Errors, fast and slow

Ludwig, Jonas; Ahrens, Fabian K.; Achtziger, Anja

Published in:
Thinking & reasoning

DOI:
10.1080/13546783.2020.1781691

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ptar20

Errors, fast and slow: an analysis of response times in probability judgments

Jonas Ludwig , Fabian K. Ahrens & Anja Achtziger

THINKING & REASONING
2020, VOL. 26, NO. 4, 627–639
https://doi.org/10.1080/13546783.2020.1781691

Routledge
Taylor & Francis Group

REPORT

Errors, fast and slow: an analysis of response times in probability judgments

Jonas Ludwig[a,b], Fabian K. Ahrens[a,c] and Anja Achtziger[a,d]

[a]Department of Political and Social Sciences, Zeppelin University, Friedrichshafen, Germany; [b]Department of Psychology, University of Education Weingarten, Weingarten, Germany; [c]Faculty of Economics and Business, University of Groningen, Groningen, Netherlands; [d]Department of Psychology, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

ABSTRACT
Based on the Dual-Process Diffusion Model, we tested three hypotheses about response times of errors and correct responses in probability judgments. We predicted that correct responses were (1) slower than errors in the case of conflicting decision processes but (2) faster than errors in the case of alignment; and that they were (3) slower in the case of conflict than in the case of alignment. A binary-choice experiment was conducted in which three types of decision problems elicited conflict or alignment of a deliberative decision process and a heuristic decision process. Consistent with the traditional dual-process architecture, the former captured computational-normative decision strategies and the latter described intuitive-affective aspects of decision making. The hypotheses (1) and (3) were supported, while no statistically significant evidence was found for (2). Implications for the generalisability of the Dual-Process Diffusion Model to slow probability judgments are discussed.

ARTICLE HISTORY Received 2 December 2019; Accepted 6 June 2020

KEYWORDS Response time; dual-process diffusion model; probability judgment

Probabilistic reasoning is heavily investigated in decision research. Violations of probability theory have been demonstrated numerously, for instance, the tendency to overestimate the joint probability of conjunct events (conjunction fallacy; Fisk, 2017) or the neglect of base-rate probabilities (Pennycook & Thompson, 2017). An ongoing theoretical debate revolves around the cognitive processes involved in probability judgments. The focal question is whether dual-process theories or a unified framework

CONTACT Jonas Ludwig jonas.ludwig@uni-wuerzburg.de

Supplemental data for this article can be accessed online at https://doi.org/10.1080/13546783.2020.1781691

prevail in explaining the observed judgment patterns. Dual-process theories (Alós-Ferrer & Strack, 2014; Evans & Stanovich, 2013) posit that information processing is either associative and automatically triggered by a stimulus, or computational and deliberately executed by cognitive control. The unified approach holds that this distinction is not warranted (Kruglanski & Gigerenzer, 2011). In this article, we provide evidence from a binary-choice probability judgment task for the validity of a recent theoretical framework that combines the dual-process perspective with core features of a single-process model (Alós-Ferrer, 2018).

The dual-process diffusion model

The Dual-Process Diffusion Model (DPDM; Alós-Ferrer, 2018) belongs to the family of sequential sampling models (see Forstmann et al., 2016). It is a parsimonious formal-analytical model for response times (RTs). Analogous to the standard architecture of dual-process models (Alós-Ferrer & Strack, 2014; Evans & Stanovich, 2013), the DPDM assumes that decisions arise from an interplay of a deliberative process (referred to as the utility process), which describes computational-normative decision strategies, and a heuristic process, which captures intuitive-affective aspects of decision making (Alós-Ferrer, 2018). Note that we diverge from the terminology proposed by Alós-Ferrer (2018), who referred to the non-heuristic decision process as the utility process. We refer to this process as the deliberative process in order to provide comparability with other versions of the dual-process model, in which the non-heuristic process is more frequently described in terms of deliberation and computational demand.

