Biasing Moral Decisions Using Eye Movements: Replication and Simulation

J. Benjamin Falandays (jfalandays@ucmerced.edu)
Cognitive and Information Sciences Department, University of California, Merced,
5200 Lake Rd.
Merced, CA 95343 USA

Michael J. Spivey (spivey@ucmerced.edu)
Cognitive and Information Sciences Department, University of California, Merced,
5200 Lake Rd.
Merced, CA 95343 USA

Abstract

A current debate concerns the degree to which moral reasoning is susceptible to bias from low-level perceptual cues. Pärnamets et al. (2015) reported that moral decisions could be biased by manipulating the timing of a prompt to respond via measurement of eye gaze, but these results were critiqued by Newell and Le Pelley (2018) as a potential design artifact. To reconcile these findings, we first replicate the previous experiments with an adjusted stimulus set. Then, we present the results of a drift-diffusion model that simulates our findings, offering an account of the mechanism by which the gaze-based timing manipulation can bias moral decision-making.

Keywords: morality; decision-making; dynamical systems; eye tracking

Introduction[1]

Moral decision making , the process of deciding whether an action is morally acceptable or unacceptable, can be a matter of grave consequence. Perhaps unfortunately, humans are not purely rational agents (Gigerenzer & Selten, 2002; Tversky & Kahneman, 1989), and even "high-level" cognition such as moral decision making has been shown to be sensitive to situation-irrelevant factors such as priming (Gu, Zhong, & Page-Gould, 2013) and framing effects (Petrinovich & O'Neill, 1996), perhaps through influences from emotional processing (Prinz, 2007). However, there is reason to suspect that emotions are not a privileged source of influence on moral cognition. Complex dynamical systems accounts of cognition (Spivey, 2008) hold that cognition in general is highly parallel and interactive, with many bidirectional linkages among subsystems involved with language, perception, action, and emotion (Falandays, Batzloff, Spevack, & Spivey, 2018). If this interactivity extends to moral decision making, there are likely many routes toward influencing moral decisions. In the present study, we explore the hypothesis that moral decision making can be biased through manipulation of the timing of a person's sensorimotor interaction with the environment.

Previous work based on the complex dynamical framework has investigated the relationship between moral reasoning and eye movements (Pärnamets et al., 2015; Newell & Le Pelley, 2018). Drift diffusion models (Krajbich, Armel, & Rangel, 2010) view decision making as the accumulation of preference via visual sampling of response alternatives, making gaze both an index of preferences and an influence upon them (Parnamets, Richardson, & Balkenius, 2014; Shimojo, Simion, Shimojo, & Scheier, 2003). In a test inspired by such models, Pärnamets et al. (2015) found that moral decisions could be biased by a gaze-based manipulation of response timing. Participants heard a moral statement over headphones, such as "Murder is sometimes justifiable." Then, two response options (e.g. "Sometimes justifiable" or "Never justifiable") appeared on the left and right sides of the screen, respectively, while an eye tracker recorded the duration of the participants' gaze to both response options. On each trial, the software randomly and secretly preselected one of the response options to be the "target." Participants were not prompted to make a choice until at least 750ms of gaze was allocated to the target, and at least 250ms of gaze to the non-target – or after a maximum of 3s if those criteria were not met. Across trials on which the gaze thresholds were successfully met, participants chose the randomly pre-selected target significantly more than chance. The authors concluded that, by interrupting a participant's deliberation when more gaze had accumulated on the software's target option, preferences were systematically biased toward that option. That is, once the gaze-based timing manipulation was engaged – because the participant had fixated the software's pre-chosen target option for at least 750ms and the non-target option for at least 250ms – the participant's cognitive deliberation was interrupted at a point when they had, on average, been spending the majority of their time considering the response option that happened to be the software's secretly pre-chosen target. If the participant's deliberation had not been prematurely interrupted by the software, it is entirely possible that it could have shifted back to the option that was *not* the software's pre-chosen target. However, when pressed to "respond now," they tended to choose the response option that was most prominent at that time in their deliberation.

