Conflict in moral and nonmoral decision making: an empirical study coupled with a computational model.

Flora Gautheron, Jean-Charles Quinton, Annique Smeding

Conflict in moral and nonmoral decision making: an empirical study coupled with a computational model

Flora Gautheron[1,2], Jean-Charles Quinton[2], Annique Smeding[1]

[1]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, LIP/PC2S, 38000 Grenoble, France.

[2]Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.

Contributing authors: flora.gautheron@univ-grenoble-alpes.fr; quintonj@univ-grenoble-alpes.fr; annique.smeding@univ-smb.fr;

Abstract

While moral psychology research has extensively studied decision making using moral dilemmas, such high-conflict situations may not fully represent all moral decisions. Moreover, most studies on the effect of conflict have focused on nonmoral decisions, and it is unclear how it applies to the moral realm. The present mixed-method research investigates how conflict impacts moral compared to nonmoral decision making. In a preregistered empirical study (N=42), participants made moral and nonmoral decisions with varying levels of conflict while their mouse trajectories were recorded. Results indicate that moral decisions were more stable in the presence of conflict, while still seeking compromise. In addition, decisions were more affected when conflict got higher. Mouse-tracking data further indicate that some factors are impacting the decision process earlier than others, supporting the relevance of tracing methods to dig into finer-grained decision dynamics. We also present a computational model that aims to capture decision mechanisms and how conflict and morality influence decision making. The model uses dynamic neural fields coupled with sensorimotor control to map a continuous decision space. Two model versions were compared: one with greater perceptual weight for moral information, and another with earlier processing of moral versus nonmoral information. The simulated data more successfully reproduced empirical patterns for the second version, thus providing insights into the underlying decision processes for both moral and nonmoral decisions, in the presence of conflict or not.

Keywords: conflict, moral decision-making, mouse-tracking, dynamic neural field model

1 Introduction

Consider a simple everyday decision: a person deciding to help her friend. This decision would no longer be as straightforward if it involves losing 100€. Indeed, one may anticipate a more conflicted and longer decision: the decision becomes more challenging, as individuals must weigh personal loss against potential harm to a relationship. Moreover, the same conflictual element (i.e., losing 100€) should not have the same impact regarding decision difficulty in a nonmoral situation, such as enjoying your favorite beverage and losing 100€. These examples illustrate how conflict may affect decision-making processes, with effects varying depending on whether the situation involves moral considerations or not.

1.1 Effect of conflict on decisions

Conflict arises when two (or more) irreconcilable elements are pushing in different directions or toward different alternatives (Evans et al, 2015; Cheng and González-Vallejo, 2018; Gürçay and Baron, 2017). The absence of a clear preference results in decisions perceived as more difficult. Conflict is even conceived as a psychological dimension of decision difficulty (Cheng and González-Vallejo, 2018) and those concepts are often associated (Anderson, 2003).

The effect of difficulty in decision making has been widely studied, especially in consumer research (Ratneshwar et al, 2003; Luce et al, 1999; Krosnick, 1991). More difficult decisions are often more extreme on a response scale (e.g., of Likert type), extremization being used as a simplifying strategy (Ford et al, 1989). The concept of more extreme decisions may not appear intuitive to readers familiar with the traditional two-alternative forced choice paradigm. However, employing a continuous response scale may prove more suitable for discerning subtle influences on decision outcomes that might not be discernible in forced binary responses, such as extremization of answers. At the processing level, increasing difficulty in decisions leads individuals to use reduced processing strategies to limit cognitive effort (Paquette and Kida, 1988). Precisely, they tend to use simplifying strategies (e.g., choosing status quo, or relying on first impression) to make the decision task more manageable (Ford et al, 1989). Thus, when facing difficult decisions, decision makers avoid compensatory evaluation and instead select the alternative that is most attractive on whatever dimension is difficult to trade off (Luce et al, 1999). These strategies, leading to extremize answers, can rely on the phenomenon of satisficing: people often engage in sub-optimal decision making strategies to reduce cognitive effort (Barge and Gehlbach, 2012). Despite these simplifying strategies, decision conflict still results in longer response times (Tyebjee, 1979). A limitation of past research when transposed to moral psychology is that difficulty has often been operationalized with cognitive load - which is not equivalent to the intrinsic difficulty of the decision itself. Besides, most research has focused on decisions regarding product purchase, and favored an effort/accuracy trade-off approach (Paquette and Kida, 1988; Leong and Hensher, 2012). These are important limitations as many everyday decisions do not necessarily involve purchase decisions or accuracy, with objective right or wrong answers.

Furthermore, those studies only investigated conflict in nonmoral situations, while we could expect different impact on moral decisions. Indeed, research has evidenced differences in decision making processes between moral and nonmoral situations (Moll et al, 2002;

Van Bavel et al, 2012; Tassy et al, 2013), suggesting that morality leads to different perceptual processes (Gantman and Van Bavel, 2014; Brady et al, 2020; Gantman and Van Bavel, 2015), neural correlates (Sommer et al, 2010), and reasoning (Greene et al, 2001). However, moral decisions have often been studied through moral dilemmas, a particular case inducing a highly conflictual decision (Bauman et al, 2014). This results in a confound between morality and conflict. Independently examining the influence of morality and conflict by comparing moral and nonmoral conflictual decisions is therefore needed. This will be implemented and tested in the present research.

1.2 Conflict in moral situations

In morality, findings suggest that people stand on extreme positions, unwilling to make compromises when decisions rely on moral convictions (Skitka and Morgan, 2014). Similarly, some authors suggest the existence of protected values, which are those moral requirements that one would hardly trade and that lead to strong positions (Bartels, 2008; Bartels et al, 2015; Dehghani et al, 2008), even though they may be malleable depending on the situation (Iliev et al, 2009). Moral dilemmas (involving conflicting moral duties) are thus perceived as difficult and leading to a hard-to-solve conflict between the different options (Hanselmann and Tanner, 2008). Conflict will therefore be higher if the concerned attributes are harder to trade-off because they both belong to the moral category, and lower if they do not belong to the same category (e.g., moral reasons with nonmoral reasons). An example of the latter would be a moral standard weighted against a monetary incentive. Two decision strategies may emerge, depending on moral intensity. If there are great moral implications, as in moral dilemmas, people are more likely to decide according to protected values and moral rules, in a non-compensatory way, resulting in extreme positions (Skitka and Morgan, 2014). Conversely, if moral implications are less intense, as in most everyday moral decisions, moral considerations may be more easily traded off, resulting in less extreme decisions in case of conflict. In this vein, and contrasting with response extremization, research on social dilemmas showed that conflict between options rather leads to less extreme decisions, along with slower response times (Kleiman and Hassin, 2011; Evans et al, 2015).

By default, everyday moral and nonmoral decisions are likely characterized by limited moral intensity and limited conflict. To generate different levels of conflict, varying monetary incentives can be introduced. A higher level of conflict in everyday moral and nonmoral decisions is expected to result in less extreme, more nuanced decisions compared to a lower level of conflict. Besides, as everyday moral decisions should elicit stronger positions and resistance to compromises (Skitka and Morgan, 2014) compared to nonmoral decisions, moral decisions should on average (regardless of level of conflict) be more extreme, less nuanced compared to nonmoral decisions. In the current research, we aim to go beyond final responses and response times, therefore adopting tracing methods to focus on decision processes.

1.3 Tracing conflict in decision making with mouse tracking

Decision making processes can be studied from a dynamical perspective using tracing tools such as mouse tracking (Sullivan et al, 2015; Freeman and Ambady, 2011; Smeding et al, 2016; Koop, 2013). Mouse tracking gives access to the dynamic evolution of the decision, compared to paradigms recording only final decisions and response times. By asking participants to move their mouse cursor from a fixed *START* button to the answer location, mouse tracking gives insights into cognition. Notably, mouse tracking translates the competition or conflict between alternative responses at the cognitive level into movements of the hand which control the mouse on the screen (Freeman et al, 2011; Spivey et al, 2005). This allows tracing the decision making process in real time from early stages onwards, revealing information about all the steps, hesitations, and attractions to one or the other response option that contribute to the final decision and response time (Hehman et al, 2015). Recent studies have shown that process measures could be used to trace decision dynamics with graded responses (using finger tracking in Dotan et al (2019), or mouse tracking in Gautheron et al (2023)), to measure ambivalence (e.g., mouse-tracking in Buttlar and Walther (2018)) and decision reversal (e.g., mouse-tracking in Koop (2013)) in moral decisions, as well as to manipulate moral decision outcomes (using eye-tracking in Pärnamets et al (2015)). Mouse tracking is therefore particularly well-adapted to trace how conflict impacts decision dynamics in moral and nonmoral situations. This applies when empirically tracing human participants' decision dynamics, but also when simulating them. Indeed, from a computational perspective, conflict can be modeled as the competition between two (or more) evidences in different locations of the decision space (i.e., the space where the model is evolving) (see section "Modeling conflict" in the Model part below). Hence it simulates how different elements are voting in different locations, and how they lead to the final decision by attracting preference to their position location during the decision process.

1.4 Contribution of computational models of decision making: A focus on Dynamic Neural Fields

Computational models of decision making provide complementary insights on the dynamics of decision making and can be of substantial interest to understand complex decisions such as moral decisions. Indeed, formal models allow studying latent subcomponents of decision making that can hardly be observed in behavioral studies and explain how their interventions in decision lead to a specific behavior (Crockett, 2016). For instance, the Drift Diffusion Model (DDM) and the larger class of sequential sampling models propose the idea of a decision process corresponding to an accumulation of evidence for one or the other alternative over time (Ratcliff and McKoon, 2008). This model has been successfully used to account for psychological results (Ratcliff and McKoon, 2008; Ditterich, 2006; Metin et al, 2013; Drugowitsch et al, 2012; Krajbich et al, 2015; Ratcliff and Tuerlinckx, 2002; Krypotos et al, 2015). Here we will focus on another class of models similar to sequential sampling models: Dynamic Neural Fields (DNF, Schöner and Spencer, 2016), that rely on a continuous decision space with attractors that will guide the preference. A final, stable, and discrete decision is thus emerging from continuous and non-linear cognitive dynamics, with continuity present in both time and space (Spivey and Dale, 2004). This model has successfully been used for motor behavior (Erlhagen and Schöner, 2002) and visual attention (Rougier and Vitay, 2006; Quinton and Girau, 2011). It is especially interesting as it encompasses DDM-related models through parameter specification and facilitates the addition of a motor module to generate mouse trajectories, similar to those obtained via mouse tracking paradigms with human participants (Lepora and Pezzulo, 2015; Quinton et al, 2013). Subsequently, simulated trajectories can be compared to human participant trajectories, together providing proof of concept for the plausibility of the computational model and a stringent test of decision making processes underlying human decisions (Smeding et al, 2016).

Evidence accumulation models have been developed in the moral and social domains (Falandays et al, 2021; Falandays and Spivey, 2020; Johnson et al, 2017) and to account for conflict in decision making (Krajbich and Rangel, 2011; Krajbich et al, 2014; Sullivan and Huettel, 2018). These models also predict that individuals with conflicting goals are slower to reach a decision and are more likely to select an intermediate, nuanced response (Krajbich and Rangel, 2011; Krajbich et al, 2014; Ratcliff and Smith, 2004; Klauer, 2014). However, these models have not been used to compare moral and nonmoral decision making dynamics at the recent exception of Falandays et al (2021). This work uses DDM to explore the difference between moral dilemmas and nonmoral choices. Their findings suggest that the specificity of moral dilemmas lies in higher conflict compared to nonmoral stimuli, with dilemmas designed to generate balanced responses, probably resulting from conflicting evidence. The encompassing class of sequential sampling models could more generally be used to investigate how decisions may emerge as a trade-off in the presence of conflicting pieces of evidence that compete for decision, manipulating and combining moral or nonmoral elements (such as monetary incentives) within stimuli. Direct competition between alternative responses (and not only through attentional resources) can be modeled with DNF as well as several sequential sampling models that integrate reciprocal inhibition between accumulators, but not with independent DDM.

As some evidence points out that moral and nonmoral elements could be differently integrated or lead to different decision processes (Gantman and Van Bavel, 2014; Brady et al, 2020; Gantman and Van Bavel, 2015; Sommer et al, 2010), we can expect such differences to be reflected in the model parameter values. Additionally, moral content seems to be more salient in perceptual streams, with a higher detection rate when the stimulus is perceptually ambiguous Gantman and Van Bavel (2014). Another effect of moral content would be to elicit fast moral intuitions that could be processed earlier compared to nonmoral content Haidt (2001). These two phenomena—an amplified or earlier influence of morality-can be tested and compared using modeling, relying on sequential sampling models to bring a fine-grained understanding of how morality is processed during decision making, compared to nonmoral elements, in the presence of conflict or not. In other words, we will develop different model versions which, while remaining parsimonious, will integrate a different parameter to reflect these phenomena. Comparison of these different model versions enables to weigh their relative validity, thereby providing proof of concept of how morality intervenes in decision making.

1.5 Overview

Taking into account existing findings on how conflict influences decision making, and the limited evidence specifically focused on everyday moral decisions, the present research aims to investigate whether conflict is managed differently in everyday moral and nonmoral decisions, resulting in distinct strategies and dynamics. By default, everyday moral decisions should be characterized by limited moral intensity and limited conflict for most people. Therefore, to generate conflict, we used a set of everyday actions from Van Bavel et al (2012) and introduced a conflicting element in the form of monetary gains or losses. We operationalized conflict intensity as a small versus high monetary gain or loss, with amounts tailored to the target undergraduate population, resulting in low versus high conflict for each decision. We hypothesize that for all actions, the greatest loss will discourage the action more than the smallest loss (and more so than the two gain conditions), while the greatest gain will encourage the action more than the smallest gain. Given the limited moral intensity of the selected actions (such as telling a white lie or being honest), trade-offs should be acceptable for both moral and nonmoral decisions, albeit less so for the former than for the latter. Indeed, the influence of conflict on moral decisions should be smaller than on nonmoral decisions, with more stable positions for moral actions, as the presence (versus absence) of a moral component should still foster stronger positions and more resistance to compromise (Skitka and Morgan, 2014). Therefore, nonmoral conflictual decisions should elicit less extreme responses compared to moral conflictual decisions (Van Bavel et al, 2012), and be more strongly impacted by a conflictual element. Additionally, the strength of the conflict should also affect response extremity, with less extreme responses for low-conflict as opposed to high-conflict decisions, and a greater increase in this effect for nonmoral decisions.