The DPDM models these processes as mathematical diffusion processes of evidence accumulation (see Ratcliff, 1978). It delivers qualitative predictions about conditional average RTs in binary and multi-alternative choices. Regarding RTs of correct responses and errors, the DPDM makes the following predictions for situations in which the processes are conflicting or in alignment, i.e. when they yield different responses or the same response (Achtziger & Alós-Ferrer, 2014; Alós-Ferrer, 2018):

> H1: In case of alignment, the response time of correct responses is smaller than the response time of errors.
>
> H2: In case of conflict, the response time of correct responses is larger than the response time of errors.
>
> H3: The response time of correct responses in case of alignment is smaller than the response time of correct responses in case of conflict.

At this point, only a few studies have examined these predictions. First, Achtziger and Alós-Ferrer (2014) reported RT differences consistent with the model's hypotheses in a Bayesian updating task (Charness & Levin, 2005). Decision makers extracted a ball from one of two urns and were rewarded when it was of a specified colour. Choices were based on two decision strategies that were either in conflict or in alignment. The rational strategy maximised payoff by integrating new information and prior beliefs following Bayes' rule (deliberative process) and the associative strategy relied on reinforcement learning through past performance (win-stay/lose-shift heuristic). Responses were slower when decision strategies conflicted. Consistent with the DPDM predictions, errors were slower than correct responses in alignment but faster than correct responses in conflict.

Second, Alós-Ferrer (2018) examined data from an experiment where the recognition heuristic supported choosing a product from a recognisable brand instead of following consumer evaluations. Participants made choices between products based on two attributes that could be conflicting or in alignment: the brand (famous vs. unknown) and customer ratings (the famous brand could have more, less, or the same number of stars reflecting consumer evaluations; see Thoma & Williams, 2013). Errors (following the brand) were slower than other responses when suggestions of both sources of product information were in alignment, but faster when they conflicted.

Third, Spiliopoulos (2018) analysed RT data from a repeated constant-sum game in which participants played against computer algorithm opponents capable of exploiting observed predictability in human behaviour (e.g. following a heuristic). In the two-person constant-sum game, the total sum of payoffs for both players is always the same, but the distribution of payoffs varies across the players (see e.g. Rapoport, 1989). Two decision strategies were either conflicting or aligned. One considered the whole history of the game, relying on working memory capacity (deliberative process), the other strategy was following the win-stay/lose-shift heuristic (heuristic process). Again, errors were slower than correct responses in alignment but faster than correct responses when the decision strategies conflicted.

Fourth, Alós-Ferrer and Ritschel (2019) studied behaviour in a Cournot oligopoly (see e.g. Offerman et al., 2002) where myopic optimisation (i.e. the deliberative process) might conflict or be aligned with interpersonal imitation (i.e. the heuristic process). Participants interacted in this economic game, taking the roles of firms and deciding on the output level of their production. The market price was determined according to the law of demand (i.e. higher quantity of all players' productions led to lower market prices). Hence, to maximise their own payoff, players must determine their individual output levels depending on the output decisions of the competitors in the market (see Alós-Ferrer & Ritschel, 2019, for more detail). The predictions were based on an extension of the DPDM allowing for more than two options. The results confirmed H1-H3.

Taken together, RTs of correct responses and errors in these studies were consistent with the DPDM. Note that Caplin and Martin (2016) developed a dual-process drift diffusion model that similarly incorporated insights from dual-process theories and sequential sampling models. This model differs in positing an endogenous choice of whether to follow the deliberative process or the heuristic process. Since it rested on evidence from a trinary-choice experiment, it was irrelevant for our examination of processes in binary-choice probability judgments.

Present research

We tested three hypotheses of the DPDM (Alós-Ferrer, 2018) on average RTs conditional on alignment or conflict of decision processes. We were interested if the DPDM applies to classic binary-choice problems such as probability judgments. These problems were different from the decisions in the previous studies in many ways. Most importantly, the previous experiments recorded much faster decisions (average mean RT here > 30 s vs. only up to 3 s in the studies by Achtziger & Alós-Ferrer, 2014; Spiliopoulos, 2018, and up to 14 s in Alós-Ferrer & Ritschel, 2019). We concentrated on decision tasks with longer RTs. Diffusion models have recently been applied to slow decisions (Lerche & Voss, 2019), challenging the long-held proposition that their use should be restricted to relatively fast decisions. Our primary aim was to test the DPDM predictions for slow decisions in base-rate neglect, conjunct probability, and ratio bias problems.

