery to lottery. If a lottery is played out, this means that you will receive exactly one of the two possible outcomes (in Euro).

In the experiment lotteries will be represented by pie charts as in the example below. The colored areas of the pie chart illustrate the probabilities for the two corresponding outcomes.



Example:

The pie chart depicted above is an example of how we present a lottery. In this example, the lottery pays €10 with a probability of 75%, which is represented by the blue area. Additionally, this information is also shown numerically above the pie chart. Accordingly, the lottery pays €0 with a probability of 25%, which is represented by the green area. The second outcome is always €0 and occurs with the remaining probability. Please note that this information is not repeated numerically on screen.

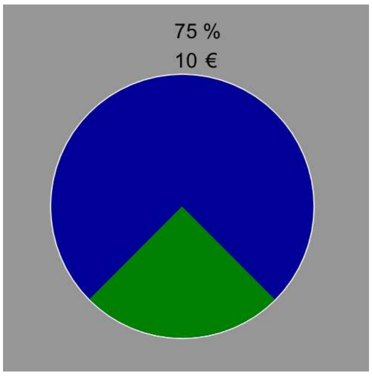
If a lottery is played out, this means that it will pay exactly one of the two outcomes. In the example above, the lottery pays €10 with a probability of 75% and €0 with the remaining probability of 25%.

Please note that the lottery shown above is only an example. The lotteries in the experiment will have different outcomes and probabilities.

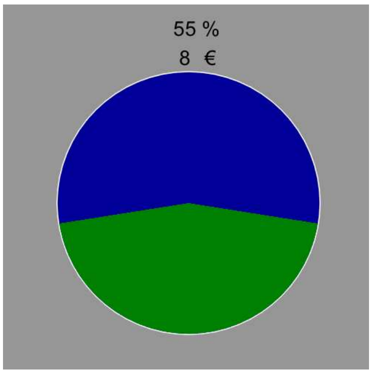
If you have a question, please raise your hand. If you have no further questions, you may proceed to the comprehension questions on the next page.

Comprehension Questions: Lotteries

Below you see examples of two lotteries, similar to the ones you will face later on in the experiment. Please note that these lotteries are only examples.



Lottery A



Lottery B

Please answer the following comprehension questions:

1. What is the probability that Lottery A pays €10?
2. What is the probability that Lottery B pays €0?
3. Which amount does Lottery A pay with a probability of 25%?
4. Which amount does Lottery B pay with a probability of 55%?

Once you have answered all comprehension questions, please raise your hand. An experimenter will then check your answers.

Specific instructions for the decision part

[These are the translated instructions for the decision part, which were distributed after all participants have answered the comprehension questions correctly. The original instructions were in German. Text in brackets [...] was not displayed to subjects.]

Your decisions: In this part of the experiment you will be presented with a series of lottery pairs. Your task is to choose one of the two lotteries from each pair.

On the screen you will see a lottery pair (consisting of two lotteries) represented by two pie charts. One of the lotteries will be shown on the left and the other will be shown on the right. You choose one of the lotteries by pressing the left or right arrow key on your keyboard. These keys are marked with a yellow sticker. To choose the lottery on the left, press the left arrow key "←." To choose the lottery on the right, press the right arrow key "→." Please note that your decisions will affect your earnings at the end of the experiment (a detailed description of how your earnings are determined will follow below).

There are no wrong or correct decisions. When you choose one of the lotteries, this simply shows that you prefer to play this lottery over the other lottery.

After you have made your decision, you will see the next lottery pair. In the decision part you will be presented with a total of 132 lottery pairs. After you have made a decision for each of the pairs, the decision part ends and we will continue with a short questionnaire.

How we determine your earnings in the decision part After you have made a decision for each of the lottery pairs, the computer will randomly select **one of the 132** lottery pairs. The computer then checks which of the two lotteries you have chosen for this randomly selected pair. The lottery you have chosen will be played out. The outcome of the lottery determines your earnings for the decision part.

If you have any further questions, please raise your hand and remain seated.