Conflict is also expected to impact decision dynamics, leading to longer response times. Employing a mouse-tracking technique is particularly relevant to measure the effect of conflict on moral and nonmoral decision dynamics. Specifically, we anticipate that conflictual decisions will result in trajectories that reveal the influence of the two conflicting elements, with deviations from one side before transitioning to the other. Moreover, these dynamics should differ between moral and nonmoral decisions, as some literature (Gantman and Van Bavel, 2014; Haidt, 2001) suggests an earlier and facilitated processing of moral content. Modeling is then particularly relevant to test those assumptions on underlying moral and nonmoral decision mechanisms. Notably, an intriguing query arising from our experimental design—which combines a nonmoral monetary incentive condition with an action that can either be moral or nonmoral—is whether the presence of a moral component compels individuals to interpret all elements from a moral perspective, subsequently processing them all as moral elements. Indeed, those stimuli elements (action and monetary incentive condition) correspond to the different factors, or evidence, taken into consideration in the decision: when at least one of them belongs to the moral domain, the decision is subsequently moral. Alternatively, it is plausible that the nonmoral condition is still processed distinctly—as a nonmoral element-while the action, in contrast, is processed as a moral element (exerting a greater or earlier influence). This conundrum highlights the intricacies of moral cognition in decision making processes. We therefore tested, using a reference DNF model, whether using the same parameters for moral and nonmoral components in conflict situations would be sufficient to fit human decisions. If so, expected differences in parameters should reflect differences in decision mechanisms between moral and nonmoral decisions. On the contrary, if such basic reference model is unable to fit human data correctly, additional parameters can be integrated to improve the fit. To keep the models as parsimonious as possible, we implemented two augmented versions of the reference model, each incorporating a single additional parameter to account for the specific influence of moral elements in decision making, either through an amplified or an earlier impact. These two versions posit that it is solely the moral component, rather than the entirety of the decision, that undergoes specific processing. Comparison of these different model versions enables to weigh their relative validity, thereby providing proof of concept of how morality intervenes in decision making.

2 Empirical study on human participants

2.1 Participants

Forty-two undergraduates (24 females, 18 males) took part individually in the study in exchange for course credits. Mean age was 25.8 years ($SD = 6.6$). Participants provided written informed consent before participation and the project was approved by the local ethics committee (CER-USMB). We estimated sample size based on statistics reported by Trémolière and Bonnefon (2014) and, as we did not find similar studies investigating interaction effects of interest, we estimated interaction effect sizes from data collected during a pretest phase of the experiment. We used PANGEA tool (Westfall, 2015), and performed a sequential analysis based on Lakens' recommendations (Lakens, 2014). We performed our first analysis at 42 participants, and obtained a significant effect for conflict and morality (p-value below the interim threshold at 0.017). We therefore stopped data collection. The study and data collection were conducted in accordance with the ethical principles of the American Psychological Association and the Declaration of Helsinki. All measures, manipulations, and exclusions are disclosed, and all participants were retained in analyzes. All data were collected prior to analysis. Preregistration is available on the Open Science Framework at the following link https://osf.io/6juex/?view_only=96c88938c98240b7bda1cfc9c0eaeed0.

2.2 Materials

One hundred and nine actions were translated from Van Bavel et al (2012). They consisted of short items with similar length for moral and nonmoral items (e.g., "write an email" or "vandalize a park"). Items were categorized as moral or nonmoral based on two criteria. First, depending on whether they include moral lexicon, for example "vandalize" or "cheat". Indeed, moral terms have specifically been used to identify moral content in natural language (Hopp et al, 2021), and more generally to compare moral and nonmoral decisions (Tassy et al, 2013; Moll et al, 2002; Van Bavel et al, 2012). They may also be used to ensure that moral considerations are intervening in decisions, as seen in studies with moral dilemmas (Greene et al, 2001; Broeders et al, 2011; Liu and Ditto, 2013; Crockett, 2013; Kahane, 2015). Second, a characteristic of moral actions is the common belief that they would lead to a beneficial outcome if universally adopted. This phenomenon, universalization, is another indicator of moral content (Levine et al, 2020). Accordingly, we used the distribution of answers (from "nobody" to "everybody") to the universalizing question "how many other people should do the action?" on the same set of items obtained from the study of Van Bavel et al (2012) to refine the moral/nonmoral categorization.

These actions were subsequently associated with a condition of gain or loss with two different amounts of money to elicit low versus high conflict: a small (1€) and a high amount (100€). See Fig. 1 for an illustration. However, as actions could endorse either a positive or negative valence, conflict would occur according both to the action valence and the money gain or loss condition. Indeed, conflict is induced when the positive valence of the action is associated with the condition of losing money (-100€, -1€), and reciprocally when the negative valence is associated with the condition of gaining money (+1€, +100€).

2.3 Procedure

The design was a 2 (action: moral, non-moral) × 2 (monetary condition: gain, loss) × 2 (amount of money: 1€, 100€) partially crossed within-participant design. The item set was indeed divided in quarters and each participant saw only three quarters to have a reasonable experiment time, resulting in four different between-participant item subsamples (which were not expected to have any significant influence on results). Each action was displayed one at a time on a screen when the participant clicked on the "START" box and remained so until the participant responded. Responses consisted in deciding whether the participant would do the action on a scale going from “Not at all” on the left, to “Absolutely” on the right. For each response, participants used a slider of varying width that became visible inside a horizontal bar once the trial was initiated, then following the mouse cursor along the x-axis of the screen (see Fig. 1). The slider had a linear color gradient to blend with the white background, and the most contrasting color corresponded to where the cursor pointed. The width of the slider varied with time: it decreased when the participant did not move the mouse and increased when they did, with a trial considered failed if the slider fully disappeared (reaching zero width). This feature encouraged quick answers and continuous moves (a prerequisite to analyze mouse trajectories (Hehman et al, 2015)). This new version of classical mouse-tracking paradigms has previously been developed for moral statements and scenarios, for which processing time is often longer than images or single words typical of most mouse-tracking studies (Gaboriaud et al, 2022; Gautheron et al, 2023).

After reading instructions, participants completed two training blocks. Training block 1 included six food images to learn how to use the slider paradigm with simple images (to be categorized as liked or disliked). Training block 2 presented 6 actions in their different versions (gain or loss of small or large amounts of money for each action), exactly as in the testing blocks. The remaining actions for these testing blocks were randomly presented within six blocks of 52 items. Each block ended with a self-paced break preceded by an attention check consisting in asking participants to either answer all right or all left on the answer scale. Upon completion, participants were thanked and fully debriefed.

2.4 Results

2.4.1 Analytic procedure

For effects to be comparable and averaged across positively and negatively valenced actions, x-coordinates ($X(t)$) were remapped based on valence. As individual differences regarding valence can be expected, we estimated for each participant each action’s valence. We expected a natural order of final x-coordinates for the decision (-100€, -1€, +1€ and +100€), from the greatest loss (discouraging the action and leading to responses towards “Not at all”, corresponding to $X = -1$) to the greatest gain (encouraging the action and leading to responses towards “Absolutely”, $X = 1$) (see Fig. 1). We therefore computed for all actions the median of x-coordinates for all conditions. Valence categorization was subsequently based on the sign of this median, which was negative (resp. positive) for negatively (resp. positively) valenced actions. This method implies that conflict is derived from the raw dependent variable and is defined as a within-participant within-action variable. To compare all conflict trials, we reversed the answer scale of all negatively valenced actions to only have positive medians (i.e., remapped to the right side of the scale). With such recoding, all effects of conflict go in the

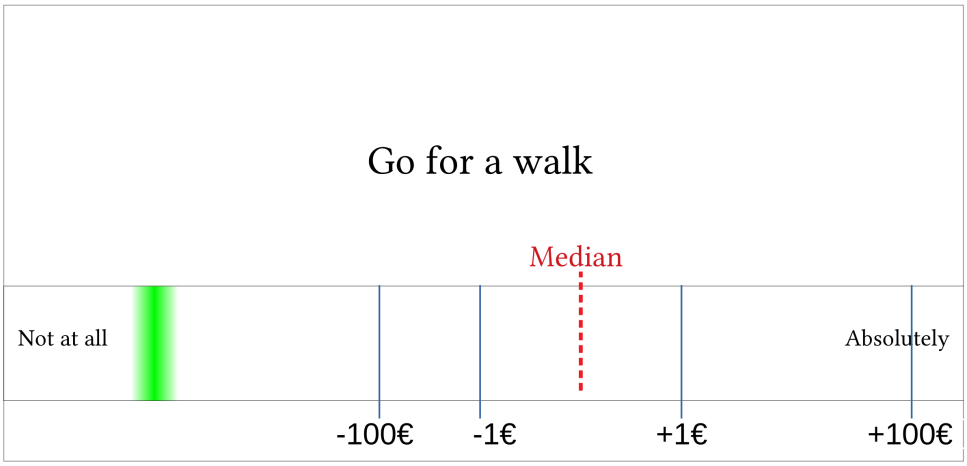

Fig. 1 Illustration of how valence is calculated for each action based on the median obtained from answers in all conditions.

same direction, although within and between participants variability is expected. A monetary loss in conflict with positively valenced actions should push responses towards the left side of the scale, resulting in negative shift in coordinates. A monetary gain in conflict with negatively valenced actions should push responses towards the right side of the scale, resulting in a positive shift in coordinates, becoming negative once remapped. This remapped variable (X' for only the final coordinate, $X'(t)$ for dynamic coordinate during decision process) will be the one used for all analyzes, as we focus on the shift of decision rather than raw decisions.

Mixed models were used to correctly estimate the different effects, hence allowing generalization at the participant and action levels. analyzes were performed using R language, lme4 package Bates et al (2015) and associated packages (e.g., DHARMa for diagnostics and emmeans for comparisons). Morality of the action (M, contrast coded -0.5 = nonmoral vs. 0.5 = moral), presence of conflict (-1€/-100€ when action positively valenced, or +1€/+100€ when action negatively valenced; C, contrast coded -0.5 = non conflictual vs. 0.5 = conflictual), strength of conflict (S contrast coded -0.5 = 1€ vs. 0.5 = 100€) were fixed factors, and participants (P) and actions (A) random factors.

2.4.2 Remapped final responses

The following maximal model specification was used (in R format):

$X' \sim M * C * S + (M * C * S|P) + (C * S|A) + (C|A : P)$

Results are illustrated in Fig. 2 as a function of morality, conflict, and strength of conflict. Results indicate a main effect of morality, with more extreme answers for moral ($M = 0.58$, $SE = 0.03$) compared to nonmoral actions ($M = 0.33$, $SE = 0.04$), $b = 0.25$, $SE = 0.03$, $\chi^2 = 63.46$, $p < .001$, 95% CI $[0.19, 0.32]$, $d = 0.56$. A main effect of conflict was found, with less extreme responses for conflictual ($M = 0.23$, $SE = 0.03$) compared to non-conflictual ($M = 0.68$, $SE = 0.04$) actions, ($b = -0.44$, $SE = 0.05$, $\chi^2 = 88.82$, $p < .001$, 95% CI $[-0.54, -0.35]$, $d = -0.98$), and a main effect of strength of conflict, with less extreme responses for the high-conflict condition ($M = 0.39$, $SE = 0.03$) compared to the low conflict condition ($M = 0.52$, $SE = 0.03$), $b = -0.13$, $SE = 0.02$, $\chi^2 = 43.03$, $p < .001$, 95% CI $[-0.17, -0.09]$, $d = -0.29$. Hence, less extreme (or more nuanced) decisions were found for nonmoral decisions, conflictual decisions, and when the conflict is more pronounced.

A two-way interaction effect between conflict and morality was observed ($b = 0.45$, $SE = 0.05$, $\chi^2 = 80.37$, $p < .001$, 95% CI $[0.35, 0.54]$, $d = 0.49$) with nonmoral actions being more impacted than moral actions by the presence of conflict. Simple main effects further indicate that there was no significant difference between moral and nonmoral actions in the no-conflict case, $b = 0.03$, $SE = 0.02$, $t(66.42) = 1.95$, $p = .06$, 95% CI $[-8e-04, 0.06]$, while this difference was significant in the presence of conflict, $b = 0.48$, $SE = 0.05$, $t(90.47) = 8.67$, $p < .001$, 95% CI $[0.37, 0.59]$. Therefore, conflict exerts a more substantial influence on nonmoral decisions, with this effect becoming increasingly pronounced as the strength of the conflict escalates. Thus, the presence of conflict led to significant changes in answers in all conditions, but with more nuanced or even reversed answers for nonmoral actions compared to moral actions.

The interaction between morality and strength of conflict is also significant, with more nuanced answers for nonmoral items compared to moral items in the low conflict condition, with an increase of this phenomenon in the high conflict condition ($b = 0.14$, $SE = 0.03$, $\chi^2 = 28.56$, $p < .001$, 95% CI $[0.09, 0.18]$, $d = 0.15$). The last interaction between the presence of conflict and its strength shows that the effect of conflict is amplified when the conflict is high rather than low, with more nuanced answers in high conflict condition only when conflict is present, and there is no significant difference in the no-conflict condition ($b = -0.42$, $SE = 0.05$, $\chi^2 = 86.77$, $p < .001$, 95% CI $[-0.51, -0.33]$, $d = -0.46$).