Previous studies investigated choice RTs in these paradigms, including situations where conflict and alignment of decision strategies were manifest (Alós-Ferrer et al., 2016; Bonner & Newell, 2010; Newman et al., 2017). However, their purpose was very different from ours. They examined individual difference effects on RTs (e.g. Faith in Intuition; Alós-Ferrer et al., 2016), or tested the speed asymmetry explanation (i.e. the assumption that intuitive responses have generally smaller RTs than deliberative responses; Newman et al., 2017). In particular, Alós-Ferrer et al. (2016) tested within-subject differences in RTs across conflict and non-conflict versions of questions from the Cognitive Reflection Test (Frederick, 2005) and several probability judgment problems. But since every participant answered each question only once, and comparisons were only within subjects for each given, fixed decision, the study could not test for H1 and H2. We extended this research by providing an additional level of analysis, i.e. RTs of correct responses and errors analysed separately for conflict and alignment of decision strategies, yielding an explicit test of the DPDM predictions.

Method

Design and participants

The experiment followed a 2 (condition: alignment vs. conflict) x 3 (problem type: base-rate neglect vs. conjunct probability vs. ratio bias) within-subjects design. We used error rate (ER) and RT in the probability judgment task as dependent variables. Ninety-five participants (47 females; $M_{age} = 23.06$, $SD = 3.90$) were recruited for a compensation of €7.

Materials and procedure

Up to ten participants were invited for laboratory group sessions[1]. The experiment was programmed in OpenSesame (Mathôt et al., 2012). Twenty decision problems were presented sequentially (black font, white background, 13.5 pt size, 1024 × 768 resolution) and remained on-screen until the participant responded. Participants decided by clicking on one of two response buttons labelled A and B. The button order (A left and B right, or reverse) was constant for each participant but randomised between participants.

We selected five base-rate neglect, one conjunct probability, and four ratio bias problems from the literature (De Neys & Glumicic, 2008; Ferreira et al., 2006; Kirkpatrick & Epstein, 1992; Miller et al., 1989; Tversky & Kahneman, 1974). Base-rate problems provoked errors based on the overestimation of stereotypical information and the neglect of base-rate probabilities (Pennycook & Thompson, 2017). For instance, the lawyer-engineer problem (Tversky & Kahneman, 1974) described a person with stereotypical characteristics of an engineer being drawn from a sample dominated by lawyers. While responses driven by the deliberative process were expected to consider the sample base-rate of lawyers (suggesting a person randomly drawn is more likely to be a lawyer), heuristic responses were expected to neglect base-rates and rely on the stereotypical information (suggesting that the person is more likely to be an engineer).

Conjunction problems elicited conflict based on the overestimation of compound probabilities (Fisk, 2017). For example, if a company's marketing department had two options to award a promotion contract (Ferreira et al., 2006): (1) a large agency known to meet deadlines with 60% probability, or (2) two smaller agencies with records of meeting deadlines of 70% and 80%, respectively. The second option had a 56% likelihood of timely

[1]We induced mindsets (Gollwitzer, 2012) in some participants prior to the task in order to explore whether they affected decision processes. Since ERs ($Fs \leq 1.07$, $ps \geq .347$) and RTs ($Fs < 1$, $ps \geq .691$) remained unaffected by the mindsets, we dropped this factor and collapsed the RT data across conditions.

delivery (conjunct probability of .80 and .70), so the first option was normatively preferable. While responses driven by the deliberative process were expected to consider conjunct probabilities, responses driven by the heuristic process were expected to show overestimation of conjunct probabilities.

Ratio bias problems exploited the preference for larger absolute numbers (Bonner & Newell, 2010). For instance, when choosing between two piles of envelopes, of which some may contain an attractive prize, the deliberative process was expected to yield the normatively correct response (the second pile with two out of ten winning envelopes), while the heuristic process was expected to bias responses toward larger absolute numbers (the first pile with 19 out of 100 winning envelopes; Ferreira et al., 2006).