In an adjusted replication, Newell and Le Pelley (2018) found a biasing effect for certain perceptual judgments, but not for these moral judgments. The authors proposed that the findings of Pärnamets et al. (2015) were a methodological artifact of excluding time-out trials on which the gaze duration thresholds were not met before the 3s time limit. Trials that timed-out may, at least in some cases, have been due to the fact that participants had a strong pre-existing preference

[1] This project was pre-registered on the Open Science Framework https://osf.io/ef6tw

for the non-target option, and therefore only briefly fixated the target (for <750ms). Consistent with this hypothesis, the authors found that participants were significantly more likely than chance to select the alternative (non-target) option on time-out trials. As such, excluding these trials may have biased the dataset in favor of trials on which the participants had a pre-existing preference for the target option. If this were the case, Pärnamets et al's (2015) finding that participants selected the target significantly more often than chance may not be attributable to the gaze-based manipulation, but rather to prior experience and opinions. Newell and Le Pelley (2018) found that the effect disappeared for moral judgments when time-out trials were included in the analysis. The authors concluded that subtle manipulations of gaze are insufficient to penetrate high-level cognition such as moral reasoning.

Interestingly, Pärnamets et al. (2015) reported that their effect was statistically robust even with the inclusion of time-out trials. As such, it is unclear why the effect held in the original study but not in the replication. However, one possibility is that Newell and Le Pelley (2018) used the same stimuli as Pärnamets et al. (2015) in a culturally different population, for which the stimuli were not normed to generate uncertainty. The new sample of participants may have had stronger pre-existing preferences for one of the two options in some items, which perhaps were too strong to be overcome by the subtle timing manipulation. This prediction follows straightforwardly from drift-diffusion models of choice: when the relative value of one option is much greater than the other, decisions will quickly evolve towards one side, even if gaze bias slightly mitigates this process.

To help reconcile these conflicting results, we conducted a replication of Pärnamets et al. (2015, experiment 2) and Newell and Le Pelley (2018, experiment 2), with stimuli renormed to optimize uncertainty for each moral question with our present population. We predicted finding an effect of the gaze-based timing manipulation of the prompt to respond for these normed stimuli, but not for a set of "filler" stimuli that were not normed to generate uncertainty. Then, we present a drift-diffusion model, with minimal assumptions, that is able to simulate the general pattern of results across all three studies (the present study, Pärnamets et al., 2015, and Newell & Le Pelley, 2018).

Experiment

Method

Participants. 56 healthy undergraduate students (39 female, 17 male; age: mean$\pm$s.d. = 19.8$\pm$1.86) were recruited from the subject pool of a university in the Western United States. Participants provided informed consent in accordance with IRB protocols and received course credit for their participation. Participation was restricted to those who reported having normal or corrected-to-normal vision and hearing.

Stimulus Selection. We began with the 98 stimuli used in Pärnamets et al. (2015) and Newell and Le Pelley (2018), of which 63 were moral statements and 35 were fillers. 37 new moral stimuli were created to reach a starting pool of 100. These 100 stimuli were then turned into 2 separate lists by making changes to wording of either the prompt or response options with the goal of maximizing potential uncertainty over the preferred answer. 40 participants (from the same population as in the experiment) responded to each list in an online pilot survey, implemented with Qualtrics. From this survey data, we selected any prompts that generated 50%$\pm$10% agreement, with the restriction that only one variant of an original prompt was included in the final list. This resulted in 36 moral/ethical prompts. One filler prompt was added to have 36 of each type. No norming was conducted on the filler stimuli. The full list of stimuli is available on our preregistration page on OSF [2].