The three-way interaction between all the fixed factors was also significant ($b = 0.38$, $SE = 0.06$, $\chi^2 = 41.28$, $p < .001$, 95% CI $[0.26, 0.49]$, $d = 0.21$). Precisely, the effect of morality, with more nuanced answers for nonmoral compared to moral actions, was amplified in the high-conflict condition (100€, $b = 0.32$, $SE = 0.04$, $t(90.37) = 8.39$, $p < .001$, 95% CI $[0.24, 0.40]$) compared to the low-conflict condition (1€, $b = 0.19$, $SE = 0.03$, $t(83.67) = 6.27$, $p < .001$, 95% CI $[0.13, 0.24]$), with even more nuanced answers in the high-conflict condition in the presence of conflict, for both moral ($b = 0.38$, $SE = 0.05$, $t(88.3) = 7.98$, $p < .0001$, 95% CI $[0.29, 0.48]$) and nonmoral ($b = -0.26$, $SE = 0.06$, $t(73.1) = -4.07$, $p = .0001$, 95% CI $[-0.38, -0.13]$) decisions. Hence, decisions are all the more nuanced as they are nonmoral (vs. moral) and highly (vs. lowly) conflicted (vs. non conflicted). Even if statistically robust, these results could be attributed to differences in response distributions between samples of nonmoral and moral actions. For instance, given the large inter-item variability observed, a few moral items may have saturated the response scale, with no potential for compromise even in the high conflict condition, resulting in more stable responses for moral actions on average. Although this would still reflect actual differences in processing and decision-making for the selected moral vs. nonmoral actions, these differences could be attributed to sampling biases. The following analyses eliminate this possibility by turning responses in the low conflict condition into a model predictor, making it possible to test the effect of conflict for nonmoral and moral actions which lead to the exact same response in the near-absence of conflict.

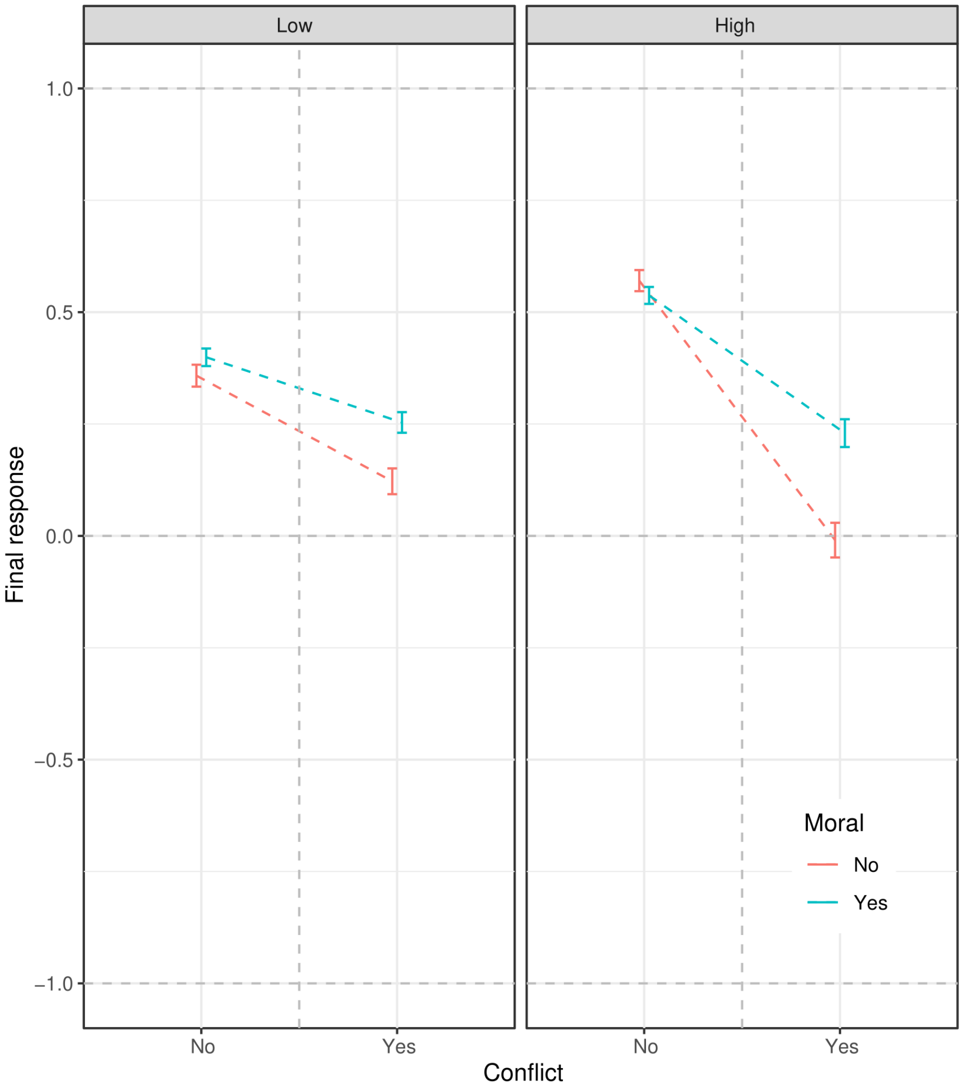

Fig. 2 Remapped final answer in the empirical study as a function of morality, conflict, and strength of conflict. Error bars correspond to *SE*. Left: condition of low conflict (1€ at stake). Right: condition of high conflict (100€ at stake).

2.4.3 Final responses conditioned by answer extremeness

As preregistered, to consolidate conclusions while simplifying the interpretation and specification of the analytical model, we also introduced separate variables for high (X'_{100} with 100€ at stake) and low (X'_1 with 1€ at stake) gain/loss, the latter serving as a baseline condition with low conflict intensity. This allowed to directly test changes in decisions based on the presence of conflict (C ; +100€ vs. -100€ depending on action valence), presence of morality in actions (M), conditioned upon responses in the low conflict conditions (X'_1 for +1€ or -1€), using the following model specification:

$X'_{100} \sim M * C * X'_1 + (M * C * X'_1|P) + (C * X'_1|A) + (1|A : P)$

Results confirm the significant interaction between morality and conflict already observed in the previous unconditional analyses ($b = 0.46$, $SE = 0.05$, $\chi^2 = 72.90$, $p < .001$, 95% CI $[0.35, 0.56]$, $d = 0.47$); there is no significant difference between moral and nonmoral answers in the no-conflict condition ($b = -0.03$, $SE = 0.01$, $t(97.86) = -1.94$, $p = .06$, 95% CI $[-0.06, 0.0006]$), but a positive effect of morality in presence of conflict ($b = 0.52$, $SE = 0.06$, $t(81.87) = 9.20$, $p < .001$, 95% CI $[0.41, 0.63]$). In presence of conflict, there is a significant effect of morality for answers made in the extreme, as the moral decisions that were already extreme can only be more nuanced but to a lesser extent compared to nonmoral decisions ($b = 0.65$, $SE = 0.08$, $t(83.32) = 8.28$, $p < .001$, 95% CI $[0.50, 0.81]$). When answers were made close to the middle of the response area, having a higher conflict did not have a significant effect on moral decisions ($b = 0.056$, $SE = 0.04$, $t(41.8) = 1.39$, 95% CI $[-0.03, 0.14]$, $p = .17$), but led to change the side of the answer for nonmoral decisions ($b = -0.33$, $SE = 0.54$, $t(51.1) = -6.16$, 95% CI $[-0.44, -0.22]$, $p < .001$). A graphical representation of these results and associated response distributions are provided in supplemental files. Together, findings confirm a different effect of conflict depending on whether the decision is moral or not, which is not explained by the extremeness of moral decisions itself. In other words, whether they are already extreme or more nuanced, decisions with a moral component result in more stable positions compared to decisions without a moral component.

2.4.4 Response times

Regarding response times, we ran the following model, which was not preregistered:

$log(RT) \sim M * C * S + (M * C * S|P) + (C * S|A) + (C|A : P)$

We observe a main effect of conflict with participants being faster for non-conflictual decisions ($M = 2.34$, $SE = 0.12$) than conflictual ones ($M = 2.55$, $SE = 0.14$) ($b = 0.09$, $SE = 0.01$, $\chi^2 = 37.55$, $p < .001$, 95% CI $[0.06, 0.11]$, $d = 0.17$), and also an effect of strength of conflict with faster answers in the low conflict ($M = 2.39$, $SE = 0.13$) vs. high conflict ($M = 2.50$, $SE = 0.13$) condition ($b = 0.05$, $SE = 0.009$, $\chi^2 = 25.01$, $p < .001$, 95% CI $[0.03, 0.07]$, $d = 0.09$). A significant two-way interaction effect was found between the presence of conflict and its strength (low or high) with even longer response times in presence of conflict in the high-conflict compared to the low conflict condition ($b = 0.05$, $SE = 0.02$, $\chi^2 = 7.01$, $p = .0081$, 95% CI $[0.01, 0.09]$, $d = 0.05$). Thus, the presence of conflict leads to longer decisions, and even longer when the conflict is high. No other interaction was significant.

2.4.5 Mouse trajectories

We analyzed mouse trajectories, for which we averaged mouse coordinates over time bins. We conducted this analysis after filtering trajectories where we could observe biased mouse movements (i.e., more than 10 back and forth movements), hence reducing noise in the data (for more details, see supplemental files). These trajectories no longer indicated hesitations but a strategy to extend trial time. They amounted to 35% of the trials. We used the same linear mixed model as in the main analysis on final responses, but for each time step (for consistency and comparison purposes). To reduce the number of tests through time, analyzes were performed on time bins on time-normalized trajectories, each time bin corresponding to 5% of the entire decision process duration. Time therefore runs between 0% when clicking on the START button to 100% when clicking on the final response. We only report significant effects over more than 2 successive time bins, hence correcting for multiple testing with dependencies over 20 time bins. In addition, to check for potential biases that could result from time normalization when participants or items highly differ in response times distributions, we estimated the same model on raw trajectories, starting from the initial click on the START button (0 second) until 2 seconds of decision time. Beyond controlling for methodological or analytical biases, this allows knowing when, during the process, the different conditions lead to significant bifurcations, while still comparing decisions that were made quickly. Results on time-normalized trajectories are represented on Fig. 3, while the convergent pattern for raw time analyses is illustrated in the supplemental files.

Regarding the main effect of morality, trajectories tend to split at 65% of the decision duration ($b > 0.05$, $SE > 0.03$, $\chi^2 > 11.99$, $d > 0.17$) and from 1s ($b > 0.04$, $SE > 0.02$, $\chi^2 > 6.37$, $d > 0.16$), with more extreme positions for moral decisions. For the conflict main effect, non-conflictual decision trajectories were significantly more extreme from 65% of the trajectories ($b < -0.06$, $SE > 0.02$, $\chi^2 > 14.73$, $d < -0.20$) and 1.2s ($b < -0.03$, $SE > 0.009$, $\chi^2 > 14.43$, $d < -0.10$). The main effect of conflict strength became significant from 75% ($b < -0.02$, $SE > 0.01$, $\chi^2 > 6.33$, $d < -0.07$) and 1.7s ($b < -0.03$, $SE > 0.01$, $\chi^2 > 9.11$, $d < -0.07$), with more extreme positions in the high conflict condition.

The interaction effect between morality and conflict showed more extreme positions for moral decisions going even more extreme in presence of conflict from 65% of the trajectories ($b > 0.06$, $SE > 0.02$, $\chi^2 > 11.02$, $d > 0.10$) and 1.3s ($b > 0.06$, $SE > 0.02$, $\chi^2 > 14.68$, $d > 0.09$). The interaction between morality and conflict strength was significant from 90% of the trajectories with even more extreme position in the moral and high conflict condition ($b > 0.05$, $SE > 0.02$, $\chi^2 > 5.69$, $d > 0.06$), but was not significant in the two first seconds of the decision. The interaction between conflict and its strength showed even more extreme positions for highly conflictual decisions from 70% of the trajectories ($b < -0.07$, $SE > 0.02$, $\chi^2 > 17.53$, $d < -0.12$) and 1.5s ($b < -0.08$, $SE > 0.02$, $\chi^2 > 18.95$, $d < -0.10$).