For the present research, we designed two base-rate neglect, six conjunct probability, and two ratio bias problems in order to diversify the pool of decision problems (see supplementary material). These newly developed decision tasks also featured less extreme base-rates than the original ones (for instance, base-rates of 99/1000 compared to 5/1000 in De Neys & Glumicic, 2008).

In all our conflict versions of the choice problems, the deliberative and the heuristic process suggested opposing decisions. In the alignment versions, the deliberative and the heuristic process yielded the same decision. Four semi-random order lists were used to counteract sequence effects. Each participant worked on all 20 decision problems, but only on one of its two versions. The first decision problem was a practice trial and was excluded from data analyses. After completing the task (i.e. after making 20 choices), participants provided demographic information, were debriefed, thanked, and paid.

Results

We studied predictions H1-H3 of the DPDM by within-subject tests. For this purpose, rather than fixing a decision problem (which would make it impossible to compare correct responses and errors within participants), we considered all decisions of one fixed individual as a set and compared errors and correct responses within this set and also controlled for problem types. We excluded trials with responses $< 10\,s$ or $> 90\,s$ because such extreme RTs were implausible (4.03% of all trials)[2]. Given the average number of words per decision task ($M = 84$ words), we assumed that RTs $< 10\,s$ signalled that participants did not read the problem properly. It would imply an implausible reading rate of more than eight words per second, not considering the required decision time. We assumed that response

[2]All analyses were also run with the full sample, i.e. without any exclusions. The results were very similar to those described in the results section.

Table 1. Mean RTs per word (milliseconds per word) for correct responses and errors in alignment and conflict trials for three problem types (SDs in parentheses).

	alignment		conflict	
	correct	error	correct	error
base-rate neglect	412.01	479.10	427.82	436.38
	(139.83)	(166.49)	(146.57)	(156.90)
conjunct probability	415.56	447.50	461.85	464.04
	(176.38)	(139.53)	(157.04)	(187.94)
ratio bias	538.75	536.58	567.57	524.43
	(184.73)	(191.66)	(165.38)	(154.90)

times > 90 s indicated that participants were distracted or disengaged from the task. To account for varying numbers of words, we divided RTs by the number of words displayed in each trial. Thereby, we created an index of RT per word (see Table 1 for descriptive statistics). We reduced the skewness of the aggregated RTs by calculating the natural logarithm, see also the online supplementary material for further detail on the properties of the RT distributions of correct responses and errors.

ERs differed between alignment ($M_{\text{Alig}} = .27$, $SD = .21$) and conflict ($M_{\text{Conf}} = .45$, $SD = .21$), $t(93) = 6.50$, $p < .001$, $d = 0.69$. This was expected because both the deliberative and the heuristic process favoured correct responses in alignment. ERs compared well to earlier research (Ferreira et al., 2006), but differed considerably across problem types both in alignment ($M_{\text{bas}} = .15$, $M_{\text{con}} = .24$, $M_{\text{rat}} = .45$) and in conflict ($M_{\text{bas}} = .58$, $M_{\text{con}} = .47$, $M_{\text{rat}} = .29$). These observations suggested substantial differences in the difficulty of the problem types.

We aggregated RT data across problem types and ran three paired-sample *t*-tests. On a test-by-test basis, we excluded participants who did not produce enough data points (e.g. no errors in alignment; which is common for these situations, see Achtziger & Alós-Ferrer, 2014). Accordingly, we excluded seven, four, and one participant for the tests of H1, H2, and H3, respectively. In support of H1 (in alignment trials, correct response RTs are shorter than error RTs), errors were slower than correct responses in alignment, $t(87) = 3.46$, $p = .001$, $d = 0.38$. In conflict trials, in line with H2, correct response RTs were larger than error RTs. However, this effect was small, and was not statistically significant, $t(90) = 1.77$, $p = .081$, $d = 0.21$. Finally, H3 (correct response RTs are smaller in alignment trials than in conflict trials) was supported, $t(93) = 4.94$, $p < .001$, $d = 0.50$.