Materials. The stimuli consisted of 72 prompts with two response options per prompt. Half of the prompts consisted of a statement expressing an opinion on some moral or ethical issue (e.g. "Murder is sometimes justifiable."). Participants indicated their agreement, by button-press, with one or the other of two possible response options (e.g. "Never justifiable' vs 'Sometimes justifiable'). These stimuli were designed with the explicit goal of generating uncertainty and conflict in choosing. To that end, response options did not necessarily represent the extreme endpoints of an opinion spectrum. For example, in response to the statement "Murder is sometimes justifiable," the extreme opinion endpoints might be "Never justifiable" and "Always justifiable," but in this case the latter response is expected to be universally undesirable, and therefore these two options would be unlikely to generate uncertainty and conflict.

The other half of the prompts were non-moral filler questions regarding opinions or facts (e.g. "Do people respect selflessness?" or "Can bacteria live in boiling water?"). Response options to these stimuli were always "Yes" or "No." As they were in the studies of Pärnamets et al (2015) and Newell and Le Pelley (2018), these non-moral items are considered "filler" items and are included mainly to prevent participants from focusing exclusively into a mindset of moral reasoning.. In principle, the filler items may also show an effect of the gaze-based timing manipulation. However, given that these stimuli were not normed to be near 50/50 uncertainty, given that the word length of the response options is much shorter than those in the moral condition, and given that the response options are identical for all filler items, we expected gaze durations to be brief, and therefore these items may quite frequently result in time-out trials. As such, we make no strong predictions regarding the presence of an effect for these non-moral filler items.

Prompts were presented auditorily over headphones at the participants' preferred volume. Response options consisted of white text centered in a 300 x 300 pixel white box on a black background. Boxes were centered vertically and placed on the left and right sides of a 1920 x 1200 pixel screen, with

[2] https://osf.io/z9r47/

a 30 pixel buffer between each box and the closest edge of the screen. Text was displayed in Times New Roman size 70 font.

Apparatus. Duration of gaze to each of the two response options was recorded using a head-mounted Eyelink II eye tracking system, sampling eye position at 250 Hz. Before beginning the experiment, the eye-tracker was calibrated using a standard nine-point grid. The subject was then shown how to perform a drift correction, which took place prior to each trial. Gaze data was collected via the Eyelink control software and custom MATLAB scripts. Data from the right eye was collected using both pupil shape and corneal reflection. Gaze durations were counted toward a given response option only when the detected x and y coordinates fell within a 300 x 300 pixel white box centered on the respective response option.

Procedure. Participants completed the experiment individually in the lab. Participants were seated in front of a computer and wearing headphones with the volume set to their most comfortable level. The experiment was run using the Psychophysics Toolbox package (Brainard, 1997) in Matlab. On each trial, a white fixation dot was displayed in the center of the screen while the audio prompt played over the headphones. Once the auditory prompt finished playing, the two response options would appear on the center-left and center-right sides of the screen. The left or right position of each response option was randomized. On 36 randomly selected trials, the left option was selected as the target, while the right option was selected on the remaining 36. After each trial, participants rated their confidence in their choice as well as their understanding (the degree to which they read and understood both response options) on a 1-7 scale.

As in Pärnamets et al. (2015) and Newell and LaPelley (2018), participants were prompted to make a decision once they had fixated the software's target option for at least 750ms, and the alternative option for at least 250ms – or after a maximum of 3s if those criteria were not met. These fixation-time thresholds were counted cumulatively, rather than sequentially. The 750/250ms thresholds were chosen so that participants at least had to fixate both options, but would on average have viewed the target option for slightly longer when the response prompt was delivered. The 3s maximum was set to make it less likely that participants would notice that the response prompt was linked to their gaze. A post-experiment survey probed for knowledge of the manipulation, and no participants reported noticing the gaze-based manipulation.

Results

To begin, we first visualize the relationship between gaze and choices in Figure 1. On the x-axis is "target bias," which is the difference between the time spent fixating the target option versus the alternative option. The y-axis shows the percentage of trials on which the target was chosen (and by complement, the percentage on which the alternative was chosen). This plot reveals that as target bias increases, participants are more likely to select the target, and vice versa. There is also an effect of the last-fixated option prior to response prompting such that participants were more likely to choose the target when they had last fixated the target, and vice versa.

