Judgment and Decision Making (2024), **Vol. 19**:e15 1–22
doi:10.1017/jdm.2024.15

EMPIRICAL ARTICLE

Measuring vacillations in reasoning

Revati Vijay Shivnekar and Nisheeth Srivastava

Department of Cognitive Science, Indian Institute of Technology Kanpur, Kanpur, India

Corresponding author: Revati Vijay Shivnekar; Email: revatis@iitk.ac.in

Received: 1 February 2024; **Revised:** 5 April 2024; **Accepted:** 17 April 2024

Keywords: conflict; vacillations; reasoning; moral reasoning; syllogistic reasoning

Abstract
Our experience of reasoning is replete with conflict. People phenomenologically vacillate between options when confronted with challenging decisions. Existing experimental designs typically measure a summary of the experience of the conflict experienced throughout the choice process for any individual choice or even between multiple observers for a choice. We propose a new method for measuring vacillations in reasoning during the time-course of individual choices, utilizing them as a fine-grained indicator of cognitive conflict. Our experimental paradigm allows participants to report the alternative they were considering while deliberating. Through 3 experiments, we demonstrate that our measure correlates with existing summary judgments of conflict and confidence in moral and logical reasoning problems. The pattern of deliberation revealed by these vacillations produces new constraints for theoretical models of moral and syllogistic reasoning.

1 Introduction

Conflict is a constant in our daily decisions, whether the choices are trivial, such as deciding between tea or coffee, or nontrivial, such as determining the best course of treatment for an ailing parent. When we deliberate such choices, we consciously consider multiple lines of arguments and counterfactuals before making a decision. Phenomenologically, these thoughts seem to appear to us sequentially with different arguments leading us to tentatively prefer different options one after another. We fluctuate in preferences mentally, vacillating back and forth between options as different considerations reveal themselves as we reason. In this article, we argue that such shifts or vacillations in our thoughts can be explicitly measured in real-time and thereby used to quantify the degree of conflict experienced during reasoning.

We investigate how observed vacillations during the decision-making process align with broader measures of conflict currently used in moral and logical reasoning research. To accomplish this purpose, we introduce a new experimental method that tracks mental vacillations by allowing reasoners to express their interim preferences during deliberation. During experimental validation of this method, participants were presented with a series of problems with 2 alternatives identified on the left or right arrow keys. They reported which of the two was the leading choice over the time-course of their reasoning process, thus revealing vacillations as switches in key presses. The sequence in which keys were pressed also allowed us to discern how many times people had changed their mind while reasoning. Figure 1 illustrates a representative deliberation phase in a trial. The green and red circles

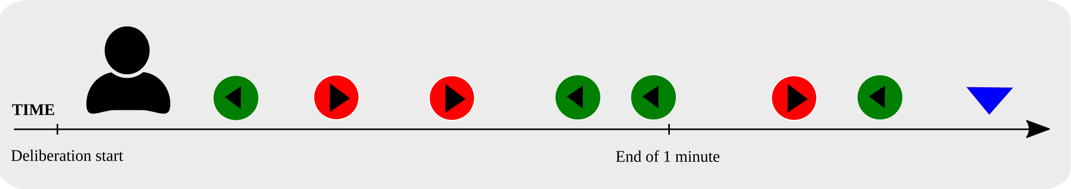


Figure 1. *From Shivnekar and Srivastava (2023). The figure depicts representative key-presses during the deliberation phase of a trial. After reading the problem, participants pressed these keys whenever they wished to record an interim preference. Green and red symbols are the LEFT and RIGHT key presses, respectively. Blue triangle indicates participant ending the deliberation to record the final judgment which they could do so only after 1 min was over.*

represent keys pressed by the participant during deliberation. In this example, the participant reported their preference 7 times and switched 4 times. Participants could conclude their deliberation at any point after 1 min had passed to record their final decision on the next page.

In statistical analyses, we used these switches or vacillations in deliberation as an indicator of conflict experienced by the reasoner. We sought to validate this new measure of cognitive conflict vis-a-vis existing measures in this experimental paradigm.

2 Existing measures of mental conflict

Reasoning research has a long history in psychology and cognitive science alike. Despite this, measuring conflict within the realm of reasoning poses significant challenges, as noted in previous studies such as Tversky and Shafir (1992). Measures of conflict usually provide 1 summary measure per trial. We call such measures trial-level measures as these can differentiate whether one trial showed more conflict than another, but not how conflict unfolds within a trial. Reaction times (De Neys and Glumicic, 2008; Greene, 2008; Greene et al., 2001, 2008; Trippas et al., 2017), normative expectations of behavior within the given choice framework (Evans and Curtis-Holmes, 2005; Greene and Haidt, 2002), and subjective ratings (Frey et al., 2018; Mevel et al., 2015; Pennycook et al., 2015a) serve as examples of trial-level conflict measures. These measures are commonly employed in both moral and logical reasoning experiments and have played a pivotal role in advancing theoretical considerations.

Conflict evolves in tandem with our thoughts that reveal contrasting arguments and choices to us. Given its dynamic nature, a more effective measurement tool for conflict needs to operate in real-time, tracking changes during deliberation while also being minimally intrusive to avoid substantial interference with the process being studied (Schulte-Mecklenbeck et al., 2017). Researchers in moral and logical reasoning fields have used mouse-tracking, eye-tracking, think-aloud paradigms, and so forth, to map the processes underlying the decisions (Bacon et al., 2003; Gürçay and Baron, 2017; Purcell et al., 2022; Skulmowski et al., 2014; Swann Jr et al., 2014). In mouse-tracking studies designed to measure conflict, a typical trial starts with the problem presented centrally on the screen and 2 alternatives displayed in the upper left and right corners. Participants then navigate the cursor with the mouse from the starting position at the center of the bottom edge of the screen to the corner corresponding to their choice. The underlying assumption of mouse-tracking methods is that motor movements in a given period reflect the cognitive processes occurring during that period (Kieslich et al., 2019; Spivey et al., 2005). Consequently, when recording a choice, more curvature in the mouse trajectory between the start point and the choice corner suggests that the non-chosen alternative had a relatively greater influence on the decision-making process than when the trajectory is straighter. Therefore, challenging choices in which both alternatives are attractive should manifest as trajectories with greater curvature. Mouse movements are also utilized to elucidate the temporal dynamics of a choice. Researchers have also examined particular mouse movements, such as the cursor pointer

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

crossing to the side of the non-chosen alternative before switching to the chosen alternative's side, to evaluate whether a specific switching pattern is more predominant than others (Gürçay and Baron, 2017; Koop, 2013).

However, mouse-tracking measures, as described above, lack clarity regarding which part of the process is captured or whether the entirety of it is reflected in the response dynamics. These measures omit a substantial portion of the process under examination by restricting the analysis of the choice or the underlying process to how a response is produced. But when reasoning about a difficult problem, reasoning itself may contain signs of conflict. Eye-tracking methods, on the other hand, are not restricted to just the response dynamics. They operate on the assumption that the decision process is dynamic but gaze reveals what information is being favored currently (Ghaffari and Fiedler, 2018; Pärnamets et al., 2015). Based on this assumption, eye tracking paradigms have used the gaze locations to decipher the moment-to-moment updating of preferences as decisions are being constructed. However, the validity of this assumption across individuals and decision contexts remains unclear. For example, a participant might focus on an alternative not to support it, but to gather evidence against it. In the case of conscious and reportable vacillation, as argued in this paper, gaze pattern shifts between alternatives are only inferential.

Think-aloud protocols offer a closer approximation to tracking the reasons and arguments produced by a reasoner during task deliberation. In these paradigms, participants are prompted to report their thoughts or explain their actions concurrently while completing a task. While thinking aloud during reasoning can provide a finer temporal resolution than mouse-tracking, the method has the potential to interfere with the reasoning itself. Indeed, when participants are asked to explain their actions while performing a task, it can alter their performance on the task (Gagne and Smith Jr, 1962). However, verbalizing thoughts during the task does not seem to improve or worsen performance but it may produce slower responses, presumably due to the additional processing time required for verbalization (Fox et al., 2011). Furthermore, verbal data analyses, such as componential analysis, demand meticulous attention to identifying the purpose and units of analysis, as well as establishing coding systems in advance (Van Someren et al., 1994).