Finally, the three-way interaction between morality, conflict, and strength was significant from 75% ($b > 0.12$, $SE > 0.04$, $\chi^2 > 10.47$, $d > 0.08$) and 1.3s ($b > 0.11$, $SE > 0.03$, $\chi^2 > 6.12$, $d > 0.07$). This interaction, similarly to the one obtained on final responses analysis, consists in more extreme trajectories for moral vs. nonmoral decisions, that are even more extreme when conflicted, and, again even more in high conflict condition. Precisely, conflictual decisions are those where hesitations and deviations from the center are earlier and more pronounced (beginning at 25%, $b > 0.12$, $SE > 0.005$, $t > 2.48$), compared to non-conflictual decisions where trajectories deviate significantly at a later time (beginning at 55%, $b > 0.45$, $SE > 0.005$, $t > 2.48$). Particularly interesting, we also observe a significant decision reversal for highly-conflictual nonmoral decisions, going from a positive mean from 40% to 50% of the trajectory, peaking at 50% ($b = 0.039$, $SE = 0.014$, $t(38.89) = 2.72$, $p = .0097$, $95\% CI$ [0.01, 0.07]), to a negative mean beginning at 80% of the trajectory increasing until the end ($b < -0.13$, $SE < 0.06$, $t < -4.20$). This means that the mouse cursor was

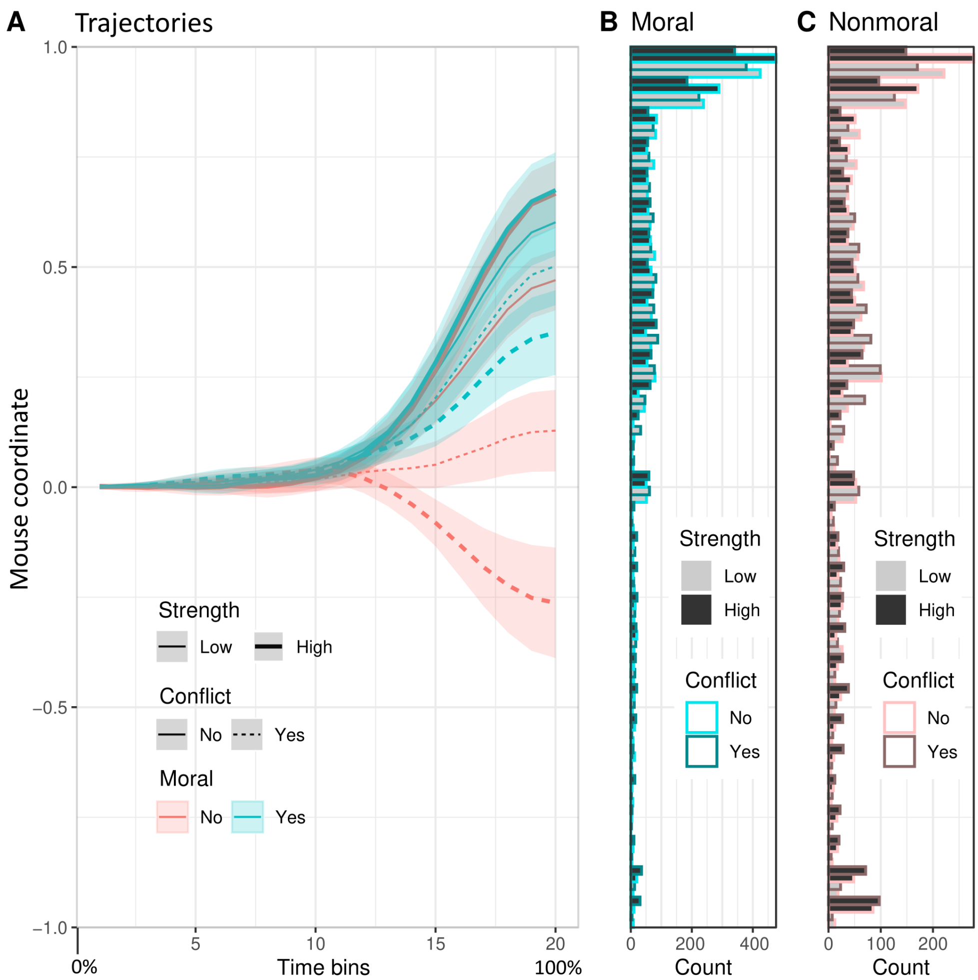

Fig. 3 (A) Estimated marginal means on filtered empirical mouse trajectories as a function of morality, conflict, and strength of conflict. Error bands correspond to *95% CI*. (B) Final answer distribution for conflictual / nonconflictual moral decisions. (C) Final answer distribution for conflictual / nonconflictual nonmoral decisions.

attracted toward one side of the response area, before bifurcating on the other side in the middle of the trajectories. Regarding nonmoral conflictual decisions, they are significantly diverging from the center from 40% to 55% of the trajectory ($b < 0.031$, $SE < 0.012$, $t < 2.65$) and 1.1s ($b = 0.36$, $SE = 0.015$, $t > 2.42$), whereas they start at 60% when there is no conflict ($b > 0.06$, $SE > 0.016$, $t > 4.04$) and also at 1.1s ($b > 0.05$, $SE > 0.015$, $t > 3.51$). On the other hand, moral conflictual decisions are significantly diverging from the center earlier, from 30% ($b > 0.19$, $SE > 0.07$, $t > 2.64$) and 0.5s ($b > 0.14$, $SE > 0.004$, $t > 3.25$) of the trajectory, whereas deviations become significant from 45% ($b > 0.02$, $SE > 0.008$, $t > 2.74$) and 0.6s ($b > 0.016$, $SE > 0.007$, $t > 2.14$) when there is no conflict. Globally, moral decisions significantly deviate from the center earlier (from

35%, $b > 0.018$, $SE > 0.006$, $t > 2.73$), than nonmoral decisions (from 55%, $b > 0.036$, $SE > 0.011$, $t > 3.21$).

2.5 Discussion

Results on final responses indicate that conflict leads to nuance for both moral and nonmoral decisions, an effect that is stronger as strength of conflict is more important. Notwithstanding this shift for moral and nonmoral decisions, moral decisions were more stable, with this shift being less pronounced compared to nonmoral decisions.

Concerning response times, longer response times were observed in the presence of conflict (vs. no conflict), and even more so when conflict was high (vs. low). This finding is in line with an interpretation in terms of cognitive load and increased deliberation induced by conflict, but here induced by stimuli per se. However, as no main effect of morality was found, findings do not align with more reasoning demands, given greater implication, when a moral component is included.

Overall, mouse-tracking data indicate that some factors are impacting the decision process earlier than others. First, contrasting with findings on response times, we observed that moral decisions deviated earlier than nonmoral decisions. Given that there was no significant response time difference between moral and nonmoral decisions, this result suggests that moral content is processed faster than nonmoral content, supporting the relevance of digging into finer-grained decision dynamics. Second, regarding normalized mouse trajectories, we observed earlier deviations from the center for conflictual decisions, albeit not confirmed with raw time trajectories. Third, highly conflictual nonmoral decisions initially followed the pattern of non-conflictual nonmoral decisions before deviating in the opposite direction, presumably due to the monetary gain/loss condition affecting the decision later, or a bias encouraging answering toward the right answer (i.e., "Absolutely"). It is also important to note that a substantial amount of data was excluded from the analyses due to oscillating trajectories. Given our repeated measures design and mixed model analyses, this did not generate any singularity or unbalancing of data across conditions. While this mouse-tracking design has already been successfully used to study moral decision-making (Gautheron et al, 2023; Gaboriaud et al, 2022), the presence of conflict in the present study might have increased the need for such oscillatory strategy in order for participants to prolong trials. We assume that, depending on the difficulty of the decision task, the basic time constraint may be adjusted to avoid these strategies, while still encouraging constant mouse movements.

To strengthen these empirical results with a complementary approach and further understand moral and nonmoral decision-making mechanisms in conflictual or nonconflictual cases, we built a computational model. Notably, it could explain how the earlier processing of moral content could lead to more stable final positions in conflicted decisions. Additionally, an intriguing query arising from our experimental design—which combines a nonmoral monetary incentive condition with an action that can either be moral or nonmoral—is whether the presence of a moral component compels individuals to interpret all elements from a moral perspective, subsequently processing them all as moral elements. Alternatively, it is plausible that the nonmoral condition is still processed distinctly—as a nonmoral element-while the action, in contrast, is processed as a moral element (exerting a greater or earlier influence, as illustrated in empirical mouse trajectories). The implementation of these features in different model versions may therefore inform about underpinning decision processes and may be a substantial help in understanding how moral and nonmoral elements are integrated during decision.

3 Model

As detailed in the introduction, we used a sequential sampling model allowing to study underlying decision mechanisms, especially how and when the evidence is accumulated. A motor module was incorporated to this model, tracing decision processes in mouse movements (similarly to empirical data). By running simulations, outputs can directly be compared with all the empirical data of interest: final decisions, response times, and mouse trajectories. Through this model, we proposed an operationalization of decisional conflict, projecting moral and nonmoral elements (or stimuli components) in a common decision space and putting them in competition. Those elements correspond to the different factors, or evidence, taken into consideration in the decision. When at least one of them belongs to the moral domain, the decision is considered as moral, then possibly leading to process nonmoral elements in a similar way to moral elements. The model thus represents how different elements are voting for different locations in the decision space, and how they lead to the final decision by attracting the preference to their position during decision process. Relying on the same dynamics, the model is able to represent how different information pieces aggregate when they are voting for close locations, reinforcing preference and commitment to answer at this specific location, or converging to an intermediate location as a trade-off between conflicting information.

In a first step, we compared parameter values fitted either for moral or nonmoral decisions, following the assumption that they underlie distinct decision processes. Thus, even if the decision process occurs within the same space (e.g., going from 'absolutely willing' to 'absolutely not willing' to take the action) and may rely on the same underlying mechanisms, the processing and integration of the different elements may nevertheless diverge between moral and nonmoral decisions. This specific moral decision processing would hold true, even when not all decision elements pertain to the moral domain (i.e., when a moral action is associated with a nonmoral monetary incentive). Indeed, the mere presence of a single moral element is sufficient to categorize the decision within the moral domain, and this categorization could influence the processing of other nonmoral elements. In a second step, we tested the need for additional parameters to describe how moral or nonmoral elements may intervene differently (i.e., with facilitated or earlier processing) in the decision making process, while still sharing the same decision space.

3.1 Dynamic evolution

Based on the neural fields theory Schöner and Spencer (2016), the core of the present model is governed by a stochastic integro-differential equation:

$$\tau \frac{\partial u(x,t)}{\partial t} = -u(x,t) + \int_X \omega(x-x')\sigma(u(x',t))dx' + i(x,t) + \varepsilon(x,t) \tag{1}$$

where u is the activation vector of the units, τ is the time constant used to control the integration speed of the system, i is the input, ω is the lateral kernel function defining how units interact (exciting or inhibiting each others), σ is an activation function (nonlinear ReLU operation) used to avoid negative values and to determine when units exchange information, ϵ corresponds to Gaussian noise. X is the input decision space (the space of the neuronal field where the decision takes place) ranging from -1 to 1 (arbitrary scale with symmetric extremes) and x is the valence parameter of u, such that $u(x,t)$ is a function of "valence" and time. Also, the decision space does not necessarily correspond to the answer space (the set of response options provided). An example of this discrepancy might occur when a participant's true preference is more extreme (e.g., "detest") than the available options (e.g., "dislike"). To address this, we set a decision space larger than the answer space, in a way that participants' true feelings, even more extreme ones, are considered, even though they can only choose from limited options (see Fig. 4 for an example). Subsequently, a mapping is then operated to select the available alternative that most faithfully corresponds to the decision or preference in the decision space. This specification relies on the idea of a decision space as borderfree as can be the neurobiological space (physically through neuronal connectivity over the entire brain, and theoretically with the conceptualization capacity allowing to think "out of frame"). From a computational perspective, it is also generally necessary to prevent border effects (as the lack of connection near borders of the decision space leads to biased probabilities of convergence on the associated decision).

For our model implementation and numerical simulation, we operated a spatial discretization (on a lattice of n units as an approximation of the continuous topology, of vector form because the answer space in one-dimensional) and temporal discretization (integration scheme of Euler-Murayama) giving for each position j in the decision space:

$$\Delta u_j = \frac{\Delta t}{\tau}\left(-u_j + \omega * \sigma(u) + i)\right) + \Delta W \tag{2}$$

During this discretization, the time step Δt must be small enough to make the approximation of the Euler-Murayama scheme good enough (here taken equal to 0.005 seconds). The Gaussian noise is discretized and integrated into a Wiener process W (with $\Delta W = \epsilon\sqrt{\Delta t/\tau}$).

The units interact with each other thanks to a weight function ω of a lateral-inhibition type neural field (in equation 1). The equivalent convolution kernel in equation 2 is modeled as a constant global inhibition combined with a local Gaussian excitation profile aiming to obtain a unique final decision (i.e., a single emergent activation bump over the vector u; see equation 3). It models the fact that we tend more to hesitate between two answers that are close (that is to say, we won't hesitate between "Not at all" and "Completely", but more likely between "Not at all", and "Not really"). In total, three different parameters make it possible to specify this kernel: the amplitude (named A) and dispersion parameters (a) of the excitatory Gaussian, and the global inhibition amplitude (B):

$$\omega(\Delta x) = Ae^{-\frac{\Delta x^2}{2a^2}} - B \tag{3}$$

Parameters A and B control the intensity of the competition between the two choices, that is to say in what proportion (accordingly to the energy of the input) the contribution of an evidence will increase the activation of the corresponding unit, similarly to the drift in the DDM. The parameter a, meanwhile, controls the spatial scale of the activation bump. In other words, the system will rapidly evolve by letting a so-called activation bump emerge, with

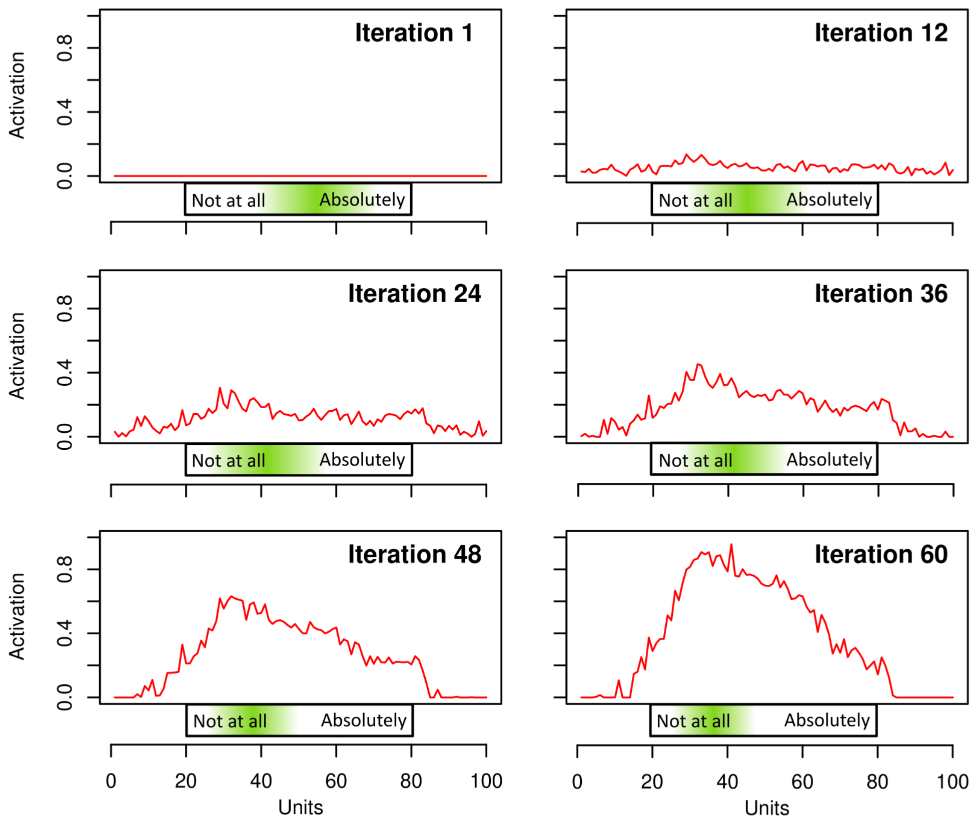

Fig. 4 Example of how unit activations can evolve during decision process. At the beginning, units are not activated. While input information is integrated, a bubble of activation begins to form somewhere in the decision space (i.e., the unit space), until reaching the activation threshold (here equal to 1) that triggers the final decision. At each iteration, the barycenter determines the current preference (represented as the green slider) in the answer space (represented as the response bar going from 'Not at all' to 'Absolutely'). As implemented in the model, the decision space is here larger than the decision space to account for more extreme reasoning that may not be present in response options.

all the units under this bump strongly activated and those around inhibited. Parameter a thus manages the size of this bubble, or the number of neighboring units strongly activated at the same time.