To account for differences between the problem types, we used linear mixed-effect models, which were fitted with maximum likelihood estimation using the lmer function in R (Bates et al., 2015). One advantage of mixed-effect models over paired-sample *t*-tests, among others, is that they deal with missing data by maximum likelihood estimation rather than excluding the cases in question. Due to the limited number of decision problems, not all participants provided data points for all cells of the experimental design (e.g. no errors in alignment trials of ratio bias problems).

First, we fitted a mixed-effect model for alignment situations, with problem type (base-rate neglect vs. conjunct probability vs. ratio bias) and performance (correct response vs. error) as fixed effects, random intercepts for participants, and random slopes for the effects of problem type and performance. Consistent with the prediction (H1), errors were slower than correct responses, $b_{\text{error}} = 0.113$, with a .95 confidence interval [0.056, 0.170]. Relative to base-rate neglect problems, responses were slower overall in ratio bias problems, $b_{\text{rat}} = 0.223$ [0.148, 0.297]. The descriptive statistics (Table 1) suggested that the overall RT difference might mainly be driven by base-rate neglect and conjunction problems. However, adding the interaction of problem type and performance did not increase model fit, $\chi^2(2) = 2.50$, $p = .287$, suggesting that RT differences did not vary reliably across the problem types. To provide additional description of the performance, separate mixed-effect models per problem type were estimated. They indicated a RT difference (correct vs. error) that was unequal to zero for base-rate neglect, $b_{\text{error}} = 0.146$ [0.074, 0.217], but no significant effect for conjunction problems, $b_{\text{error}} = 0.094$ [-0.020, 0.208], or for ratio bias problems, $b_{\text{error}} = -0.033$ [-0.127, 0.062].

We repeated this procedure for conflict situations to test H2. There was no difference between RTs of correct responses and errors, $b_{\text{error}} = -0.005$ [-0.061, 0.051]. As the *t*-test had suggested, there was no support for H2. Responses were slower overall in ratio bias, $b_{\text{rat}} = 0.287$ [0.220, 0.354] and conjunction problems, $b_{\text{con}} = 0.078$ [0.013, 0.144], relative to base-rate neglect. There was no interaction of problem type and performance, $\chi^2(2) = 1.93$, $p = .381$. Yet, separate analyses per problem type were run to provide additional description of the effect of performance. There was a non-significant RT difference in the expected direction for ratio bias problems, $b_{\text{error}} = -0.081$ [-0.192, 0.030], while no RT differences were observed for base-rate, $b_{\text{error}} = 0.014$ [-0.061, 0.090], and conjunction problems, $b_{\text{error}} = 0.002$ [-0.100, 0.104].

Lastly, we modelled correct responses across alignment and conflict to test H3. We repeated the procedure used for H1 and H2, but condition (alignment vs. conflict) was entered as a fixed effect instead of performance. Consistent with H3, correct responses were slower in conflict than alignment, $b_{\text{conflict}} = 0.080$ [0.029, 0.130]. Again, responses were slower in ratio bias, $b_{\text{rat}} = 0.302$ [0.239, 0.366], relative to base-rate neglect. The interaction of condition and problem type was not significant, $\chi^2(2) = 3.29$, $p = .194$. Separate analyses were used to describe the effect of condition on RT for each problem type. Correct responses were slower in conflict than alignment in conjunction problems, $b_{\text{conflict}} = 0.116$ [0.037, 0.195], an effect in the predicted direction was observed in ratio bias problems but this was non-significant, $b_{\text{conflict}} = 0.066$ [-0.036, 0.168], and no effect in base-rate neglect problems, $b_{\text{conflict}} = 0.027$ [-0.057, 0.111].

Discussion

We used RT data from probability judgments to test three hypotheses derived from the DPDM (Alós-Ferrer, 2018). Our decision problems required relatively long RTs. The results supported the hypotheses H1 and H3 (but not H2) and therefore strengthened the predictive validity of the DPDM.