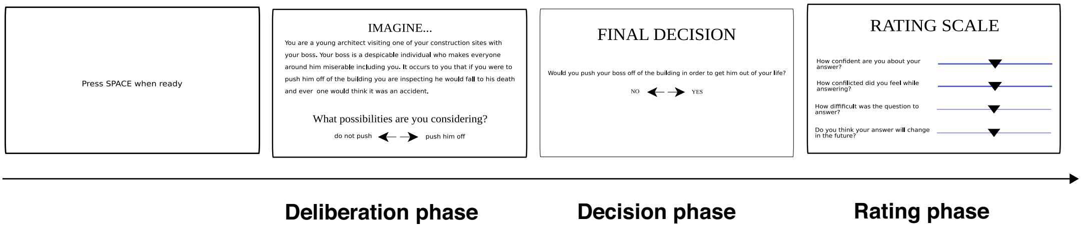
2.1 A key-press paradigm for measuring mental vacillations

Measuring mental vacillations within choice trials is crucial for a realistic assessment of cognitive conflict and the differentiation of plausible theories of the reasoning process. As we note above, existing measures of mental conflict are either insufficiently granular or overly intrusive to adequately measure such vacillations. We introduce a novel method for measuring vacillations. Our method captures participants' instantaneous preferences during reasoning unobtrusively.

In the experiments described below, participants were instructed to report the direction their thoughts were leaning while deliberating on a problem with 2 possible choice alternatives. They were encouraged to express their preferences whenever they felt them building and as frequently as desired. The trial structure was explained to participants using a relatable example of choosing between a favorite and highly rated dish in a restaurant. In a trial, the decision process was divided into reasoning and committing to a final decision. The problem was presented at the center of the screen, with each of the 2 alternatives identified by either the right or left arrow keys. While reasoning, participants pressed the arrow keys whenever they felt they were strongly considering the corresponding choice. This allowed for multiple presses of the same key and the freedom to press them in any order. Participants were explicitly informed that the key presses they made during reasoning did not necessarily have to align with their final decision. This approach was implemented to mitigate the potential bias associated with feeling compelled to reason in line with the normative choice.

The primary dependent variable in our paradigm was the switches in preference. When a participant pressed the right key after pressing the left key, or vice versa, we inferred that during that period, the participant changed their mind. We hypothesized that participants would exhibit more switches

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press



You are a young architect visiting one of your construction sites with your boss. Your boss is a despicable individual who makes everyone around him miserable including you. It occurs to you that if you were to push him off of the building you are inspecting he would fall to his death and ever one would think it was an accident.

Would you push your boss off of the building in order to get him out of your life?

How confident are you about your answer?

How confilicted did you feel while answering?

How diffiicult was the question to answer?

Do you think your answer will change in the future?

Figure 2. *Trial structure in all experiments involves a new problem displayed centrally on the screen during the deliberation phase. Here, a moral dilemma is displayed. In Experiment 3, participants saw the syllogism's 2 premises and the conclusion on separate lines. Participants record their final decisions in the decision phase on a separate screen. Every trial concludes with rating the reasoning experience on subjective measures.*

while deliberating over conflicting problems compared to those with a straightforward choice (see Section 3.1.1 and Figure 2 for detailed description of the paradigm).

Our objective in employing this paradigm was to establish both the internal and external validity of our vacillation measurement as a gauge of cognitive conflict. To establish internal validity, we aimed to demonstrate that people vacillate more when they subjectively feel conflicted during a choice. For external validity, we wanted to see how vacillations map onto previously proposed measurements of conflict in the literature. Experiments 1 and 2 tested 2 operationalizations of conflict in moral reasoning proposed by Koenigs et al. (2007) and Bago and De Neys (2019), respectively. Next, we wanted to test the generalizability of the paradigm when the rules of reasoning are familiar. We used categorical syllogisms in Experiment 3 in place of moral dilemmas, while keeping other details unchanged. Our results demonstrate that directly measuring vacillations in reasoning can help differentiate theories of both moral and logical decision-making.

3 Moral reasoning

Moral dilemmas have been widely employed to study moral reasoning. These dilemmas often pit deontological and utilitarian principles against each other such that choosing one option rules out the reasoner endorsing principle behind the option not chosen. We tested whether 2 specific operationalizations of moral conflict from literature align with our indicator of conflict, that is, vacillations. We expected to see more vacillations in trials which are proposed to be conflicting. Experiment 1 explored whether conflicting dilemmas, identified as those that result in disagreement or dissimilarity in final judgments at the group level, also result in vacillations within the individual. However, such a conceptualization may not reflect the internal conflict within the process of making a choice for an individual.[1] Experiment 2 investigated whether individuals vacillate more when the 2 ethical principles cue separate choices rather than the same one. Bago and De Neys (2019) propose that when deontological and utilitarian principles prompt different choices, a conflict arises in resolving a moral dilemma as opposed to when these principles cue the same choice leading to minimized conflict in decision making.

We were also interested in investigating the order in which deontological and utilitarian responses (henceforth D and U, respectively) are preferred in moral dilemmas. Dual-process theory models (DPT) often hypothesize a specific order in which people consider ethical principles, particularly because they are supported by a fast, emotional system and a slow, deliberative system (see, for a review, Evans and

[1] Consider, for example, a scenario where a group is tasked with choosing between tea and coffee. The group may be evenly divided, implying that the decision on beverage choice is potentially high-conflict. However, each individual member within the group may have encountered no internal conflict in making their personal beverage selection.

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

Stanovich, 2013). In the default-interventionist or corrective model of DPT applied to moral decision-making, wherein the systems engage sequentially, the emotional system is proposed to activate first. This model attributes D responses to the emotional system, while U judgments are attributed to the deliberative system. Hence, the temporal order prediction about the judgments is that reasoners first consider the D choice before the deliberative System 2 overriding it and emitting U as the final choice. However, such a pattern in responding has reportedly been observed only in some select dilemmas (Cushman et al., 2006; Greene et al., 2009; Moore et al., 2008). Thus, we hypothesize that people, if they are to be consistent with the theoretical expectations of the corrective model of DPT in our experimental paradigm, would primarily switch from D to U options, but negligibly from U to D options during choices. The hybrid model of DPT of moral decision-making entails that the emotional system simultaneously generates both D and U responses, albeit with different activation strengths. In other words, reasoners can have D and U inclinations from the beginning, and which gets selected as the initial response will depend on which one of the two has the strongest activation. Whether this response gets updated further depends on the relative difference in the strengths of these activations. Prior research has garnered more support for the hybrid over the corrective model (Bago and De Neys, 2019; Baron and Gürçay, 2017; Gürçay and Baron, 2017; Koop, 2013). However, even in a hybrid DPT account, it is unclear what mechanism might account for people vacillating back and forth between D and U options.

3.1 Experiment 1

In this experiment, we assess our key-press paradigm, which gauges vacillations, alongside evaluating its concordance with conflict defined as cohort-level disagreement and the subjective sense of conflict. The stimuli for this experiment are taken from Koenigs et al. (2007).

3.1.1 Method

Participants

Twenty-five participants were recruited for this experiment (13 females; mean age = 25.3 years). The sample size was derived from a pilot study, where the effect size of the difference between high-conflict personal and low-conflict personal dilemmas was 0.67 (Cohen's d), achieving a power of 0.8 with a significance level of $\alpha = 0.05$.

Materials

We selected 16 problems from Koenigs et al.'s (2007) paper which were divided in 4 conditions: non-moral (NM), impersonal (IM), low-conflict (LC), and high-conflict (HC). All moral problems in their stimuli set had a mean emotionality rating. We ranked these moral problems and selected 4 moral problems for each condition, taking into consideration anticipated familiarity of participants. NM problems were selected only considering familiarity.

In each of the presented problems, participants were tasked with making a 2-alternative forced choice between performing an action and refraining from it. In the NM trials, the scenarios did not invoke any moral principles. The actions in NM condition involved routine tasks like scheduling appointments, choosing between routes (2 problems featured this action), and deciding to purchase product A instead of B.

On moral trials, the stimuli contained the context of the problem in which the action and inaction were made clear along with their consequences. Specifically, actions in the LC and HC conditions involved saving a larger group at the expense of injuring or killing a smaller number of people. These actions were considered personal, as they directly caused harm to individuals or groups such as breaking someone's arm, smothering a baby and so forth (for a more detailed discussion of 'personal' actions in this context, refer to Greene et al., 2001, 2004). Six out of 8 of these trials involved scenarios where death was a possible outcome.

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

Koenigs et al. (2007) classified a personal problem as ‘LC’ post hoc after almost all participants in their study disagreed with endorsing the U action. In ‘HC’ cases, varying degrees of disagreement were observed, with no dilemma recording complete agreement among the participants to endorse the action. IM dilemmas did not include any problems in which the victim died directly from carrying out the action. These scenarios offered a choice between inaction and actions that aimed at benefiting the actor’s welfare, for example, stealing cash from a wallet on the ground, bribing to win a case, and so forth.

At the beginning of the experiment, participants practiced with 2 problems from the same set of stimuli (1 NM and 1 LC). All 18 dilemmas can be found in https://osf.io/3tbgw/?view_only=2f64b229232a4f6899ab59254cb6a90f or in the Supplementary Material.