The Wiener process W models different kinds of stochastic variations, likewise the diffusion in DDM/LCA. This could include, not exhaustively, noise in the input (inter and intra-individual variations), noise in the decision process (influence of the experimental context, or the influence of the previous trial), noise due to the evaluation metric on the output (uncertainty in the measurement of the trajectory, or the fact that the trajectory is only an indirect measure of the decision-making process), or any noise in neural signals. The time constant τ affects the speed of response. The lower it is, the faster the integration of information, but also the higher the noise integration. Finally, the input vector i incorporates bumps positioned where the evidence supports associated responses on the answer space. It thus corresponds to the information that is taken into consideration and which will guide the preference toward an alternative or a region in the answer space. It can be diffuse, as an evidence can vote for an extended region if not precise and/or certain. An illustration of the evolution of the $u(x)$ vector during the decision process is presented in Fig. 4.

3.2 Modeling conflict

Conflict arises when two (or more) irreconcilable elements are pushing in different directions or toward different alternatives in the decision. We set up our experiment in order to have the action and the condition with money at stake each voting for a location in the answer space (the continuum between 'Not at all' and 'Absolutely'). They were sometimes in agreement in the same direction, sometimes in the opposite direction. We model the conflict as the phenomenon of two (or more) bumps / sources in input competing across different locations in decision space (the space where the model is evolving) (see Fig. 5 for an illustration). With the previously introduced lateral kernel and adequately chosen model parameter values, this competition process leads to the emergence of a single attractor in the decision space. The distance between the two bumps in input leads to induce more or less conflict, with very distant bumps leading to *high conflict*, and reversely less distant bumps to *low conflict*. It thus represents how different elements are voting for different locations in the decision space, and how they lead to the final decision by attracting the preference to their position during decision process. Relying on the same dynamics, the model is able to represent how different information pieces aggregate when they are voting for close locations, reinforcing preference and commitment to answer at this specific location, or converging to an intermediate location as a trade-off between conflicting information.

In order to model how different elements of the decision are impacting the decision process, we considered that the action and the condition (gain/loss of money) were projected as two Gaussians in the input vector. In order to determine where those Gaussians are situated, we calculated their valence in order to have an idea of the polarity of each item. For that, we used a set of rated valence words (Stadthagen-Gonzalez et al, 2017) and calculated the total valence when the action encompassed more than one word. This simplification ensures that we have a unique a priori valence for each action in our set of stimuli, as they relate to a single idea if taken as a whole. We also used the variance of the set of rated valence words to specify the variance of the associated Gaussian. For the condition, we modulated the valence of the words 'gain' and 'loose' according to the amount of money at stake. For that, we set plausible a priori parameters. As our participants were students, we can easily assume that they would always really want to gain 100€ (or avoid losing 100€) but were way less interested by gaining 1€ (and be little affected by losing 1€). We therefore set an extreme value when 100€ was at stake (value of 0.9 for a gain and -0.9 for its loss), and a moderate value, closer to the neutral position, for 1€ at stake (0.3 for a gain, and -0.3 for a loss). We also checked for possible bias in valence distribution between moral and nonmoral items. We found that the mean valence was significantly more positive for nonmoral items than for moral items on the same scale used for the experiment (i.e., with ratings going from -1 to +1) ($b = 0.52$, $t(88.9) = -5.23$, $p < .001$, 95% CI $[-0.72, -0.32]$). However, they did not differ in valence extremeness (comparing extremeness for both negative and positive valences by computing the absolute value) ($b = 0.105$, $t(71.66) = -1.55$, $p = .13$, 95% CI $[-0.23, 0.03]$).

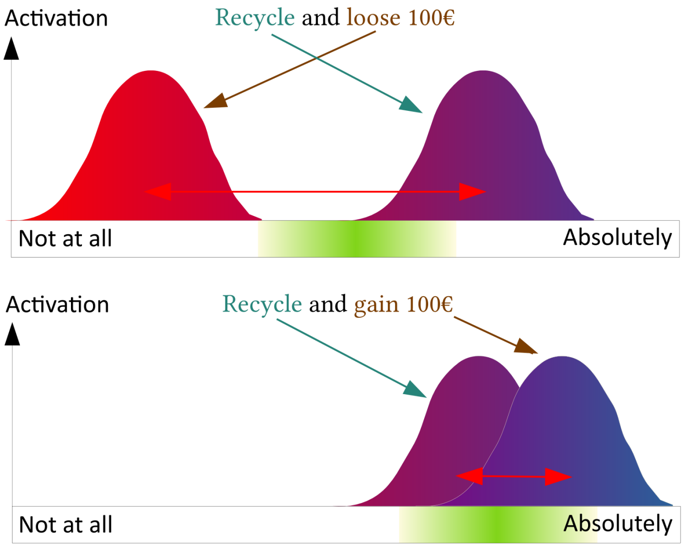

Fig. 5 Example of how conflict intervenes depending on the distance between locations associated to each piece of evidence (here actions in the stimuli set). Conflict strength also directly depends on this distance, increasing when the distance is high and reciprocally reducing when the distance is low.

Here, our focus is mainly on extremeness, given that the same analyzes as in the empirical study were conducted. Hence, as we did not observe a significant difference between distribution of valence extremeness or moral and nonmoral actions, this specification was not sufficient to explain moral and nonmoral decision differences.

3.3 Modeling moral content

As evidenced in the literature (Gantman and Van Bavel, 2014; Moll et al, 2002; Brady et al, 2020; Van Bavel et al, 2012; Gantman and Van Bavel, 2015; Haidt, 2001), moral content could be processed differently from nonmoral content. Modeling of decision-making processes can shed light on mechanistic explanations of such differences through differences in parameter values or parametrizations. Sequential sampling models could thus reveal underlying decision components and functioning. However, the reference model introduced in section 3.1 and 3.2, as it is, may not adequately or completely capture how morality is processed during decision. A single additional feature may be needed (for testing a facilitated or early integration of moral information), aiming at the most parsimonious model possible to explain the observed pattern of responses in empirical data.

3.3.1 Version 1, amplitude model - Amplified moral perception

In order to model how moral factors could lead to more stable and stronger positions, we decided to implement this phenomenon with an increased weight for moral evidence in the decision process, as it seems that there is an increased perceptual awareness for moral elements (Gantman and Van Bavel, 2014). We thus specified the amplitude of the Gaussian in the

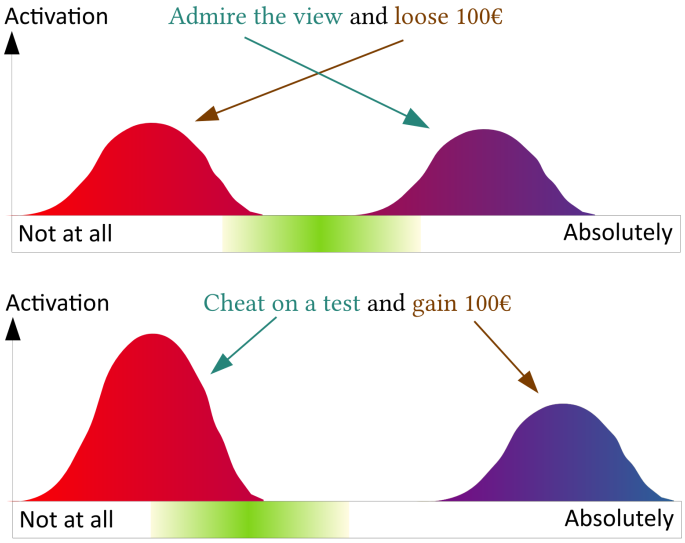

Fig. 6 Example of how moral information could intervene differently in the decision making process, here with a greater amplitude in the input signal.

input vector corresponding to moral actions as higher than for nonmoral components (nonmoral actions and conditions of gain/loss of money) according to a certain factor $ampl_m$ (see Fig. 6 for an illustration).

3.3.2 Version 2, delay model - Early moral integration

Other conceptual models consider that moral stimuli are processed faster, in an intuitive or emotional way (Haidt, 2001). Even if this idea has been nuanced, emphasizing the importance of the situation in how intuitive or rational is moral cognition (Monin et al, 2007), the intuitive response could occur especially when facing strongly morally-laden stimuli (such as "murder" or "cheating"). Corroborating this hypothesis, empirical trajectories showed an earlier significant deviation from the center for moral decisions compared to nonmoral decisions.

We also implemented this version of the model - with an earlier impact of moral factors in the decision-making process, in order to compare the fit and results from the two extended models. The nonmoral factors thus appear in the input after a certain number of iterations (modeled with the $delay$ parameter). This specification (earlier impact of moral elements) leads to a stronger influence of moral elements, due to nonlinear bifurcations in our DNF-based dynamical system. Disregarding stochastic variability, at any given time during the decision process simulation, more evidence should be accumulated for locations in the input with a stronger or earlier activation, giving them a competitive advantage, leading to an increased self-excitation and inhibition of alternative decisions. The two extended versions of the original model are therefore both linked to the “saliency” of moral content, but in different ways. Indeed, their dynamics differs: the impact of nonmoral elements on decision process is delayed compared to the first model version, thus impacting decision temporality.

This could be reflected through response times and in a finer grained manner in generated decision trajectories.

For the two versions (amplitude and delay models), we suppose that the underlying decision mechanism is the same for moral and nonmoral decisions (i.e., that there is no variation in the internal parameter values of the model). We assumed that the minimal addition of a single parameter could be sufficient to explain the impact of morality on decision-making. This approach relies on the idea of a specific processing of moral elements, contrasting the assumption that the whole decision process could be impacted by moral situations.

3.4 Motor part

The model has also been enriched with a motor part to generate mouse trajectories during the decision-making process, with the aim of reflecting decision making dynamics found in human mouse-tracking data.

We used a method inspired from Lepora and Pezzulo (2015), and Quinton et al (2013) (eq. 5 & 6), where the model for mouse trajectory production takes as input a preferential position or temporary answer at each time. The trajectory is then produced by the attraction of the cursor to this position.

In our DNF-based model, the position is calculated as a barycenter in the decision space in the following way: at each time step, the units have a certain activation u_j. The position x_{tgt} is the weighted average of units position according to their activation:

$$x_{tgt} = \frac{\sum_{j=1}^{n} x_j u_j}{\sum_{j=1}^{n} u_j} \tag{4}$$

Each x_j thus corresponds to a position on the decision space, that will be remapped in the response axis. The target position x_{tgt} reveals the preference in the response space, given the actual gathered evidence and decision step. Hence, this specific point is attracting the mouse cursor in order to have an adequacy between the position of the mouse and the decision preference when the final decision will be taken (e.g., through a mouse click in human experiments).

To model this attraction, we based our model on the following equation, discretized temporally in the same way as the equation of the global system:

$$\frac{dx_p}{dt} = -k(x_p - x_{tgt}) \tag{5}$$

with x_p the mouse coordinate on screen and x_{tgt} the coordinate computed out of the DNF part of the model at each time step. Here the time step dt is the same as in the equation equation 2. The stiffness k (if using the metaphor of a physical spring) makes it possible to control the speed at which the cursor is attracted by the calculated final position: the higher it is, the faster the cursor will move. This parameter is dynamic as it depends on the mean activation of all units, associated to a scaling parameter λ to control maximal movement speed. It models the fact that the higher the unit activations are, the higher is the preference for the corresponding choice area, and thus the faster the movement will be to reach this area. It is defined as:

$$k = \frac{\lambda}{n} \sum_{j=1}^{n} u_j \tag{6}$$

3.5 Final decision - stop criterion

In order to simulate the final mouse click in response areas, we consider that the decision is reached when (i) at least one of the units is sufficiently activated - with a corresponding set threshold (dynamic criterion determined by the non-linearity and parameters in equation 2, which also corresponds to the formation of a stable bubble of activation); (ii) and reaching the target position with the mouse in the motor part was achieved (motor criterion is met when the difference between the target and the current cursor coordinates is smaller than a certain value).

These two criteria correspond to the convergence in the differential equations controlling both the decision and motor components of the model. A time limit is also set to interrupt the simulation if needed, in order to prevent configurations in which the model would not reach convergence quickly enough. Indeed, very long decision times and trajectories seldom occur in empirical data, as humans regulate their behavior to satisfy task constraints.

3.6 Results

3.6.1 Model fit

The model parameters of interest are the convolution kernel parameters A, B, a, the noise ε, the time constraint τ and the scale parameter λ. The remaining parameters are set to a priori fixed values, derived from previous model fits, obtained on similar paradigms but different tasks, relying on the flexibility and robustness of DNF models dynamics (Gautheron et al, 2023). The parameter number is then reduced at the minimum to reduce the model's complexity and computational processing time required for fitting. To fit human data and be able to compare the different versions of our model, we calculated the root-mean-square error (RMSE) between the empirical and model generated final response coordinates.