The DPDM's (Alós-Ferrer, 2018) suitability to judgments in base-rate neglect, conjunct probability, and ratio bias problems had not been examined before. Our data speak for the model's generalisability to a variety of decision problems beyond its previous applications in Bayesian updating paradigms (Achtziger & Alós-Ferrer, 2014), constant-sum games (Spiliopoulos, 2018), and Cournot oligopolies (Alós-Ferrer & Ritschel, 2019). This is noteworthy because the demand to restrict drift diffusion and related models to relatively fast decisions (e.g. around one second) was recently challenged by the validation of a diffusion model based on slow decisions (Lerche & Voss, 2019).

The cognitive processes in probability judgments may, in principle, be appropriately modelled as dual diffusion processes of evidence accumulation. Yet, the DPDM's generalisability to different types of probability judgments may be limited. Not all DPDM predictions held equally well for all types of decision problems, the predicted effects varied considerably in size. We found robust evidence for H1 and H3, while there was no significant difference between RTs of correct responses and errors in conflict situations (H2).

The results of the separate mixed-model analyses per problem type should not be taken as strong evidence for or against the model's predictive validity for different problem types (as none of the interactions with problem type was significant). Yet, these descriptive analyses suggested possible differences between decision tasks, which could be examined further with more highly powered studies. Future research should specify the relation of the DPDM's deliberative (utility) and heuristic processes, and the cognitive processes involved in solving a variety of decision tasks.

The DPDM in its current form may be open for further development. The model remains silent with respect to parameters like a relative starting point (to model individual differences), or dynamic thresholds (changes over time in the amount of evidence required for one option). But it is one major advantage of the DPDM that it explains fast and slow errors without the requirement of additional parameters (see e.g. Forstmann et al., 2016;

Ratcliff, 1978, relying on additional parameters to explain this pattern), thus avoiding the risk of over-parameterisation. The predictions based on the current parsimonious version were, within the scope of the outlined limitations, supported at large. But an extension of the model considering additional parameters might foster its predictive validity. Such extensions could add important conceptual features to describe the processes in probability judgments and binary-choice more generally (see also Alós-Ferrer, 2018, p. 216).

One could argue that the ER and RT differences observed in our study reflected a choice difficulty manipulation rather than providing evidence for the DPDM. Since alignment and conflict trials represented easier and more difficult decision problems, respectively, one possibility is that ER and RT differences arose from the different levels of choice difficulty rather than from the interplay of a deliberative process and a heuristic process. In this case, one would expect fewer errors and much faster responses in alignment than in conflict trials, for both correct responses and errors. Indeed, in case of alignment (reported above) we observed fewer errors and faster correct responses than in case of conflict (as predicted by H3). However, there was no difference in error RTs between alignment and conflict tasks, $t(84) = 0.97$, $p = .333$, $d = 0.106$. A difference would have been expected if choice difficulty determined RTs in both alignment and conflict trials. Therefore, there is no clear support for choice difficulty as an alternative explanation of our findings.

One limitation of our study was that the design did not allow for disentangling reading time from decision time. There are three approaches to solve this issue. First, non-decisional processes like reading and encoding can be integrated in diffusion models by adding a response time constant parameter (e.g. Lerche & Voss, 2019). Second, reading time can be accounted for mathematically, as we did by dividing RTs by the number of words (e.g. Cuetos & Suárez-Coalla, 2009). Finally, participants could be prompted to make a choice only after they have read the problems. When decisions are relatively slow, however, it remains unclear whether this approach could capture the decision process entirely, or if decisions would be initiated prematurely without waiting for an explicit command. Therefore, we argue that it was a practical solution to account for reading time by setting RTs in ratio to the number of words.

Although our data compared well to previous research (see Ferreira et al., 2006), the error rates in alignment trials were relatively high. One reason could be the difficulty of the decision problems. We developed new problems to ensure that their solution was unknown. Furthermore, the experimental setup did not help improving performance since there was no feedback on decisions. Negative feedback could have raised the awareness that the tasks were not easy, which could have enhanced performance.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The authors gratefully acknowledge funding from the German Research Foundation (DFG) under grant number FOR 1882, AC 183/4-1, awarded to Anja Achtziger, and from the Zeppelin University's Student Research Department. We like to thank Carlos Alós-Ferrer for his comments on an earlier version of the manuscript and two anonymous reviewers for their helpful comments and suggestions.