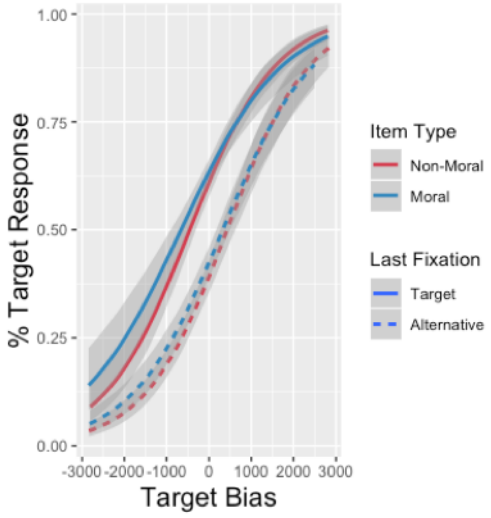

Figure 1: Percent of trials on which participants selected the target option as a function of target bias and last-fixated object before response prompting.

While Figure 1 clearly demonstrates that fixations are at least an *index* of preferences, we next analyzed whether the gaze-based manipulation also *influenced* decisions. Following Pärnamets et al. (2015) and Newell and Le Pelley (2018),

Figure 2: Percent of trials on which participants selected the target option for non-moral (red, left) and moral (blue, right) statements when time-out trials are excluded.

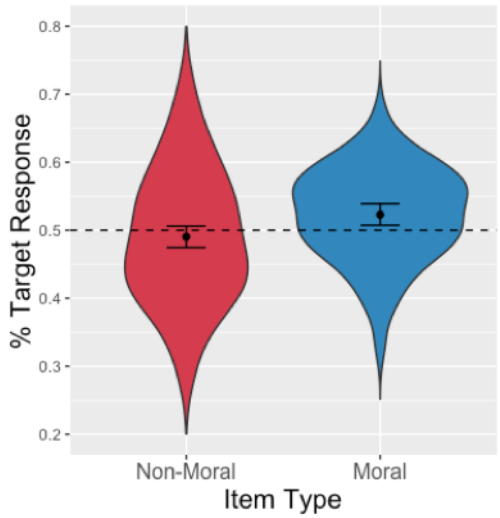

Figure 3: Percent of trials on which participants selected the target option when time-out trials are included.

the data were analyzed using one-sided, one-sample t-tests of the hypothesis that participants selected the target more often than 50%. We first analyzed the data set *excluding* time-out trials. These data are plotted in Figure 2. Time-outs occurred on 44.58% of trials overall, 23.59% of moral items, and 64.87% of non-moral (filler) items. We found that participants selected the target option 63.78% of the time overall ($t_{(55)}$ = 11.403, p $<$ .001.), 62.67% of the time for moral items ($t_{(55)}$ = 9.129, p $<$ .001) and 65.67% of the time for filler items ($t_{(55)}$ = 6.692, p $<$ .001).

However, consistent with the critique made by Newell and Le Pelley (2018), participants were significantly more likely to select the *non*-target option on time-out trials. Among time-out trials, participants chose the target only 20.18% of the time for moral items ($t_{(55)}$ = 12.76, p $<$ .001) and 40.16% of the time for filler items ($t_{(55)}$ = 6.4, p $<$ .001). Therefore, we next analyzed the full data set including both successful trials and time-out trials. These data are plotted in Figure 3. For moral items, participants selected the target option 52.3% of the time ($t_{(55)}$ = 2.537, p = .007), but only 49% of the time for filler items.

To account for subject- and item-level variability, the full data (including time-out trials) was also analyzed using separate logistic mixed-effects models for the moral and filler stimuli. The only fixed effect was the intercept term. The random effects structure included random intercepts for participants and items. These analyses revealed no significant difference from chance for filler stimuli, and only a marginally significant difference for moral items (b = .093, SE = 0.054, z = 1.713, p $<$ .087).