Procedure

All trials were self-paced. A trial started with a deliberation phase, then the decision phase and ending with a rating screen (refer to the trial structure in Figure 2). During the deliberation phase, participants read the problem which was identical to the body of the problem described in Koenigs et al. (2007) except for the question at the end. The text made clear what choices participants had on that trial. At the bottom of the deliberation phase screen, choices were presented under the prompt ‘What possibilities are you considering?’ Each choice was associated with either the left or right arrow key. The left key represented the deontological option, while the right key represented the utilitarian option on moral trials with these choice alternatives (henceforth referred to as D and U, respectively; see Supplementary Material for list of these problems). Throughout the deliberation phase, participants were instructed to pay attention to their thoughts and indicate which choice they preferred at the moment by pressing the corresponding arrow key. This allowed participants to actively and continuously express their evolving preferences as they deliberated on the given problem. They could report their preferences multiple times (but at least once) and at any point during this phase in a trial. Trials where no key was pressed during this time were excluded. The deliberation phase lasted a minimum of 1 min, although participants could take longer if needed.

After the deliberation phase, participants proceeded to the next screen to make their final decision using the arrow keys, which corresponded to the same options as before. Finally, participants rated their experience of reasoning on the following four 5-point scales: (a) How conflicted did you feel while answering? (b) How confident do you feel about your answer? (c) How difficult was the question to answer? (d) Do you think you will change your mind about your answer?

Upon filling in the consent form, participants were provided with task instructions, which were supplemented with examples for clarity. The initial set of instructions that participants read is as follows:

> Welcome to the experiment!
>
> This experiment will take 30 minutes to complete. The aim of this experiment is to understand how individuals arrive at their choices. Let’s take an example to understand this:
>
> Imagine being in an unfamiliar restaurant, faced with a tempting menu offering options like farm-fresh pasta, pizza, and garlic bread with spread. As you contemplate your choice, various arguments may cross your mind, such as the comfort of pizza or the lighter option of bread and spread when not very hungry. Your task is to pay close attention to these arguments, categorize your thoughts based on the preferred option, and indicate your choice at the end of each scenario.
>
> In this experiment, you will read a few stories and will be asked to think and make a choice at the end. When you are deciding you have to categorize your thoughts based on which choice they indicate. Hence, the task for you is to be attentive to your thoughts and indicate which option you are preferring while reasoning. Whenever you find your thoughts leaning towards one of the options, express your preference by pressing the corresponding key on the keyboard. Feel free to press these keys multiple times and in any order.
>
> Now, let’s walk you through an example trial within the experiment.

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

After these preliminary instructions, we showed participants screenshots of a dummy trial with an NM problem to help navigate the task. Each phase of the experiment carried specific instructions. The experimenter read the following instructions out loud while displaying the screenshots to the participant.

> All trials are self-paced. A trial is made of 4 screens. Press SPACE to continue.
>
> This indicates start of a new trial. You can rest on this screen between trials. [First screen from the left in Figure 2]
>
> This screen contains the context of the scenario. At the end of the scenario, you will be asked to report the possibilities you are considering. LEFT and RIGHT ARROW keys will indicate different kinds of considerations. When you catch yourself thinking about one of them, press the respective key. After about 1 minute, you can press SPACE to go to next page. You can take longer if you have not decided by then. [Second screen from the left in Figure 2]
>
> This screen indicates you have to report your final decision by single key press of LEFT or RIGHT ARROW key. [Third screen from the left in Figure 2]
>
> Finally, you must indicate how it felt to answer the question. There will be four scales: (a) How conflicted did you feel while answering? (b) How confident do you feel about your answer? (c) How difficult was the question to answer? and (d) Do you think you will change your mind about your answer?

Following this, participants completed 2 practice trials with NM and LC problems. Any difficulties encountered during these trials could be addressed by asking the experimenter for clarifications. Experimenter exited the room after commencing the experiment.

3.1.2 Results and discussion

The categorization of stimuli into LC and HC was done in a post hoc fashion by Koenigs et al. (2007). Nonetheless, such division of dilemmas into LC and HC conditions based on cohesiveness of cohort-level judgments was replicated in our experiment. LC dilemmas demonstrated fewer commitments to the U action, with no participant agreeing to take the action in 2 out of 4 of them. There was also a greater variability in endorsing the action in HC dilemmas (Figure 3a). Our primary focus was connecting the cohort-level conceptualization of conflict to an internalized experience of it. Participants' momentary preferences, which are frequently subject to modification during deliberation, were identified with shifts in key presses during the deliberation phase. If on a trial, a participant presses dissimilar keys one after the other then it was counted as a switch. All 4 HC dilemmas showed frequent switching in preferences compared to LC (Figure 3b). This observation aligns with the prediction that for these stimuli, cohort-level disagreements may indicate internal conflict within the individual. To account for effects of both participant and item on switching, we ran a linear mixed effects model treating participants and items as random intercepts, revealing that LC dilemmas, indeed, recorded fewer switches than HC (Table 1a).

Following that, we investigated the pattern in which D and U inclinations are considered. Bago and De Neys (2019) employed the 2-step paradigm to discern the temporal order in inclinations by requiring a quick response at the beginning of the trial, followed by a reasoned response when participants had their final judgment ready. While this method allowed for the dissection of the process to a certain extent, constraining the investigation to specific time windows excludes a significant portion of the reasoning process that follows the initial inclination. To demonstrate this limitation, we created key-press pairs of the first and the last keys pressed on a trial during the deliberation phase, resulting in 4 possible pairs: DD, DU, UD, and UU. Here, the first letter in each couplet denotes the first key, and the second letter signifies the last key pressed during this period.

Notably, all 4 response change types were reported in moral dilemmas (see Table 2 for how frequently these pairs were observed in moral trials). We aimed to determine if there is a predictable order in which these inclinations come to the reasoners' minds. According to the corrective model, reasoners should be inclined toward the D alternative at the beginning of the trial and switch over to U if

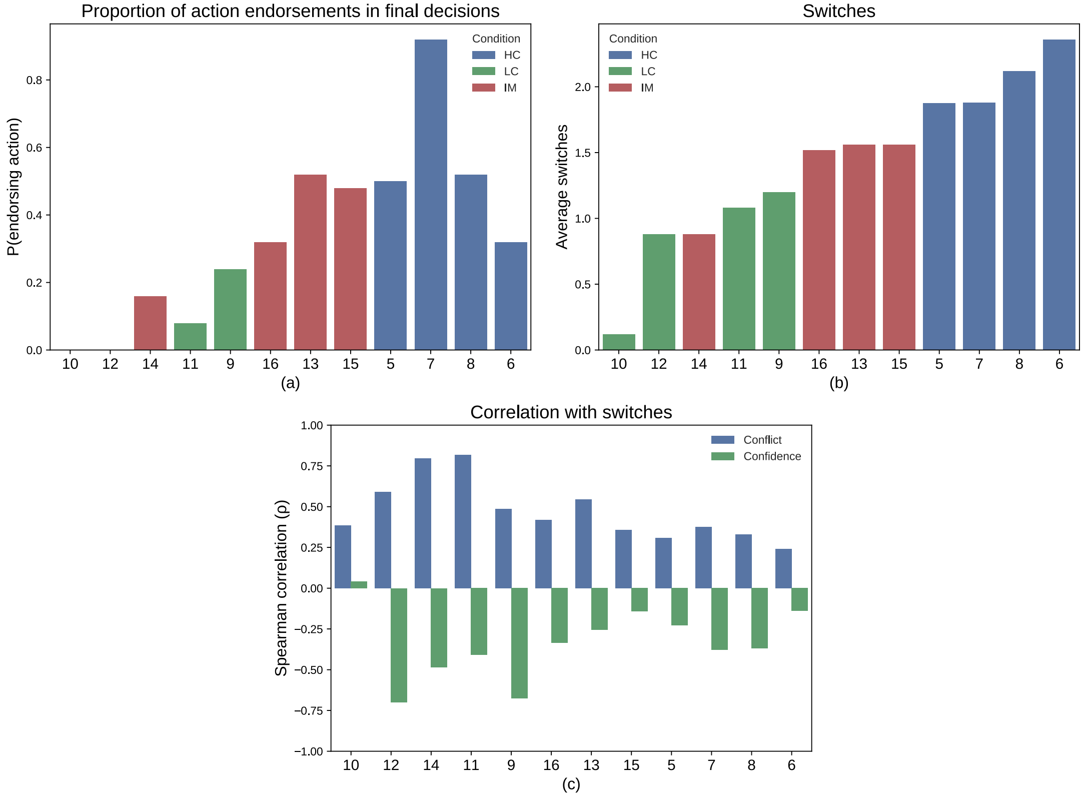


Figure 3. *Results of Experiment 1 across all moral items are depicted in the figure, where each bar represents an item and is color-coded by condition. Panel (a) represents the proportion of trials in which the given action (usually U; see Supplementary Material for more details) was endorsed in the final decision. Panel (b) represents the average number of switches. Panel (c) shows Spearman correlations of switches within a trial to the subjective ratings of Conflict and Confidence reported at the end.*

mental resources permit which should make the DU transitions more prevalent than UD (Greene et al., 2001; Paxton et al., 2012). However, in all 3 moral dilemmas, the frequency of DU was not significantly more than UD (see Table 2 for frequencies of key-press pairs; IM: $\chi^2(1) = 12.78, p < 0.001$; LC: $\chi^2(1) = 3.33, p = 0.07$; HC: $\chi^2(1) = 1.93, p = 0.17$). Furthermore, although DD and UU were the most frequently observed key-press pairs, between the first and the last keys pressed, participants had switched at least twice on 58% and 33% of trials, respectively, to land on the same key they started with. Such oscillations in reasoning are challenging to explain under the sequentiality assumption made by some models of DPT.