3.6.2 Reference model - Moral and nonmoral decisions as two different decision processes

First, we applied our reference model without introducing any specific differentiation for moral and nonmoral items. Then, this model assumes that the presence of a moral situation compels individuals to interpret all elements from a moral perspective, subsequently processing them all as moral elements. We therefore conducted independent fitting of model parameters for moral and nonmoral decisions, thus facilitating a comparison to ascertain whether the whole underlying decision-making processes might differ between moral and nonmoral situations.

We can observe that the model fit errors are very similar (see Table 1), which can be interpreted as a fraction of scale size (and error of 0.11 corresponding to 5.5% of scale or screen width). When focusing on parameter values, we notice that in nonmoral decisions, the convolution kernel amplitude seems to be higher (bigger A value) than for moral decisions, then giving more weight to input evidence. Also, for both models, a value is minimal (not 0

Model	Parameters	A	B	a	ε	τ	λ	$ampl_m$	$delay$	$RMSE$
Only moral	Fitted	4.11	3.08	0.05	0.05	3.64	0.55	-	-	0.11
Only nonmoral	Fitted	4.51	2.82	0.05	3.97	0.98	0.91	-	-	0.14
Amplitude model	A priori	0.9	0.4	0.2	2	0.9	0.2	2.5	-	0.16
Delay model	A priori	0.9	0.4	0.2	2	0.9	0.2	-	15	0.28

Table 1 Values of parameters and error for model fitted only on moral and nonmoral decisions, and for amplitude and delay models.

to prevent simulation issues). This limits the integration of close input evidence and favors a "selection behavior". Nonmoral decisions also seem to be much noisier (possibly reflecting higher variance in decisions). Finally, there is a faster integration of the information for nonmoral decisions, with a smaller τ value. However, the τ value does not alone determine the convergence speed of the decision, as it also depends on A and λ parameters and input distribution.

However, our attempt to fit the reference model on moral decisions did not yield realistic results, particularly in generating plausible trajectories and final answer distribution (with a substantial portion of the simulated trials corresponded to extreme final responses, i.e., $x = 1$ values). On the other hand, when applied to nonmoral decisions, the fitted model demonstrated coherent results. Meanwhile, we did not observe a significant difference in extremeness between moral and nonmoral inputs. This suggests that the reference model may fail in accurately accounting for moral decisions. Hence, the underlying assumption of this model, that morality affects the whole decision process rather than only moral components processing, can be questioned, although our simulation results do not invalidate this possibility. It therefore seemed relevant to introduce and test additional parameters, to account for differences in how moral and nonmoral information influenced the decision making process when reading the stimuli. Those additional features assume that the nonmoral condition is processed distinctly - as a nonmoral element, while the action, in contrast, is processed as a moral element when moral.

3.6.3 Additional feature for moral information processing

For this part, we used a priori values for all the reference model parameters. In fact, this time we do not assume different values for moral and nonmoral decisions, as the processing difference between them would lie at the element level (action or condition). Instead, we used values obtained from the model fit of a previous study, where participants had to make decisions in a similar experimental design (Gautheron et al, 2023). Those values are presented in the Table 1. We then chose the value of the additional parameter explaining difference in moral processing ($ampl_m$ for the amplitude model and $delay$ for the delay model) in a way to minimize $RMSE$, because we had no a priori information helping to estimate their values. For the sake of clarity, we reported only effects of interest for the two model versions, while detailed results replicating our full reporting on human data can be found in supplemental files. Particularly, we do not present here analyzes of final responses conditioned by the extremeness of the answer. Potential biases that must be tested and eliminated in empirical human studies can indeed be eliminated by design in computational models. Also, following the conflict modeling approach presented earlier, we did not find any significant difference in aggregated valence between nonmoral and moral items (as reported at the end of section 3.2), removing a possibly confound from the expected pattern of results.

Analysis on final responses

Version 1, amplitude model - Amplified moral perception. This version of the model was able to reproduce qualitatively empirical results with a $RMSE$ of 0.16 and a $ampl_m$ parameter equal to 2.5 (see Table 1 and Fig. 7). Notably, the different effects of interest observed in the empirical study were also found: (i) main effect of conflict, where answers are more nuanced in presence of conflict ($b = -0.35$, $SE = 0.006$, $\chi^2 = 2984.10$, $p < .001$, 95% CI $[-0.37, -0.34]$, $d = -1.54$); (ii) main effect of an increased strength of conflict ($b = 0.05$, $SE = 0.003$, $\chi^2 = 223.44$, $p < .001$, 95% CI $[0.04, 0.05]$, $d = 0.21$); (iii) interaction effect with increased difference between moral and nonmoral conditions in high-conflict case compared to low-conflict case ($b = 0.02$, $SE = 0.006$, $\chi^2 = 6.41$, $p = .01$, 95% CI $[0.004, 0.03]$, $d = 0.04$); (iv) no significant difference in no-conflict case between moral and nonmoral trials ($b = -0.005$, $SE = 0.04$, $t(91.99) = -0.12$, $p = .91$, 95% CI $[-0.09, 0.08]$); (v) finally, a bigger impact of conflict on nonmoral items, whereas moral decisions seem to be more stable even in conflictual situation ($b = 0.28$, $SE = 0.01$, $\chi^2 = 468.00$, $p < .001$, 95% CI $[0.26, 0.31]$, $d = 0.61$). Also, the main effect of morality was significant ($b = 0.14$, $SE = 0.05$, $\chi^2 = 8.54$, $p = .0035$, 95% CI $[0.04, 0.23]$, $d = 0.59$), as well as the interaction between conflict and its strength ($b = -0.29$, $SE = 0.008$, $\chi^2 = 1409.39$, $p < .001$, 95% CI $[-0.30, -0.27]$, $d = -0.62$). Then, moral decisions were more extreme than nonmoral decisions, while the presence of conflict led to more nuanced decisions—and even more nuanced decisions when the conflict was high rather than low.

Version 2, delay model - early moral integration. This version of the model could also reproduce empirical effects with a $RMSE$ of 0.28, and a $delay$ of 15 iterations (see Table 1 and Fig. 7), corresponding to 75ms. The error is higher than in the amplitude model, seemingly because decisions are less extreme. However, as we used a priori values, they may not minimize the error of this model version. Notably, the introduction of a delay could significantly alter decision processes, compared to the fitted model, which would consequently change the optimal parameter values. This model was similarly able to account for: (i) main effect of conflict, where answers are more nuanced in presence of conflict ($b = -0.28$, $SE = 0.004$, $\chi^2 = 4513.51$, $p < .001$, 95% CI $[-0.29, -0.27]$, $d = -1.64$); (ii) main effect of an increased strength of conflict ($b = 0.04$, $SE = 0.003$, $\chi^2 = 185.36$, $p < .001$, 95% CI $[0.04, 0.05]$, $d = 0.24$); (iii) interaction effect with increased difference between moral and nonmoral conditions in high-conflict case compared to low-conflict case ($b = 0.01$, $SE = 0.006$, $\chi^2 = 5.13$, $p = .02$, 95% CI $[0.002, 0.03]$, $d = 0.04$); (iv) no significant difference in no-conflict case between moral and nonmoral trials ($b = 0.05$, $SE = 0.03$, $t(92.01) = 1.66$, $p = .10$, 95% CI $[-0.009, 0.11]$); (v) finally, a bigger impact of conflict on nonmoral items, whereas moral decisions seem to be more stable even in conflictual situation ($b = 0.10$, $SE = 0.008$, $\chi^2 = 146.91$, $p < .001$, 95% CI $[0.08, 0.12]$, $d = 0.30$). Similarly to the amplitude model, the main effect of morality was also significant ($b = 0.10$, $SE = 0.03$, $\chi^2 = 11.85$, $p < .001$, 95% CI $[0.04, 0.16]$, $d = 0.58$), as well as the interaction

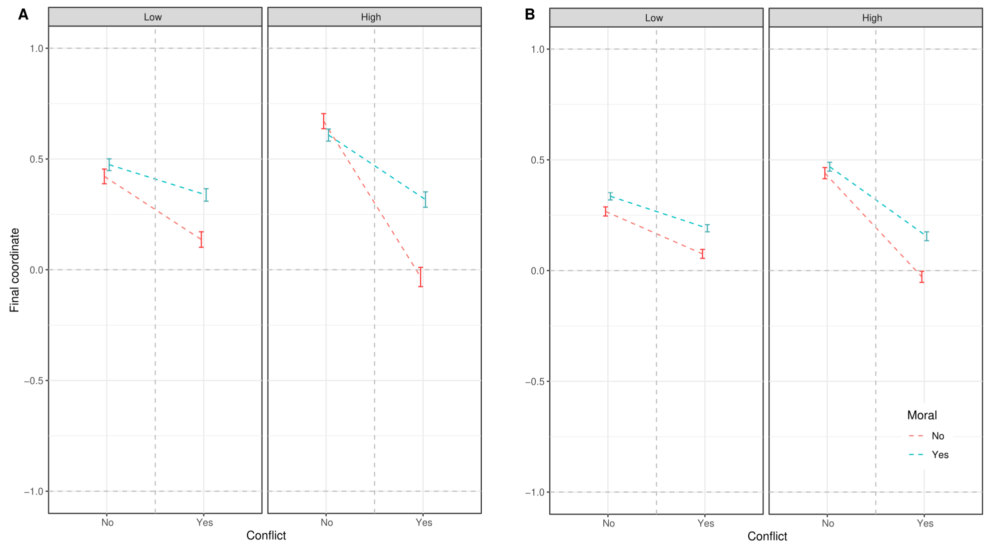

Fig. 7 Remapped final answers obtained with the model (A: amplitude model, B: delay model), for moral and non-moral conditions, and when conflict present or absent. In each subfigure, the left panels corresponds to the condition of low conflict (close attractors), and the right the condition of high conflict (distant attractors). Error bars correspond to *SE*.

between conflict and its strength ($b = -0.22$, $SE = 0.004$, $\chi^2 = 3035.61$, $p < .001$, 95% CI $[-0.23, -0.22]$, $d = -0.65$).

Response time

Given the time-continuous nature of differential equations 1 and 2, real time decision-making is directly embedded in the model. In the time-discretized computational implementation, response times are simply obtained by multiplying the time step parameter value with the number of iterations of the model, as it represents the time needed for the decision process to integrate stimulus information and converge to a final decision. It is worth noting that the models have been fitted only on final answers, which does not include the response time in the fit criteria, so that time scaling (e.g., through parameters τ and λ) obtained from earlier data is required to generate plausible response times. A matching response time pattern between empirical and simulated trajectories would then support the validity and forecasting capability of the model.

Version 1, amplitude model - Amplified moral perception. Response times of this model version showed similarities with those obtained during the experiment. In particular, we observe, similarly to empirical data, shorter decisions for non-conflictual condition ($b = 0.26$, $SE = 0.01$, $\chi^2 = 572.65$, $p < .001$, 95% CI $[0.24, 0.28]$, $d = 1.74$), and for low-conflict condition ($b = 0.12$, $SE = 0.01$, $\chi^2 = 153.86$, $p < .001$, 95% CI $[0.10, 0.14]$, $d = 0.80$). We also found interaction effects between the presence of conflict and its strength (high or low) ($b = -0.10$, $SE = 0.02$, $\chi^2 = 26.56$, $p < .001$, 95% CI $[-0.14, -0.06]$, $d = -0.33$), between the presence of conflict and morality ($b = -0.20$, $SE = 0.02$, $\chi^2 = 85.74$,

$p < .001$, 95% CI $[-0.25, -0.16]$, $d = -0.67$), and between the strength of the conflict and morality ($b = -0.10$, $SE = 0.02$, $\chi^2 = 26.56$, $p < .001$, 95% CI $[-0.14, -0.06]$, $d = -0.33$). Moreover, unlike in empirical data, we observed a significant effect of morality, with shorter moral decisions ($b = -0.61$, $SE = 0.03$, $\chi^2 = 595.10$, $p < .001$, 95% CI $[-0.66, -0.56]$, $d = -4.04$).

Version 2, delay model - early moral integration. We found the same pattern as in the amplitude model. Precisely, shorter decisions for non-conflictual condition ($b = 0.18$, $SE = 0.008$, $\chi^2 = 559.23$, $p < .001$, 95% CI $[0.17, 0.20]$, $d = 1.31$), and for low-conflict condition ($b = 0.14$, $SE = 0.005$, $\chi^2 = 798.95$, $p < .001$, 95% CI $[0.13, 0.15]$, $d = 1.03$). We also found interaction effects between the strength of the conflict and morality ($b = -0.05$, $SE = 0.01$, $\chi^2 = 26.04$, $p < .001$, 95% CI $[-0.07, -0.03]$, $d = -0.19$). However, there was no significant interaction between the presence of conflict and its strength (high or low) ($b = -0.004$, $SE = 0.01$, $\chi^2 = 0.14$, $p = .71$, 95% CI $[-0.03, 0.02]$, $d = -0.01$), and between the presence of conflict and morality ($b = 0.005$, $SE = 0.02$, $\chi^2 = 0.10$, $p = .76$, 95% CI $[-0.03, 0.03]$, $d = 0.02$). Moreover, unlike in empirical data, we observed a very significant effect of the morality, with way shorter moral decisions ($b = -0.18$, $SE = 0.02$, $\chi^2 = 60.55$, $p < .001$, 95% CI $[-0.22, -0.13]$, $d = -1.29$).

Mouse trajectories

As the model was designed to generate mouse trajectories, we were able to compare trajectories generated for each condition, using the same time bin analyzes as in the empirical data analyzes. Again, dynamic parameters of mouse trajectories that are of interest here were not included in the fitting criteria. Therefore, results presented below derive from model predictions informed by theory and a priori input distributions.