Across all trials, the mean rating for understanding was 6.43 (on a 1-7 scale); SD = 1.14) indicating that participants were able to read and understand both response options on most trials. The mean confidence rating was 5.4 (SD = 1.46) overall, 5.19 for non-moral items (SD = 1.68), and 5.6 for moral items (SD = 1.44). The relationship between confidence and response was analyzed using a linear mixed effects model with target response (coded as 0 or 1, for whether a participant selected the non-target or target, respectively) as a fixed effect and a random intercept for each participant. This analysis revealed no significant difference in confidence when selecting the target vs the non-target. However, when the timing manipulation was engaged (i.e. excluding time-out trials), participants were more confident when choosing the target than the non-target (b = .245, SE = .062, χ^2 = 3.942, p $<$ .001).

Discussion

We sought to reconcile inconsistent results reported by Pärnamets et al. (2015) and Newell and Le Pelley (2018) through a replication with an adjusted stimulus set. While Pärnamets et al. (2015) found that the effect of the gaze-based timing manipulation remained statistically significant (though reduced in size) when including time-out trials, Newell and Le Pelley (2018) found that the effect disappeared completely for both moral and for filler statements. We considered the possibility that the lack of an effect in the latter case was due to the re-use of the original stimulus set without re-norming it for the experimental population, which could have resulted in items for which participants had strong pre-existing preferences that washed out the effect of the manipulation.

The results of our experiment are consistent with the findings of Pärnamets et al. (2015), showing a clear effect of the gaze-based timing manipulation for experimental items when time-out trials are both excluded and included (though this effect was significant when using t-tests, in keeping with previous work, but not when using mixed-effects analysis, most likely due to insufficient power when accounting for participant- and item-level variability). While the effect disappears completely for our filler statements when including time-out trials, this is unsurprising since, as mentioned above, the response options for all filler trials consisted of a simple "yes" and "no," with the result that participants need only fixate one of the two options for a brief period of time in order to know what *both* options were. As a result, the manipulation was not engaged for most of those trials (time-outs occurred on $\sim$65% of fillers). This lends support to the notion that the lack of an effect found in Newell and Le Pelley (2018) was the result of re-using the original stimulus set without re-normalizing the stimuli to generate uncertainty in the sampled population.

Simulation

To provide some insight into the potential mechanisms and processes underlying the effect of this gaze-based timing manipulation on moral decision making, we designed a simple model of the cognitive deliberation process and how it might get perturbed by an interruption that triggers a premature response. The drift diffusion model (DDM) is a standard model

for simulating choices and response times in a two-alternative forced choice task (e.g. Pärnamets et al., 2015; Ratcliff & McKoon, 2008). The DDM assumes that decisions are made through the stochastic accumulation of perceptual evidence until a decision threshold is exceeded. The standard DDM represents the relative evidence for one of two alternatives at time t as $x(t)$ according to the following equation (Bogacz et al., 2006):

$$x_{t+1} = x_t + A + W \quad (1)$$

When x is 0, the two options have equal relative evidence, while positive or negative values indicate greater evidence for one option than the other. The change in evidence over time is the result of a constant "perceptual evidence" factor, A, plus Gaussian noise, W. For an initially undecided choice, $x = 0$ at $t = 0$, indicating equal support for each of the two response options.

While the standard DDM is designed to represent perceptual decisions based on a single stimulus, Krajbich, Armel, and Rangel (2010) adapted this model to the context of choosing between two displayed stimuli through visual sampling. Their model, which provided a close fit to human data, allowed the rate of change in x to vary as a function of the currently-fixated option, supporting the claim that fixations modulate preferences. For this version of the model,

$$x_{fixated,t+1} = x_{fixated,t} + d(A_{fixated} - \theta A_{unfixated}) + W \quad (2)$$

where θ is a value between 0 and 1 which discounts the value of the currently unfixated option, and d represents the rate of information accumulation.