ORCID

Jonas Ludwig http://orcid.org/0000-0002-0635-3195
Fabian K. Ahrens http://orcid.org/0000-0001-8279-9482
Anja Achtziger http://orcid.org/0000-0001-7207-9342

References

Achtziger, A., & Alós-Ferrer, C. (2014). Fast or rational? A response-times study of Bayesian updating. *Management Science*, *60*(4), 923–938. https://doi.org/10.1287/mnsc.2013.1793

Alós-Ferrer, C. (2018). A dual-process diffusion model. *Journal of Behavioral Decision Making*, *31*(2), 203–218. https://doi.org/10.1002/bdm.1960

Alós-Ferrer, C., & Ritschel, A. (2019). *Multiple behavioral rules in Cournot oligopolies*. Working Paper No. 331, Department of Economics, University of Zurich.

Alós-Ferrer, C., Garagnani, M., & Hügelschäfer, S. (2016). Cognitive reflection, decision biases, and response times. *Frontiers in Psychology*, *7*, 1402. https://doi.org/10.3389/fpsyg.2016.01402

Alós-Ferrer, C., & Strack, F. (2014). From dual processes to multiple selves: Implications for economic behavior. *Journal of Economic Psychology*, *41*, 1–11. https://doi.org/10.1016/j.joep.2013.12.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using {lme4}. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition*, *38*(2), 186–196. https://doi.org/10.3758/MC.38.2.186

Caplin, A., & Martin, D. (2016). The dual-process drift diffusion model: Evidence from response times. *Economic Inquiry*, *54*(2), 1274–1282. https://doi.org/10.1111/ecin.12294

Charness, G., & Levin, D. (2005). When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect. *American Economic Review*, *95*(4), 1300–1309. https://doi.org/10.1257/0002828054825583

Cuetos, F., & Suárez-Coalla, P. (2009). From grapheme to word in reading acquisition in Spanish. *Applied Psycholinguistics*, *30*(4), 583–601. https://doi.org/10.1017/S0142716409990038

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*(3), 1248–1299. https://doi.org/10.1016/j.cognition.2007.06.002

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology*, *91*(5), 797–813. https://doi.org/10.1037/0022-3514.91.5.797

Fisk, J. E. (2017). Conjunction fallacy. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (pp. 25–43). Psychology Press.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, *67*(1), 641–666. https://doi.org/10.1146/annurev-psych-122414-033645

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Gollwitzer, P. M. (2012). Mindset theory of action phases. In P. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 1, pp. 526–545). Sage Publications.

Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: Further evidence for two conceptual systems. *Journal of Personality and Social Psychology*, *63*(4), 534–544. https://doi.org/10.1037/0022-3514.63.4.534

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, *118*(1), 97–109. https://doi.org/10.1037/a0020762

Lerche, V., & Voss, A. (2019). Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research*, *83*(6), 1194–1209. https://doi.org/10.1007/s00426-017-0945-8

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Miller, D. T., Turnbull, W., & McFarland, C. (1989). When a coincidence is suspicious: The role of mental simulation. *Journal of Personality and Social Psychology*, *57*(4), 581–589. https://doi.org/10.1037/0022-3514.57.4.581

Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(7), 1154–1170. https://doi.org/10.1037/xlm0000372

Offerman, T., Potters, J., & Sonnemans, J. (2002). Imitation and belief learning in an oligopoly experiment. *Review of Economic Studies*, *69*(4), 973–997. https://doi.org/10.1111/1467-937X.00233

Pennycook, G., & Thompson, V. A. (2017). Base-rate neglect. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (pp. 44–61). Psychology Press.

Rapoport, A. (1989). *Decision theory and decision behaviour: Normative and descriptive approaches*. Springer.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

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. https://doi.org/10.1016/j.cognition.2017.11.006

Thoma, V., & Williams, A. (2013). The devil you know: The effect of brand recognition and product ratings on consumer choice. *Judgment and Decision Making*, *8*(1), 34–44.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science (New York, N.Y.).)*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124