Finally, we examined participants' subjective ratings recorded at the end of each trial. While participants rated trials on 4 scales, our primary focus was on conflict and confidence ratings, commonly used indicators of conflict in decision-making (see Frey et al., 2018; Mevel et al., 2015; Pennycook et al., 2015b). Overall, more switches were associated with increased reported conflict and decreased confidence in the final answer. This trend is broadly reflected at the item-level, where conflict is positively correlated and confidence is negatively correlated with vacillations. However, given the item-wise variability in these associations, as evident in Figure 3c, we modeled the subjective ratings by switches accounting for participant- and item-level random effects. Trials with more switches were correlated with increased level of conflict, reporting the problem as difficult and subjective sense that their mind about the answer might change in the future (panels b, d, and e of Table 1, respectively). On the other hand, confidence ratings dropped with more vacillations (panel c of Table 1).

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

Table 1. *Experiment 1 results: Linear mixed effects model of (a) switches, (b) conflict, and (c) confidence ratings by conditions with participants and items as random effects.*

(a) Switches ~condition

Fixed effects					Random effects	
	Estimate	SE	df	*t*		Variance
Intercept	0.8200	0.24	20.03	3.416**	Participant	0.60
HC	1.2357	0.26	9.04	4.76**	Item	0.05
IM	0.5600	0.26	8.98	2.16 .	Residual	2.09

(b) Conflict rating ~switches

Fixed effect					Random effects	
	Estimate	SE	df	*t*		Variance
Intercept	1.1614	0.2	19.95	5.82***	Participant	0.21
Switches	0.3605	0.04	293.27	9.07***	Item	0.3
					Residual	0.98

(c) Confidence rating ~switches

Fixed effect					Random effects	
	Estimate	SE	df	*t*		Variance
Intercept	3.0634	0.16	20.45	18.60***	Participant	0.15
Switches	−0.2246	0.03	293.84	−6.94***	Item	0.2
					Residual	0.656

(d) Difficulty rating ~switches

Fixed effect					Random effects	
	Estimate	SE	df	*t*		Variance
Intercept	1.1920	0.22	17.57	5.46***	Participant	0.21
Switches	0.3009	0.04	291.85	8.42***	Item	0.41
					Residual	0.78

(e) Change of mind rating ~switches

Fixed effect					Random effects	
	Estimate	SE	df	*t*		Variance
Intercept	0.6612	0.14	25.28	4.57***	Participant	0.17
Switches	0.2599	0.03	296.45	7.96***	Item	0.12
					Residual	0.66

Note: Conditions are dummy coded with LC as the reference level. Significant codes: 0, '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1, ' ' 1.

Table 2. *Response changes during deliberation in Experiments 1 and 2.*

Experiment 1					Experiment 2				
	DD	DU	UD	UU		DD	DU	UD	UU
IM	41	4	22	33	Non-conflict	4	8	4	50
LC	80	4	12	4	Conflict	8	5	7	46
HC	32	19	11	37					

In summary, like Koenigs et al. (2007), our results suggest that highly conflicting moral dilemmas observed diverse judgments with most individuals disagreeing with taking action in LC scenarios, while responses were mixed in HC as well as IM dilemmas. Vacillations which are internalized within the reasoner mapped reasonably well with the overall cohort-level disagreements in final decisions as well as subjective feeling of conflict. Furthermore, we employed vacillations as an analytical tool to examine the reasoning process and scrutinize models of moral reasoning by outlining how preferences evolve during deliberation. These patterns revealed that infrequent transitions and frequent oscillations between options were common in reasoning challenging to explain under certain theoretical models.

3.2 Experiment 2

The results of Experiment 1 show that our measurement of mental vacillations, when summed across a trial, correlate well with the expectation of people experiencing conflict during reasoning, as operationalized via cohort-level disagreement by Koenigs et al. (2007). In Experiment 2, we sought to validate our findings from Experiment 1 with a more recent definition of conflict.

Bago and De Neys (2019) manipulated conflict in moral decisions in terms of convergence of deontological and utilitarian principles on a choice. They propose that when these 2 principles contradict each other, people feel conflicted. We tested this definition of conflict in moral decisions in a pre-registered study below.

3.2.1 Method

Participants

Sample size and hypotheses for Experiment 2 were pre-registered (see https://osf.io/ut3yb). We collected data from 27 participants. Four participants' data did not record reliably due to a technical issue in the software and 1 participant failed to qualify our inclusion criterion (i.e., did not record any key press during the deliberation phase in any of the trials). We analyzed data of 22 participants (8 females; Mean age = 20.3 years).

Materials

In Experiment 2, we utilized a moral stimuli set from Bago and De Neys (2019) and NM stimuli from Koenigs et al. (2007). All moral dilemmas were impersonal and required participant to choose between an action and its omission. Consequence of each action within a dilemma was such that it killed a group of people as a side-effect of saving another group (for clarification on the distinction between personal and impersonal dilemmas, as implied here, refer to Greene (2014). Also note that the consequence of impersonal actions here is different from the stimuli used in Experiment 1). Bago and De Neys (2019) manipulated conflict in moral dilemmas based on the convergence of deontological and utilitarian principles. In conflict moral problems, the consequence of the action was such that it saved a larger group of people at the cost of harming or killing a smaller group. These trade-offs mirror familiar trolley problems and their variations such that the choice is between U and D (Greene, 2016; Thomson, 1984). In non-conflict moral dilemmas, the action led to the death of a larger group

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

to save a smaller group. Following is an example of choice alternatives and their trade-offs in a non-conflict dilemma:

'If you activate the emergency circuit to transfer the oxygen, these 11 miners will be killed, but the 3 miners will be saved. Would you activate the emergency circuit to divert the oxygen in the shaft?'

Bago and De Neys (2019) call these dilemmas non-conflict because both deontological and utilitarian principles converge on the choice of not endorsing the action. We refer to this converging choice as U in non-conflict trials for the ease of discussion.

In conflict trials, the U choice refers to the action that saves many by killing a few (see Bago and De Neys (2019) and Conway and Gawronski (2013) for a detailed description of these problems). Following excerpt from a conflict dilemma demonstrates that in these dilemmas the choice is between a utilitarian action and a deontological inaction:

'If you push the button and divert the fire into the sideline, this building will explode and kill the 4 people in it, but the 12 in the building above the main line will be saved. Would you push the button to divert the fire explosion?'

All the stimuli can be found in https://osf.io/64f3z/?view_only=8f87a33c530c4e6b9945acc34b637d3f or in the Supplementary Material.

Procedure

Experiment 2 retained the trial structure from Experiment 1 with a minor adjustment in the rating phase by including only confidence and conflict scales. The instructions remained unchanged from those provided in Experiment 1.

3.2.2 Results and discussion

In Experiment 2, we expected people to be highly cohesive in their final answers at the group level on both conflict and non-conflict items based on response pattern reported in the original paper. Although people produced highly coherent final decisions in both conflict and non-conflict cases, the vacillations reveal that the reasoning process was indeed distinct. Participants switched more frequently on conflict than non-conflict items (Figure 4b). A linear mixed effects model of switches with random effects of the participants and items corroborated this observation (panel a of Table 3). Vacillations also correlated with subjective measures of conflict such that trials on which participants switched more often, they reported greater conflict in reasoning and lesser confidence in the final judgment (panels b and c of Table 3).

However, a more nuanced narrative emerged when examining response transitions. Bago and De Neys (2019) hybrid model of the DPT predicts that individuals who provide a U response both rapidly and after careful consideration (i.e., UU transitions) would not need to alter their preference during deliberation. Contrary to this prediction, our findings suggests that even if they start and end with the same response, participants may not necessarily adhere to that choice throughout the deliberative process. Approximately, 35.42% of all UU trials exhibited at least 2 switches in between. Hence, once again, vacillations in preferences offer a more informative metric than simple response transitions, capturing the dynamic nature of the reasoning process between the initial and final decision points.