Version 1, amplitude model - Amplified moral perception. Results can be seen on Fig. 8. First, nonmoral trajectories were significantly deviated from the beginning to the end of the trajectory ($b < 0.30$, $SE < 0.036$, $t < 8.21$, $p < .001$), as well as moral trajectories ($b < 0.43$, $SE < 0.029$, $t > 14.82$, $p < .001$). In particular, nonmoral highly conflictual trajectories significantly deviate from the middle only from 10% to 45% of the trajectories ($b > -0.032$, $SE < 0.015$, $t < -2.14$, $p < .04$). Then, in all conditions, except for high conflict nonmoral condition, trajectories deviate from the very beginning until the end. The difference between moral and nonmoral trials was significant since the first time bin to the end of the trajectories ($b < 0.13$, $SE < 0.05$, $\chi^2 < 8.24$, $p < .0041$, $d < 0.58$), as well as the presence of conflict ($b > -0.35$, $SE < 0.006$, $\chi^2 < 3066.07$, $p < .001$, $d > -1.53$), the strength of the conflict ($b < 0.05$, $SE < 0.003$, $\chi^2 < 225.69$, $p < .001$, $d < 0.20$), and the interaction between morality and conflict ($b < 0.28$, $SE < 0.01$, $\chi^2 < 471.35$, $p < .001$, $d < 0.60$). Hence, the main effects of morality, presence of conflict, and conflict strength, as well as the interaction effect between morality and conflict were also significant from the beginning of the trajectories.

Version 2, delay model - early moral integration. The simulated trajectories for this version of the model are presented in Fig. 8. This time, due to delayed integration of nonmoral factors, nonmoral trajectories were deviated from the center only from 40% to the end of the trajectory ($b < 0.19$, $SE < 0.023$, $t(89.68) < 8.25$, $p < .001$), compared to moral trajectories that are significantly deviated from the beginning ($b < 0.29$, $SE < 0.018$, $t(89.68) > 15.78$,

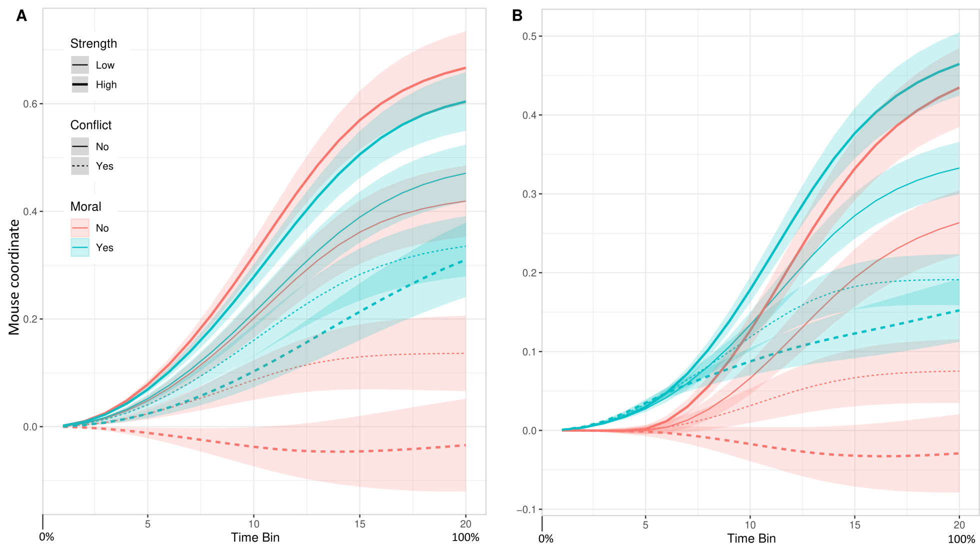

Fig. 8 Mean simulated trajectories for each condition (A: amplitude model, B: delay model). Error bands correspond to *95% CI*.

$p < .001$). Again, the nonmoral highly conflictual trajectories never come to significantly deviate from the middle. The difference between moral and nonmoral trials was significant since the first time bin to the end of the trajectories ($b < 0.10$, $SE < 0.03$, $\chi^2 < 89.02$, $p < .001$, $d < 1.00$), as well as the presence of conflict ($b > -0.28$, $SE < 0.004$, $\chi^2 < 4543.53$, $p < .001$, $d > -1.62$), and the interaction between morality and conflict ($b < 0.10$, $SE < 0.008$, $\chi^2 < 145.81$, $p < .001$, $d < 0.29$). The strength of the conflict becomes significant from 20% of the trajectories ($b < 0.04$, $SE < 0.003$, $\chi^2 < 181.17$, $p < .001$, $d < 0.23$). Again, the main effects of morality, presence of conflict, and the interaction effect between morality and conflict were significant from the beginning of the trajectories, while the effect of conflict strength appears later.

3.7 Discussion

Our attempt to fit the reference model on moral decisions did not yield realistic results, particularly in generating plausible trajectories. On the other hand, when applied to nonmoral decisions, the fitted model demonstrated coherent results. Meanwhile, we did not observe a significant difference in extremeness between moral and nonmoral inputs. This suggests that the reference model may fail in accurately accounting for moral decisions. Hence, the underlying assumption of this model, that morality affects the whole decision process rather than only moral components processing, can be questioned, although our simulation results do not invalidate this possibility. It therefore seemed relevant to introduce and test additional parameters, to account for differences in how moral and nonmoral information influenced the decision making process when reading the stimuli.

The two versions of the model each introducing a single additional parameter were able to reproduce empirical effects in the final responses analysis. We observed a substantial difference of goodness of fit between the two versions, the amplitude model being closer to the empirical data. However, as we used a priori parameters (not directly fitted on each model version) it just might have favored the amplitude model. To confirm this, we manually tuned the parameters to minimize the error of the delay model and obtained an error similar to the amplitude model's one. In the amplitude model, we were able to explain differences in the effect of conflict between moral and nonmoral decisions only by changing the weight of the evidence taken into consideration (amplitude of the bumps in input). This means that the difference of treatment of the information would just give more weight to information when it recognizes it as moral; *or* moral information has intrinsically more weight when processed. This could explain the perceptual saliency of moral stimuli (Gantman and Van Bavel, 2014).

On the other hand, we were also able to reproduce empirical final answer results with the delay model involving a delayed integration of nonmoral factors compared to moral factors. This could indicate that more stable moral decisions could also be caused by an earlier consideration of moral elements. This earlier perceptual consideration of moral elements could result from the moral pop-out effect (Gantman and Van Bavel, 2014), suggesting a facilitated perceptual awareness of morally-laden stimuli that could lead to their earlier processing. The delay minimizing the difference between empirical and modeled data (i.e., nonmoral evidence being processed 75ms after moral evidence) can be informative of the actual time difference between moral and nonmoral processing, but should be considered with caution. Indeed, we used fixed a priori values for other parameters based on earlier studies to test the predictive value of the model instead of fine-tuning parameters to fit empirical data. Also, we were able to find similar patterns of results for response times, except for the effect of morality. Indeed, the model presented significantly shorter response times for moral trials, whereas this effect was not significant in empirical data.

The major difference between the two versions of the model really appears when we look at simulated trajectories. Not surprisingly, in the delay model, we notice through visual inspection, confirmed by analytical results, that nonmoral decision trajectories show a delayed deviation from the center. This deviation is present in empirical data, albeit less pronounced. On the other side, this difference in the timing of deviations does not appear in the amplitude model. Hence, these results would rather favor the delay model, or a model integrating features of both versions (i.e., an earlier *and* an enhanced impact of moral factor in the decision process). Yet, a model combining both versions was not tested here, as it would also yield higher complexity through additional parameters, thus possibly leading to overfit of the general data pattern. Also, we can see that, globally, deviations from the center appear earlier than in the empirical mouse trajectory. Indeed, in our implementation of the model, the different factors are influencing the decision process since the beginning, or almost (all conditions are significantly different from 0 since the first time bin in the amplitude model, except for the nonmoral highly conflictual condition). In empirical data, participants first need to integrate the information, by reading the sentence, understanding it, etc., exploiting it progressively to make their decision. For a more realistic model, it is possible to set the influence of the different factors after a certain iteration step. However, we are not interested in this part of the decision process, which is mainly driven by decision noise, and since it would require introducing additional parameters.

Globally, our model specifications were quite straightforward, as we resumed evidence taken into consideration in the decision making process as a mere instant summation of one or more words' valence. However, the sentence processing during decision making is undeniably more complex. More developed models could take it into account by integrating a natural language processing module. The simplicity of our model is nevertheless a strength, as a simple mechanism with simple inputs could reproduce and explain a substantial part of empirical results. Furthermore, the a priori specification of the model succeeded in reproducing empirical effects. Moreover, it did not only account for final decisions, but also for their dynamics by reproducing response time and mouse trajectories patterns.

4 General Discussion

Findings of the empirical study suggest that, when conflict arises in everyday decisions, individuals tend to find compromise, as higher level of conflict resulted in less extreme decisions, for moral and nonmoral situations. Moreover, when facing a moral situation, participants made more stable decisions, less affected by the presence of conflict than nonmoral decisions. As such an interaction pattern could be explained by differences in samples of nonmoral and moral actions, we alleviated confounds and refined our results by conducting conditional analyses. We were therefore able to confirm our conclusions by estimating the effect of conflict at the item-by-participant level, comparing the extremeness of responses given in presence of high vs. low conflict. These effects on empirical data were reproduced through two versions of our computational model, each implementing a different processing of moral elements compared to nonmoral ones. The first version, the amplitude model, implemented a greater amplitude for moral factors. The second version, the delay model, implemented an earlier impact of moral factors compared to nonmoral factors. These models were set a priori, reinforcing their general applicability and bolstering the validity of their predictions. Importantly, they were devised separately from the experiment itself. The models therefore serve a double purpose: they help to test the validity of assumptions regarding the underlying decision-making process and to illustrate how such processes might impact the dynamics and final outcomes of decisions. Specifically, we examined how conflict can influence decisions, as manifested by multiple evidence supporting different options, leading to the selection of one option over others or a compromise between them. The models revealed how various pieces of information aggregate to form a compromise. By adding distinct processing for moral and nonmoral factors, we were able to examine and provide possible explanations for the greater stability in moral decisions observed in empirical results, especially when conflict arises.

Response times also showed interesting patterns, in line with the literature. In particular, we found that conflicted decisions resulted in longer response times in the experiment, as well as in both model versions. Observing this result in the models supports their validity, as only the final decisions and not the mouse trajectories nor decision dynamics were used in the fit criteria. Notably, the implementation of decision conflict in both models successfully reproduced associated empirical dynamics, with decisions taking more time to integrate evidence voting for distant (more conflicting) responses in the decision space. The formal model provides a mechanistic account of conflict and response time differences, as such situations lead to stronger competition between distant inputs, through mutual inhibition rather than excitation between response options. Additionally, despite the experiment revealing a marginal, non-significant effect of morality on response times, with nonmoral decisions taking descriptively more time, this effect was substantial and significant in both versions of the model. It is noteworthy that this difference can be explained by the lack of simulated inter-participant variability, facilitating the generalizability compared to the empirical data when using mixed effect models with both participants and items random factors. Indeed, even simulating the same number of observations (number of conditions, participants and trials), the inferential tests applied to the simulated data yield higher statistical power due to reduced variance. Hence, the potential mechanistic explanations offered by the model warrant further examination in subsequent empirical studies, with experimental designs calibrated to test the model predictions with sufficient statistical power. Finally, when focusing on the decision dynamics, we observed empirically earlier deviations for moral decisions, and for conflictual decisions. The earlier influence of morality was only reproduced in the delay model, not in the amplitude model.

Taking into account all these results, we would rather favor the delay model, with a processing delay between moral and nonmoral factors, as simulated trajectories were more in accordance with empirical results. However, it is not contradictory with the existence of both decision mechanisms leading to a difference between moral and nonmoral decisions. We also suspect that one mechanism over the other could have more effect depending on the situation. For example, in the moral decision making literature, the dual-process model supposes that moral rules are processed earlier than moral considerations for consequences or implications (Greene et al, 2008; Bartels, 2008). Hence, the earlier intervention of moral factors in the decision may occur especially when moral intensity is high (e.g., moral dilemmas). Reciprocally, for situations where the moral dimension is salient, the moral factor may have a bigger weight in the decision process. Testing these different assumptions would be a relevant avenue for future research.

Despite its mixed-method contributions, the present research is not without limitations. A first technical and empirical limitation lies at the mouse-tracking paradigm level, as we collected a non negligible proportion of mouse trajectories exhibiting a problematic oscillatory pattern. This pattern appeared as a strategy to lengthen the trial duration, as the mouse cursor shrank over time. While we already discussed that the removal of these problematic trajectories did not globally impact statistical validity, it is still a feature of the paradigm that should be improved to avoid this situation in subsequent studies, by adapting the shrinking pace of the slider, for instance.

At the more theoretical level, one could argue that the conflict experienced in moral and nonmoral situations is not equivalent, as the same condition (same amount of money) was used in both cases. Indeed, the moral action may have greater implications than the nonmoral action and might require a more significant incentive to produce the same level of conflict (e.g., harming someone as a consequence, thus involving a moral aspect). This is consistent with the literature suggesting that decisions are more difficult when conflicting attributes belong to the same rather than different categories (Hanselmann and Tanner, 2008). Thus, when a moral attribute conflicts with a nonmoral attribute, the decision should be easier than when it conflicts with another moral attribute (i.e., moral dilemma case). This may be an inherent aspect of moral decisions that is impossible to control experimentally. For example, introducing a moral condition (involving harm or death) for moral items would have created vastly different conditions between moral and nonmoral trials, complicating comparison while not definitively addressing the issue. However, this could explain why we were not able to fit the reference model only on moral decisions, without specifying a specific processing of moral factors. This suggests that moral actions would be processed differently from the non-moral actions and conditions used in the experiment. The computational study contributes to answer to the query arising from our experimental design, whether the presence of a moral situation compels individuals to interpret all elements from a moral perspective, subsequently processing them all as moral elements, or whether each elements are processed distinctly as moral or nonmoral. Yet given that results from computational models directly depend on their specification and parameter values, additional studies investigating this specific question would be needed to further support this conclusion. In a similar way, the choice to consider distinct moral and nonmoral elements was theoretically driven, and to guarantee parsimony in both the computational and preregistered empirical designs as well as analyses. However, there might be a continuum between moral and nonmoral decisions, for instance relying on the degree of 'moral intensity' (May and Pauli, 2002), which would be translated by varying amplitudes and delays in the proposed computational models. Yet, estimating parameters at the condition and action level for each stimulus would greatly complexify the model without bringing any new notion or concept. Even if neglecting between-participant variability in moral sensitivity (i.e. consideration of the morality of a situation, see Sparks, 2015), quantitative estimates of moral saliency of individual words, conditions or actions would involve complex natural language processes, and constitutes an avenue left open for further research.