4 Logical reasoning

Just like moral dilemmas in moral decision-making research, logical reasoning has been studied widely by using categorical syllogisms. In this kind of syllogisms, participants see 2 premises and a conclusion with categorical propositions. The task is to judge whether the conclusion logically follows assuming the premises are true when, often, the logical validity of the syllogism and the believability of the conclusion are manipulated across trials. Decades-long research shows that people are more likely to accept conclusions that are valid rather than invalid and believable than unbelievable. Particularly, there is also an interaction effect between the validity and believability such that the rates of conclusion

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

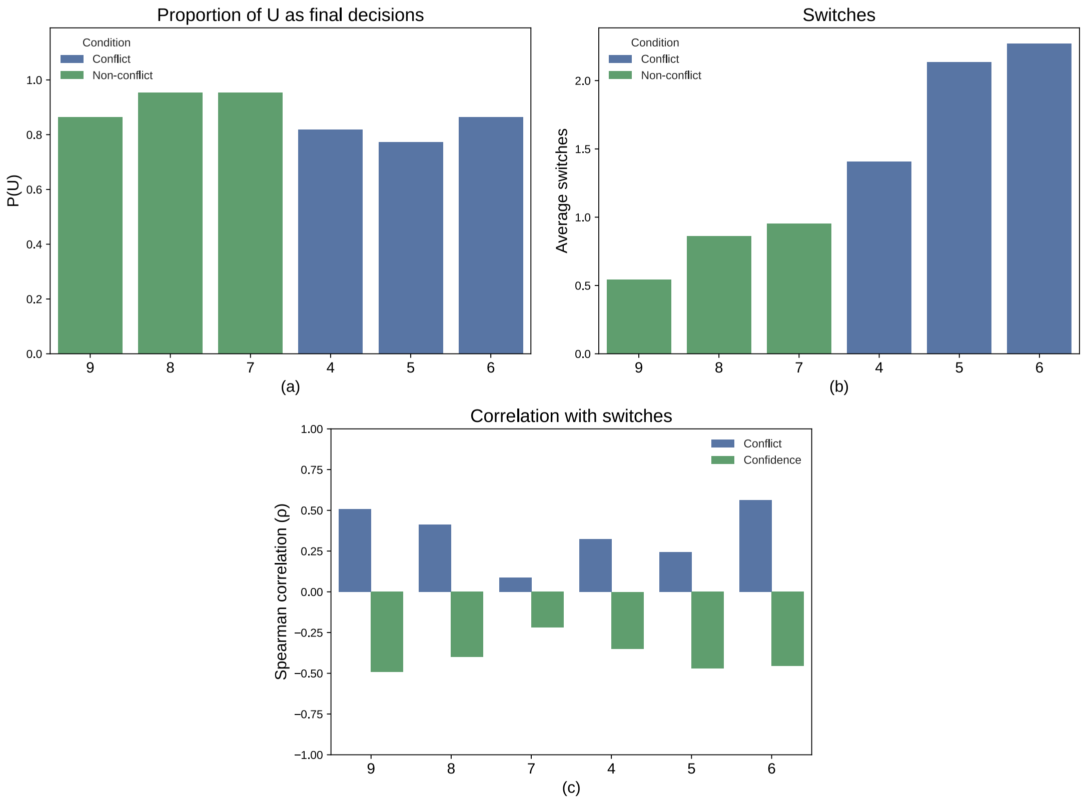


Figure 4. *Results of Experiment 2 across moral trials are depicted in the figure, where each bar represents an item and is color-coded by condition. Panels (a) and (b) represent item-wise proportion of U responses as the final decision and switches, respectively. Panel (c) shows Spearman correlations of switches within an item to the subjective ratings of conflict and confidence reported at the end.*

endorsement differ in valid and invalid trials depending on the believability of the conclusion (Evans et al., 1983; Janis and Frick, 1943; Morgan and Morton, 1944; Oakhill and Johnson-Laird, 1985). This belief bias effect is often attributed to the conflict between the reasoner's belief system and the logical status of the conclusion. When these 2 factors clash, De Neys and Van Gelder (2009) hypothesized that reasoners are less accurate because people have to inhibit the response favored by their beliefs. In no-conflict problems in which both these factors are consistent, no such inhibition has to take place and, hence, people are generally highly accurate. To suppress belief-based responses, reasoners must identify the conflict between belief and logic, a process contingent upon their familiarity with the logical rules and the application of their own priors within the task (refer to De Neys, 2013 for a comprehensive review). We investigate efficacy of vacillation as an indicator of such conflict between belief and logic in syllogisms for our final experiment.

Different theoretical frameworks, such as selective scrutiny, misinterpreted necessity, and the theory of mental models, present varying predictions regarding the sequence of preferences over time (Barston, 1986; Dickstein, 1980; Evans, 1989; Johnson-Laird and Bara, 1984). With the interim preferences at our disposal, an examination of how reasoners are influenced by the believability of conclusions becomes possible. According to the selective scrutiny model, reasoners purportedly begin by assessing the likelihood or believability of the conclusion. If the conclusion is deemed unlikely, only then do they scrutinize its logical status. Conversely, the misinterpreted necessity model proposes that people inspect syllogisms based on logic first, but rely on the believability of the conclusion if it does not necessarily

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

Table 3. *Experiment 2 results: Linear mixed effects models of (a) switches by conditions with participants and items as random effects, (b) conflict, and (c) confidence by switches.*

(a) Switches ~condition						
Fixed effects					Random effects	
	Estimate	SE	df	*t*		Variance
Intercept	0.7879	0.31	13.01	2.53*	Participant	1.18
Conflict	1.1515	0.3	4.0	3.9**	Item	0.02
					Residual	2.44
(b) Conflict rating ~switches						
Fixed effect					Random effects	
	Estimate	SE	df	*t*		Variance
Intercept	2.8049	0.21	22.6	13.09***	Participant	0.77
Switches	0.1730	0.05	90.38	3.50***	Item	0.0001
					Residual	0.83
(c) Confidence rating ~switches						
Fixed effect					Random effects	
	Estimate	SE	df	*t*		Variance
Intercept	3.5449	0.19	22.03	18.94 ***	Participant	0.49
Switches	−0.1507	0.04	118.85	−3.35 **	Item	0.03
					Residual	0.65

Note: Panels (b) and (c) are modeled with only the participant as the random effect to avoid over-fitting. Conditions are dummy coded with non-conflict dilemma-type as the reference level. Significant codes: 0, ‘***’ 0.001, ‘**’ 0.01, ‘*’ 0.05, ‘.’ 0.1, ’ 1.

follow from the premises. In other words, on invalid syllogisms, if the premises lead to the conclusion being valid in some cases and invalid in others, then reasoners would rely on the believability of the conclusion in their judgments.

Along the same lines, the mental model theory argues that individuals initiate reasoning by constructing mental representations by integrating given premises, and subsequently evaluating the conclusion based on its logical validity. If the conclusion lacks logical alignment, it is rejected; only logically fitting conclusions prompt consideration of their believability, with unbelievable conclusions leading the reasoner to form alternate models from the premises. Such a reasoning process is only possible in cases of syllogisms that permit the construction of alternate models, called multiple-model syllogisms. The key-press paradigm serves as a testing ground for evaluating the predictions about the logical reasoning process as it provides a glimpse into the pattern through which inclinations are formed during the process of deliberation. If participants initially express a preference based on the logical validity of the syllogism, the selective scrutiny model can be dismissed. If we observe more switches on valid but unbelievable syllogisms, then the explanation provided by the mental models theory is plausible. However, if participants tend to switch between preferences more frequently on invalid syllogisms regardless of their validity, then the misinterpreted necessity model may be a more suitable explanation. In Experiment 3, we evaluate whether first preference is belief or logic based. Additionally, we explore the frequency of individuals switching between preferences, especially in multiple-model syllogisms with valid but unbelievable conclusions. The results of Experiment 3, indeed, reveal that

Table 4. *Conclusions in 8 syllogisms used in Experiment 3.*

	Conclusions	Believable?
1	All lollipops are made out of sugar.	Yes
2	All animals that are able to swim spend majority of their lives in the water.	Yes
3	Some highly trained dogs are not police dogs.	Yes
4	Some addictive things are not cigarettes.	Yes
5	All college professors have medical degrees.	No
6	All objects with sides of equal area are objects with 6 sides.	No
7	Some priests are not religious people.	No
8	Some deep sea divers are not good swimmers.	No

Note: Conclusions with universal quantifier 'all' (1, 2, 5, and 6) are taken from Robison and Unsworth (2017). Rest are from Evans et al. (1983).

individuals may not strictly adhere to the reasoning patterns predicted by any of the above-mentioned theories.

Switching in reasoning can also be indicative of the quality of thinking. Actively Open-Minded Thinking (AOT) asserts that good thinkers give fair consideration to new information, irrespective of its alignment with existing beliefs (Baron, 2023). This form of thinking is also claimed to be predictive of belief bias (Toplak et al., 2014, 2017). In Experiment 3, we put this prediction to the test by examining whether the frequency of participants' switches was positively correlated with AOT.

4.1 Experiment 3

Experiment 3 uses syllogisms as stimuli. One of the primary motivations for employing a different type of stimulus was to investigate the generalizability of the key-press paradigm to various forms of reasoning. Additionally, theories related to syllogistic reasoning provide hypotheses concerning the underlying cognitive processes involved in problem-solving that can be tested with o. Particularly, we investigate belief bias in 1 and 2-model syllogisms. Through the application of our paradigm to trace this process, we aim to gain insights into syllogistic reasoning with minimal interference, ultimately contributing to a more nuanced understanding of the cognitive mechanisms at play.

4.1.1 Method

Participants

Based on a pilot study, we calculated the sample size of 25 (for a detailed pre-registration report, see https://osf.io/q2n56/?view_only=bd0a1f9894a740d0bcc06679f5be3d6e). Expecting some data loss, we collected data from 30 participants (7 females; mean age = 21.7 years).

Materials

In this experiment, participants solved 8 syllogistic reasoning problems. We employed a within subject 2 × 2 design with 2 factors: validity (whether the conclusion follows logically from the premises) and believability (whether the conclusion is believable). To manipulate the believability of the conclusions, we used the materials from Robison and Unsworth (2017) and Evans et al. (1983) (see Table 4 for all 8 conclusions).

Our stimuli were also divided in one of the 2 forms, single- or multiple-model (Newstead et al., 1992). Single-model syllogisms were taken from Robison and Unsworth (2017) and had the form:

All A are B.
All B are C.
Therefore, all A are C. (valid)
OR
Therefore, all C are A. (invalid)

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

Multiple model syllogisms are those for which more than one conceptually distinct model can be constructed, and to necessarily conclude about the validity of the conclusion, all models need to be considered. These were of the following form, taken from Evans et al. (1983):

Some A are B.
No B are C.
Therefore, some A are not C. (valid)

OR

No A are B.
Some B are C.
Therefore, some A are not C. (invalid)

In Experiment 3, conflict trials (invalid-believable and valid-unbelievable) were those in which there was a conflict between logic and believability, while on non-conflict trials, logic and believability both coincided on the same choice (valid-believable, invalid-unbelievable). List of the stimuli used can be found in the Supplementary Material.

Procedure

The overall trial structure remained the same as in Experiments 1 and 2. In the deliberation phase, participants saw 2 premises and the conclusion in the middle of the screen with each statement on a separate line. The prompt in the deliberation phase was slightly altered to read ‘Which option are you considering?’ for better interpretability. The right and left arrow keys corresponded to the conclusion being TRUE and FALSE, respectively, in both deliberation and decision phases of the trial. On the last screen, participants reported on a 5-point scale how conflicted they felt while solving the syllogism and how confident they were with their final answer.

Since we used syllogisms which has a correct and an incorrect answer, the instructions were altered to include a dummy syllogism to explain the task. The preliminary instructions for Experiment 3 read as follows:

> ‘In this experiment, we are investigating how people solve a particular kind of problem. Let me explain it with an example.
>
> Imagine that you are trying to solve a multiple-choice question with only 2 options: A and B. Sometimes you know the answer right away. However, some other times, you may think A could be right before correcting yourself and saying B is right. Perhaps, you switch again to A and record it as your final answer.
>
> In this experiment, the task is to pay close attention to these thoughts before you settle on your final answer. As you notice yourself leaning toward one of the options, you have to indicate it by pressing a key on the keyboard.
>
> Any questions so far? Please ask now.’

After clarifying any doubts participants had so far, they were explained the task by using a dummy syllogism.

> ‘You will see 3 statements. The first two statements will give you some information about the third one. You have to say if the third statement following “Therefore’. . .’ is a logically valid statement assuming that the first two statements are true.
>
> For example:
> All mammals are zephrodytes.
> All zephrodytes fly.
> Therefore, all mammals fly.
> Is the third statement “TRUE” or “FALSE”?

Table 5. *Generalized linear models for single-model and multiple-model syllogisms, from Experiment 3.*

Single-model					Multiple-model			
Effect	DF	F	p	$\eta^2_{partial}$	DF	F	p	$\eta^2_{partial}$
Validity	(1, 29)	102.9	< 0.001	0.78	(1, 29)	19.12	< 0.001	0.4
Believability	(1, 29)	6.37	0.02	0.18	(1, 29)	13.05	0.001	0
Interaction	(1, 29)	0.14	0.71	0.01	(1, 29)	10.63	0.003	0

> You will solve 8 such problems. For each problem, you will have at least 1 minute to solve, but you can take longer if needed. Remember: while you are thinking about this problem you also have to indicate which answer you are considering.
>
> Any questions so far? Please ask now.’

After reading these instructions, participants saw screenshots of a dummy trial (with the same example used above) along with description of the trial structure. Instructions were followed by the main block with 8 syllogisms. At the end of the experiment, they filled out the actively-open minded thinking questionnaire (AOT; Baron, 2019). We also asked them if they were familiar with syllogisms.

4.1.2 Results and discussion

Twenty-six out of 30 participants were acquainted with syllogisms and had prior experience solving similar problems; however, none of the participants were excluded from the analyses. Despite their familiarity, participants still showed belief bias effect in their judgments. We operationalized participants’ final decisions recorded following the deliberation phase in terms of accepting or rejecting the conclusion. We conducted 2 × 2 repeated measures ANOVAs with believability and validity as predictors for both single- and multiple-model syllogisms. According to the belief bias studies, the difference in acceptance rates between valid and invalid trials is larger when the conclusion is unbelievable than when believable. However, this particular pattern of interaction remained predictive of acceptance rates only in multiple-model syllogisms (Table 5). Reasoners were highly accurate in judging the validity of single-model syllogisms which contained universal premises and conclusions which is in line with representative results from literature in belief bias studies (Newstead et al., 1992).

Although people show belief bias in their judgments produced at the end of a trial, the reasoning process itself is much more textured. Consistent with Experiments 1 and 2, syllogisms on which participants shifted their preferences often were likely to be rated higher on conflict and lower on confidence (Table 6). Multiple-model syllogisms which allow premises to be modeled in more than one way saw more switches in preferences than when they could be arranged only in one way like in single-model syllogisms, consistent with the mental models theory and misinterpreted necessity models (paired t-test: $t(29) = 3.43, p = 0.002, CI = [0.23, 0.90]$). Models of logical thinking discussed before also purport a temporal order in which arguments are considered. When believability and logical validity converge on the same answer (valid-believable and invalid-unbelievable syllogisms), participants were likely to consider the convergent choice first more than chance (proportion = 0.78; one proportion $z(1) = 7.23$, $p < 0.01$). Contrastingly, when these 2 factors contradict, like in invalid-believable and valid-unbelievable syllogisms, participants’ first preferences were more influenced by the logical validity than the conclusion’s believability (proportion = 0.69; one proportion $z(1) = 4.47$, $p < 0.01$). This aligns with the mental models and misinterpreted necessity theories because both theories predict that individuals initially assess validity of conclusions, but contrasts with the selective scrutiny model, which expects that individuals will examine a syllogism’s believability before checking its logical validity.

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

Table 6. *Experiment 3 results: Linear mixed effects models of switches in syllogistic reasoning predicting (a) conflict and (b) confidence ratings with participants as a random effect.*

(a) Conflict rating ~switches						
Fixed effect					Random effects	
	Estimate	SE	df	t		Variance
Intercept	2.1138	0.15	34.18	14.29***	Participant	0.44
Switches	0.5172	0.07	232.74	7.29***	Residual	1.24
(b) Confidence rating ~switches						
Fixed effect					Random effects	
	Estimate	SE	df	t		Variance
Intercept	4.4087	0.09	38.88	50.74***	Participant	0.1
Switches	−0.2954	0.05	237.90	−5.55**	Residual	0.74

Note: Significant codes: 0, '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1, ' ' 1.

According to the mental models theory, people should vacillate more while deliberating on valid-unbelievable syllogisms because people entertain alternate models only in such cases. However, participants switched the most on the invalid-believable syllogism (last bar in Figure 5b). Switches recorded on invalid-believable trials were significantly more than switches on valid-unbelievable ($t(29) = 2.76, p = 0.01, CI = [0.12, 0.78]$) and invalid-unbelievable syllogisms ($t(29) = 2.57$, $p = 0.015, CI = [0.07, 0.63]$). While the misinterpreted necessity model anticipates reasoners to vacillate when making judgments on invalid multiple-model syllogisms, in our experiment, they only do so in invalid-believable syllogisms. Switching between preferences is comparable in valid syllogisms (both valid and invalid) and invalid-unbelievable cases. Therefore, the observed pattern of vacillations was not entirely consistent with either the mental model theory or the misinterpreted necessity model. However, our stimuli were limited to one problem each in each condition within single- and multiple-model syllogisms. It remains to be seen whether these predictions extend beyond the examples we have used in our experiment.