5 Conclusion

In sum, conflict, as operationalized in the present research, tends to impact decision making, encouraging compromise between the two conflicting elements for both moral and nonmoral decisions. Conflict also lengthens decision times, and leads to earlier deviations in mouse trajectories. Additionally, moral decisions happen to be more more stable, presumably because they appeal to moral standards that are less flexible than nonmoral decisions.

The computational model associated with the empirical study offers a possible explanation of how conflict intervenes in moral and nonmoral decision processes. First, conflict can indeed be understood as a competition between two (or more) evidence taken into consideration during the decision. Second, the difference between moral and nonmoral decisions may be explained either by 1) moral factors having a more important weight on the decision process; or 2) an earlier impact of moral processing. These two a priori versions of the model were able to reproduce almost every effect observed in the empirical study. However, only the second version accounted for empirically observed delayed deviations of nonmoral trajectories, compared to moral trajectories. We propose that the two described mechanisms could both intervene more or less depending on the situation. The model thus serves as an interesting basis to make predictions about the implication of conflict in moral and nonmoral decision making. More importantly, these two model versions are compatible with specific processing of moral elements during decision making rather than with a whole different underlying decision process in moral situations.

Statements and Declarations

Financial disclosure

This work was supported by the Pôle Grenoble Cognition (FR 3381 CNRS, Univ. Grenoble Alpes, Grenoble INP), and the French National Research Agency in the framework of the "Investissements d'avenir" programs ANR-15-IDEX-02 and ANR-11-LABX-0025-01.

Conflict of interest

The authors declare no potential conflict of interests.

Data availability statement

Data that we are allowed to share that support the findings of the studies reported in this paper are openly available on OSF at https://osf.io/6juex.

Supporting information

Additional supporting information may be found in the online version of the article at the publisher's website, and on OSF.

Ethical statement

The present research was conducted in accordance with the Declaration of Helsinki and was approved by the local ethics committee (CER-USMB). Participants were informed about its aim and confidentiality of the data collection, and gave their consent to participate. Participants could withdraw at any time during the study.

References

Anderson CJ (2003) The psychology of doing nothing: forms of decision avoidance result from reason and emotion. Psychological bulletin 129(1):139

Barge S, Gehlbach H (2012) Using the theory of satisficing to evaluate the quality of survey data. Research in Higher Education 53(2):182–200

Bartels DM (2008) Principled moral sentiment and the flexibility of moral judgment and decision making. Cognition 108(2):381–417

Bartels DM, Bauman CW, Cushman FA, et al (2015) Moral judgment and decision making. The Wiley Blackwell handbook of judgment and decision making 63:478–515

Bates D, Mächler M, Bolker B, et al (2015) Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67(1):1–48

Bauman CW, McGraw AP, Bartels DM, et al (2014) Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. Social and Personality Psychology Compass 8(9):536–554

Brady WJ, Gantman AP, Van Bavel JJ (2020) Attentional capture helps explain why moral and emotional content go viral. Journal of Experimental Psychology: General 149(4):746

Broeders R, Van den Bos K, Müller PA, et al (2011) Should I save or should I not kill? How people solve moral dilemmas depends on which rule is most accessible. Journal of Experimental Social Psychology 47(5):923–934

Buttlar B, Walther E (2018) Measuring the meat paradox: How ambivalence towards meat influences moral disengagement. Appetite 128:152–158

Cheng J, González-Vallejo C (2018) Unpacking decision difficulty: Testing action dynamics in intertemporal, gamble, and consumer choices. Acta psychologica 190:199–216

Crockett MJ (2013) Models of morality. Trends in cognitive sciences 17(8):363–366

Crockett MJ (2016) How formal models can illuminate mechanisms of moral judgment and decision making. Current directions in psychological science 25(2):85–90

Dehghani M, Tomai E, Forbus KD, et al (2008) An integrated reasoning approach to moral decision-making. In: AAAI, pp 1280–1286

Ditterich J (2006) Stochastic models of decisions about motion direction: behavior and physiology. Neural networks 19(8):981–1012

Dotan D, Pinheiro-Chagas P, Al Roumi F, et al (2019) Track it to crack it: Dissecting processing stages with finger tracking. Trends in Cognitive Sciences 23(12):1058–1070

Drugowitsch J, Moreno-Bote R, Churchland AK, et al (2012) The cost of accumulating evidence in perceptual decision making. Journal of Neuroscience 32(11):3612–3628

Erlhagen W, Schöner G (2002) Dynamic field theory of movement preparation. Psychological review 109(3):545

Evans AM, Dillon KD, Rand DG (2015) Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. Journal of Experimental Psychology: General 144(5):951

Falandays J, Spivey MJ (2020) Biasing Moral Decisions Using Eye Movements: Replication and Simulation. In: 42nd Annual Virtual Meeting of the Cognitive Science Society (CogSci 2020), pp 2553–2559

Falandays JB, Spevack S, Pärnamets P, et al (2021) Decision-making in the human-machine interface. Frontiers in Psychology 12:99

Ford JK, Schmitt N, Schechtman SL, et al (1989) Process tracing methods: Contributions, problems, and neglected research questions. Organizational behavior and human decision processes 43(1):75–117

Freeman J, Dale R, Farmer T (2011) Hand in motion reveals mind in motion. Frontiers in Psychology 2:59

Freeman JB, Ambady N (2011) A dynamic interactive theory of person construal. Psychological review 118(2):247

Gaboriaud A, Gautheron F, Quinton JC, et al (2022) The effects of intent, outcome, and causality on moral judgments and decision processes. Psychologica Belgica 62(1):218

Gantman A, Van Bavel J (2015) Behavior is multiply determined and perception has multiple components: The case of moral perception. Available at SSRN 2695248

Gantman AP, Van Bavel JJ (2014) The moral pop-out effect: Enhanced perceptual awareness of morally relevant stimuli. Cognition 132(1):22–29

Gautheron F, Quinton JC, Muller D, et al (2023) Paradigm constraints on moral decision-making dynamics. Journal of Behavioral Decision Making 36(4):e2324

Greene JD, Sommerville RB, Nystrom LE, et al (2001) An fMRI investigation of emotional engagement in moral judgment. Science 293(5537):2105–2108

Greene JD, Morelli SA, Lowenberg K, et al (2008) Cognitive load selectively interferes with utilitarian moral judgment. Cognition 107(3):1144–1154

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

Haidt J (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychological review 108(4):814

Hanselmann M, Tanner C (2008) Taboos and conflicts in decision making: Sacred values, decision difficulty, and emotions. Judgment and Decision making 3(1):51–63

Hehman E, Stolier RM, Freeman JB (2015) Advanced mouse-tracking analytic techniques for enhancing psychological science. Group Processes & Intergroup Relations 18(3):384–401

Hopp FR, Fisher JT, Cornell D, et al (2021) The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. Behavior Research Methods 53(1):232–246

Iliev R, Sachdeva S, Bartels DM, et al (2009) Attending to moral values. Psychology of learning and motivation 50:169–192

Johnson DJ, Hopwood CJ, Cesario J, et al (2017) Advancing research on cognitive processes in social and personality psychology: A hierarchical drift diffusion model primer. Social Psychological and Personality Science 8(4):413–423

Kahane G (2015) Sidetracked by trolleys: Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. Social neuroscience 10(5):551–560

Klauer KC (2014) Random-walk and diffusion models. Dual process theories of the social mind pp 139–152

Kleiman T, Hassin RR (2011) Non-conscious goal conflicts. Journal of experimental social psychology 47(3):521–532

Koop GJ (2013) An assessment of the temporal dynamics of moral decisions. Judgment and decision making 8(5):527

Krajbich I, Rangel A (2011) Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. Proceedings of the National Academy of Sciences 108(33):13852–13857

Krajbich I, Oud B, Fehr E (2014) Benefits of neuroeconomic modeling: New policy interventions and predictors of preference. American Economic Review 104(5):501–06

Krajbich I, Hare T, Bartling B, et al (2015) A common mechanism underlying food choice and social decisions. PLoS Comput Biol 11(10):e1004371

Krosnick JA (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied cognitive psychology 5(3):213–236

Krypotos AM, Beckers T, Kindt M, et al (2015) A Bayesian hierarchical diffusion model decomposition of performance in Approach–Avoidance Tasks. Cognition and Emotion 29(8):1424–1444

Lakens D (2014) Performing high-powered studies efficiently with sequential analyses. European Journal of Social Psychology 44(7):701–710

Leong W, Hensher DA (2012) Embedding decision heuristics in discrete choice models: A review. Transport Reviews 32(3):313–331

Lepora NF, Pezzulo G (2015) Embodied choice: how action influences perceptual decision making. PLoS computational biology 11(4):e1004110

Levine S, Kleiman-Weiner M, Schulz L, et al (2020) The logic of universalization guides moral judgment. Proceedings of the National Academy of Sciences 117(42):26158–26169

Liu BS, Ditto PH (2013) What dilemma? Moral evaluation shapes factual belief. Social Psychological and Personality Science 4(3):316–323

Luce MF, Payne JW, Bettman JR (1999) Emotional trade-off difficulty and choice. Journal of marketing research 36(2):143–159

May DR, Pauli KP (2002) The role of moral intensity in ethical decision making: A review and investigation of moral recognition, evaluation, and intention. Business & Society 41(1):84–117

Metin B, Roeyers H, Wiersema JR, et al (2013) ADHD performance reflects inefficient but not impulsive information processing: A diffusion model analysis. Neuropsychology 27(2):193

Moll J, de Oliveira-Souza R, Bramati IE, et al (2002) Functional networks in emotional moral and nonmoral social judgments. Neuroimage 16(3):696–703

Monin B, Pizarro DA, Beer JS (2007) Reason and emotion in moral judgment: Different prototypes lead to different theories. Do emotions help or hurt decision making? A hedgefoxian perspective pp 219–244

Paquette L, Kida T (1988) The effect of decision strategy and task complexity on decision performance. Organizational behavior and human decision processes 41(1):128–142

Pärnamets P, Johansson P, Hall L, et al (2015) Biasing moral decisions by exploiting the dynamics of eye gaze. Proceedings of the National Academy of Sciences 112(13):4170–4175

Quinton JC, Girau B (2011) Predictive neural fields for improved tracking and attentional properties. In: The 2011 International Joint Conference on Neural Networks, IEEE, pp 1629–1636

Quinton JC, Volpi NC, Barca L, et al (2013) The cat is on the mat. Or is it a dog? Dynamic competition in perceptual decision making. IEEE Transactions on Systems, Man, and Cybernetics: Systems 44(5):539–551

Ratcliff R, McKoon G (2008) The diffusion decision model: theory and data for two-choice decision tasks. Neural computation 20(4):873–922

Ratcliff R, Smith PL (2004) A comparison of sequential sampling models for two-choice reaction time. Psychological review 111(2):333

Ratcliff R, Tuerlinckx F (2002) Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. Psychonomic bulletin & review 9(3):438–481

Ratneshwar S, Mick DG, Huffman C (2003) The why of consumption: Contemporary perspectives on consumer motives, goals, and desires, vol 1. Psychology Press

Rougier NP, Vitay J (2006) Emergence of attention within a neural population. Neural Networks 19(5):573–581

Schöner G, Spencer J (2016) Dynamic thinking: A primer on dynamic field theory. Oxford University Press

Skitka LJ, Morgan GS (2014) The social and political implications of moral conviction. Political psychology 35:95–110

Smeding A, Quinton JC, Lauer K, et al (2016) Tracking and simulating dynamics of implicit stereotypes: A situated social cognition perspective. Journal of personality and social psychology 111(6):817

Sommer M, Rothmayr C, Döhnel K, et al (2010) How should I decide? The neural correlates of everyday moral reasoning. Neuropsychologia 48(7):2018–2026

Sparks JR (2015) A social cognitive explanation of situational and individual effects on moral sensitivity. Journal of Applied Social Psychology 45(1):45–54

Spivey MJ, Dale R (2004) On the continuity of mind: Toward a dynamical account of cognition. The psychology of learning and motivation: Advances in research and theory 45:87–142

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

Stadthagen-Gonzalez H, Imbault C, Sánchez MAP, et al (2017) Norms of valence and arousal for 14,031 Spanish words. Behavior research methods 49(1):111–123

Sullivan N, Hutcherson C, Harris A, et al (2015) Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. Psychological science 26(2):122–134

Sullivan NJ, Huettel SA (2018) Dietary self-control depends on the latency and rate of information accumulation. bioRxiv p 465393

Tassy S, Oullier O, Mancini J, et al (2013) Discrepancies between judgment and choice of action in moral dilemmas. Frontiers in psychology 4:250

Trémolière B, Bonnefon JF (2014) Efficient kill–save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. Personality and Social Psychology Bulletin 40(7):923–930

Tyebjee TT (1979) Response time, conflict, and involvement in brand choice. Journal of Consumer Research 6(3):295–304

Van Bavel JJ, Packer DJ, Haas IJ, et al (2012) The importance of moral construal: Moral versus non-moral construal elicits faster, more extreme, universal evaluations of the same actions. PloS one 7(11):e48693

Westfall J (2015) PANGEA: Power analysis for general ANOVA designs. Unpublished manuscript, available at https://osfio/x5dc3/download