Finally, participants' AOT scores did not demonstrate any correlation with switches ($r_{switches} = 0.25, p = 0.17$). While there is a trend in the anticipated direction for switches, suggesting that a reasoner who considers more diverse information is more likely to score higher on AOT, it remains unclear whether the lack of effect is due to a small sample size. Additional experimentation is required to directly assess this claim.

5 General discussion

Experiencing conflict during reasoning is a common phenomenon, yet its empirical measurement poses challenges. Researchers often rely on trial-level metrics of conflict such as final judgments or choices produced after reasoning and RTs, which limit our insight into how the conflict evolves. Recent decision-making experiments have incorporated process-tracking methods, such as mouse-tracing to get a closer insight into reasoning (see, for a general review, Maldonado et al., 2019; particularly for moral reasoning, see Gürçay and Baron, 2017; Koop, 2013). In these assessments, the trajectories of mouse movements reveal changes in inclinations proportionally toward the available options. However, pinpointing when these adjustments occur during a given trial remains challenging. With real-time process-tracing methods such as think-aloud protocols, there is some level of distortion risk, meaning that the measurement of the phenomenon could itself hinder it. On the other hand, the reliability

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

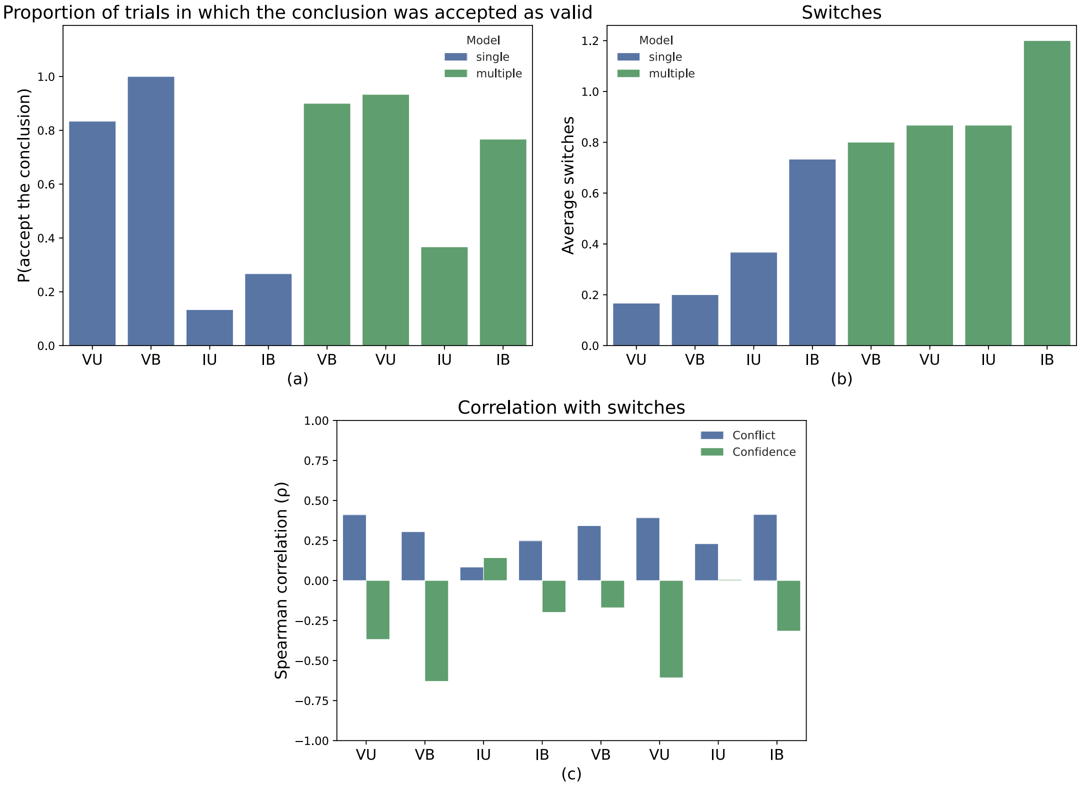


Figure 5. *Results of Experiment 3 across all syllogisms. Each bar is a syllogism which is either valid (V) or invalid (I) and has either a believable (B) or unbelievable conclusion (U). All bars are color-coded by the model type. Panels (a) and (b) depict the proportion of correct trials and number of times participants switched between options while deliberating, respectively. Panel (c) shows average item-wise conflict and confidence ratings recorded at the end of each trial on a scale.*

of eye-tracking methods across different decision contexts and individuals is uncertain due to their indirect inference about the subjectively experienced conflict. Our conflict measurement method, in contrast, allows experimenters to observe the complete time-course of a respondent's decision, thereby facilitating more nuanced analyses. We believe it is also less intrusive than the think-aloud method because reasoners just have to press a key to report their inclination without having to verbalize or give an explanation for their behavior. Additionally, our main dependent variable, switches in preference, can be readily analyzed without the need for inter-rater reliability checks or complex coding techniques. This opens avenues for exploring finer details, including the potential incorporation of gaze or neurophysiological markers to measure the experience of conflict in future research.

In 3 experiments, we demonstrate a reasonable correlation between our measurement and participants' subjective experience of conflict in 2 different fields of reasoning. On problems that people vacillated while reasoning were also those which they rated as conflicting. Their confidence in their final answer was also predicted by how often they vacillated while reasoning.

Measurement of mental vacillations also offer constraints for theories of reasoning and decision-making. Our results from Experiment 3 illustrate that while the final acceptance rates aligned well with predictions from the mental models theory and misinterpreted necessity models, the pattern of vacillations between choices did not support either of them completely. Similarly, while Bago and De Neys's (2019) 2-step method might lend support to a hybrid model of dual-process moral reasoning, our results from Experiment 2 unveil a less straightforward narrative as individuals frequently revisited alternatives considered before.

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

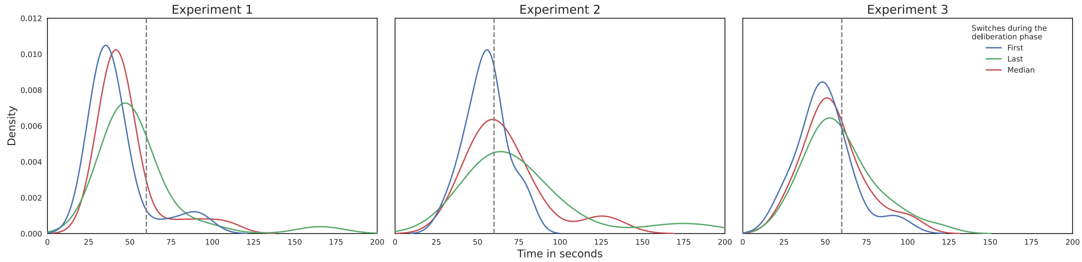


Figure 6. *Density plots of response times of the first, median, and last switches for each participant in 3 experiments. X-axis is time in seconds. The deliberation phase starts with the presentation of the problem in text format to participants at X = 0. The dashed vertical lines are at X = 60 seconds after which participants could report their final decision.*

While not determinative, our findings are inconsistent with dual process accounts that overlay a slow deliberative process on a faster heuristic process, and consistent with simpler race-to-threshold accounts of preference formation, with duality emerging as a property of the set of hypotheses under consideration. As an example of such a single process account, Srivastava and Vul (2015) present an interesting demonstration that the characteristic signatures of both System 1 and System 2 decisions can be elicited from a single race-to-threshold model simply by changing the set of options an observer is selecting between, with fewer choices yielding System 1-like behavior, and more choices yielding System 2-like behavior.

Complementarily, Gürçay and Baron (2017) propose that the choice structure of some problems, particularly moral dilemmas, is such that it evokes feelings of conflict. They conceptualize the reasoning process as a competition among the alternatives to control the final decision such that the reasoner may favor any of these alternatives, in no particular order, before settling on a choice. This idea is consistent with preference switching or vacillations in reasoning. These single process models may be re-examined as potential explanations for results in moral and logical decision-making.

In this paper, we introduce a new method to measure and employ vacillations as an indicator of conflict. Although our paradigm allows a more granular insight into the process of reasoning, there is room for enhancement to extract even more information about the intricacies in this process. To continuously monitor preferences, joysticks can be employed, with the direction of movement indicating the current preference and the distance displaced reflecting the extent of certainty associated with that preference simultaneously. Future studies may also explore modifying instructions to align with specific task demands such as tracking changes in confidence instead of preferences during reasoning.

We had imposed a 1-min interval for deliberation to prevent inattentive responding and ensure active reasoning. However, such enforced temporal constraints on reasoning might introduce confounding factors. The actual time participants take to decide is unclear as a decision may have been reached well before the 1-min mark. A participant may continue to engage in reasoning even after reaching a decision, likely increasing the frequency of switching. A point to note here though is that this additional reasoning, forced or not, may not be ineffectual as it might alter the confidence in the judgment with more information considered.

The imposition of forced deliberation conditions may also impact the timing of key presses, potentially complicating the interpretation of switches as indicators of internalized preference shifts. Alerting participants that the 1-min period has ended could influence their key-pressing behavior. This is more problematic for our measurement if it pushes participants to switch when close to the prompt. We plotted the timing of switches during the deliberation period in Figure 6 to examine when these switches occur. In Experiment 1, participants' initial switches appear well before the 1-min mark (illustrated by gray lines in all 3 panels). However, in Experiments 2 and 3, switches seem to align

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

with this mandated deliberation time. In future research, a more comprehensive understanding of the effectiveness of vacillations could be achieved by eliminating the imposed minimal time for responding. Without the time restriction, it would be possible to investigate the timing of vacillations, how it interacts with contextual factors and whether there are any individual-level differences. Furthermore, drift-diffusion models (DDMs) could provide additional insights into these variables by using this primary data to model preference shifts within decision-making as an evidence accumulation process.

In summary, we present a new experimental paradigm for measuring how individuals vacillate between choices while deliberating. Empirical evidence obtained across 3 experiments indicates that this measurement holds the potential for greater insights than can be gleaned from trial summary statistics such as response times or cohort-disagreement levels. Our results demonstrate that prevailing theoretical accounts of reasoning struggle to adequately explain the sequence of vacillations seen in peoples' judgments. We anticipate that the enhanced visibility into the deliberation process afforded by our paradigm will contribute to refining and improving these theoretical models.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/jdm.2024.15.

Data availability statement. Stimuli, data, and analysis files for Experiment 1, Experiment 2, and Experiment 3 are made available online.

Acknowledgments. We thank Prof. Narayanan Srinivasan for providing insights throughout the project. R.V.S. thanks Akshay Bose for the much-needed coffee breaks.

Funding statement. This research received no specific grant funding from any funding agency, commercial, or not-for-profit sectors.

Competing interests. The authors declare none.

References

Bacon, A., Handley, S., & Newstead, S. (2003). Individual differences in strategies for syllogistic reasoning. *Thinking & Reasoning*, *9*(2), 133–168.

Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, *148*(10), 1782.

Baron, J. (2019). Actively open-minded thinking in politics. *Cognition*, *188*, 8–18.

Baron, J. (2023). *Thinking and deciding*. Cambridge: Cambridge University Press.

Baron, J., & Gürçay, B. (2017). A meta-analysis of response-time tests of the sequential two-systems model of moral judgment. *Memory & Cognition*, *45*(4), 566–575.

Barston, J. L. (1986). *An investigation into belief biases in reasoning*. (Doctoral dissertation, University of Plymouth). University of Plymouth Research Theses. https://pearl.plymouth.ac.uk/bitstream/handle/10026.1/1906/JULIELINDABARSTON.PDF?sequence=1

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*(2), 216–235. https://doi.org/10.1037/a0031021

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17*(12), 1082–1089.

De Neys, W. (2013). Heuristics, biases and the development of conflict detection during reasoning. In H. Markovitz (Eds.), *The developmental psychology of reasoning and decision-making* (pp. 130–147). Psychology Press, New York.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*(3), 1248–1299.

De Neys, W., & Van Gelder, E. (2009). Logic and belief across the lifespan: The rise and fall of belief inhibition during syllogistic reasoning. *Developmental Science*, *12*(1), 123–130.

Dickstein, L. S. (1980). Inference errors in deductive reasoning. *Bulletin of the Psychonomic Society*, *16*(6), 414–416.

Evans, J. S. B., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306.

Evans, J. S. B. (1989). *Bias in human reasoning: Causes and consequences*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Evans, J. S. B., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*(4), 382–389.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*(2), 316.

Frey, D., Johnson, E. D., & De Neys, W. (2018). Individual differences in conflict detection during reasoning. *Quarterly Journal of Experimental Psychology*, *71*(5), 1188–1208.

Gagne, R. M., & Smith Jr, E. C. (1962). A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology*, *63*(1), 12.

Ghaffari, M., & Fiedler, S. (2018). The power of attention: Using eye gaze to predict other-regarding and moral choices. *Psychological Science*, *29*(11), 1878–1889.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, *6*(12), 517–523.

Greene, J. D. (2008). The secret joke of Kant's soul. *Moral Psychology*, *3*, 35–79.

Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (neuro) science matters for ethics. *Ethics*, *124*(4), 695–726.

Greene, J. D. (2016). Solving the trolley problem. In J. Systma & W. Buckwalter (Eds.), *A companion to experimental philosophy* (pp. 173–189). Wiley-Blackwell, West Sussex.

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364–371.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fmri investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108.

Gürçay, B., & Baron, J. (2017). Challenges for the sequential two-system model of moral judgement. *Thinking & Reasoning*, *23*(1), 49–80.

Janis, I. L., & Frick, F. (1943). The relationship between attitudes toward conclusions and errors in judging logical validity of syllogisms. *Journal of Experimental Psychology*, *33*(1), 73.

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, *16*(1), 1–61.

Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis 1. In M. Schulte-Mecklenbeck, A. Kuhberger & J. G. Johnson (Eds.), *A handbook of process tracing methods* (pp. 111–130). Routledge.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*(7138), 908–911.

Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and Decision Making*, *8*(5), 527–539.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19*(6), 549–557.

Morgan, J. J., & Morton, J. T. (1944). The distortion of syllogistic reasoning produced by personal convictions. *Journal of Social Psychology*, *20*(1), 39–59.

Newstead, S. E., Pollard, P., Evans, J. S. B., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition*, *45*(3), 257–284.

Oakhill, J., & Johnson-Laird, P. N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *Quarterly Journal of Experimental Psychology*, *37*(4), 553–569.

Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J., & Richardson, D. C. (2015). Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences*, *112*(13), 4170–4175.

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*(1), 163–177.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015a). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, *24*(6), 425–432.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015b). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72.

Purcell, Z. A., Howarth, S., Wastell, C. A., Roberts, A. J., & Sweller, N. (2022). Eye tracking and the cognitive reflection test: Evidence for intuitive correct responding and uncertain heuristic responding. *Memory & Cognition*, *50*, 348–365.

Robison, M. K., & Unsworth, N. (2017). Individual differences in working memory capacity and resistance to belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, *70*(8), 1471–1484.

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-tracing methods in decision making: On growing up in the 70s. *Current Directions in Psychological Science*, *26*(5), 442–450.

Shivnekar, R. V., & Srivastava, N. (2023). Measuring moral vacillations. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45), 464–470.

Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014). Forced-choice decision-making in modified trolley dilemma situations: A virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience*, *8*, 426.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398.

Srivastava, N., & Vul, E. (2015). Choosing fast and slow: explaining differences between hedonic and utilitarian choices. In *CogSci*.

https://doi.org/10.1017/jdm.2024.15 Published online by Cambridge University Press

Swann Jr, W. B., Gómez, Á., Buhrmester, M. D., López-Rodríguez, L., Jiménez, J., & Vázquez, A. (2014). Contemplating the ultimate sacrifice: Identity fusion channels pro-group affect, cognition, and moral decision making. *Journal of Personality and Social Psychology*, *106*(5), 713.

Thomson, J. J. (1984). The trolley problem. *Yale Law Journal*, *94*, 1395.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Rational thinking and cognitive sophistication: Development, cognitive abilities, and thinking dispositions. *Developmental Psychology*, *50*(4), 1037.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2017). Real-world correlates of performance on heuristics and biases tasks in a community sample. *Journal of Behavioral Decision Making*, *30*(2), 541–554.

Trippas, D., Thompson, V. A., & Handley, S. J. (2017). When fast logic meets slow belief: Evidence for a parallel-processing model of belief bias. *Memory & Cognition*, *45*, 539–552.

Tversky, A., & Shafir, E. (1992). Choice under conflict: The dynamics of deferred decision. *Psychological Science*, *3*(6), 358–361.

Van Someren, M. W., Barnard, Y. F., Sandberg, J. A. (1994). *The think aloud method: A practical approach to modelling cognitive processes*. London: Academic Press, *11*, 29–41.

Cite this article: Shivnekar, R. V. and Srivastava, N. (2024). Measuring vacillations in reasoning. *Judgment and Decision Making*, e15. https://doi.org/10.1017/jdm.2024.15