In modeling gaze behavior, we adopted the following simplifying assumptions: (1) one alternative is fixated at any given time, (2) the first fixation on any trial is random, (3) there is a minimum fixation length, after which fixation switches are determined by competition between current preferences and attentional fatigue, and (4) saccades are instantaneous. The minimum fixation time was set at 200ms, which is approximately the time required to plan and launch a saccade (Salthouse & Ellis, 1980). After this period, we assume that attention decays at a rate determined by current preferences until gaze is switched to the alternate option. To model this, we introduced an "attentional fatigue" parameter. After the currently fixated object has accumulated >200ms of gaze consecutively:

$$a_{t+1} = x_t - f, \qquad x > 0 \quad (3)$$

$$a_{t+1} = x_t + f, \qquad x < 0 \quad (4)$$

where a represents the current attentional state and f represents attentional decay. For the first 200ms of any fixation, a is exactly equal to x, but after this time begins to move towards zero. When a crosses the zero-value and changes sign, gaze is directed towards the alternate option. Because a is coupled to x, greater magnitudes of x can offset the decay from f, such that the model looks longer at options that it "prefers," despite some attentional fatigue. Similar attentional parameters are commonly used in dynamical systems models of bi-stable perceptual phenomena, such as the Necker cube, to account for perceptual reversals (Ditzinger & Haken, 1995; Fürstenau, 2007). The red lines in Figures 4 and 5 show how attention decays as compared to decision preference (black lines), leading to gaze-changes (alterations between blue and yellow regions). Note that, while attentional fatigue can lead to gaze switches, unless there is a corresponding switch of preferences, gaze will switch back to the preferred option after the minimum fixation time (e.g. in Figure 4 the brief period of fixating the target from ~1200-1400ms).

On each simulated trial, the pre-chosen target was randomly assigned to one of the two response options. Each trial was run for a maximum of 3000 timesteps, where each time step represents 1ms, analogous to the 3s time limit in our experiment. We recorded the number of time steps spent "fixating" each alternative. If at least 750 time steps of gaze accumulated on the target side and at least 250 time steps on the alternative side (analogous to the 750ms/250ms threshold in the experiment), the trial was ended, and a positive x value resulted in choosing the reference option (coded as +1) while a negative x value resulted in choosing the other option (coded as -1). Figure 4 shows an example trial where the simulation met the gaze-time thresholds, time out after 2245 timesteps, and selected the target option. Figure 5 shows an example trial where the simulation did not fixate the target for long enough, leading to a time-out after 3000ms, after which the simulation selected the alternatives.

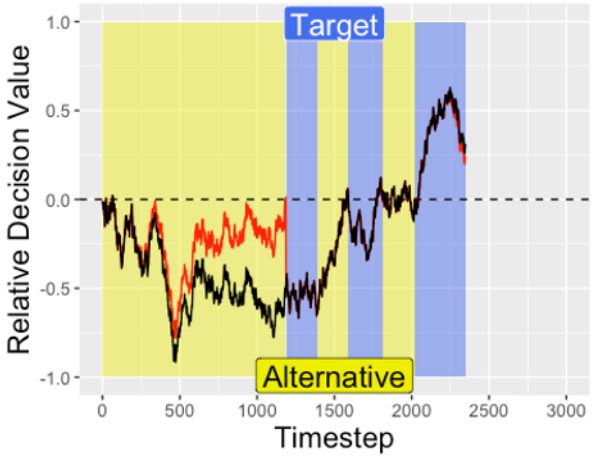

Figure 4: An example simulated trial on which the DDM was trending toward preferring the non-target, but once it achieved the gaze-time thresholds, it selected the target. Periods of fixating the target are marked in blue, with fixations to the non-target alternative in yellow. The red line represents attention, which decays faster than the relative decision value (in black).

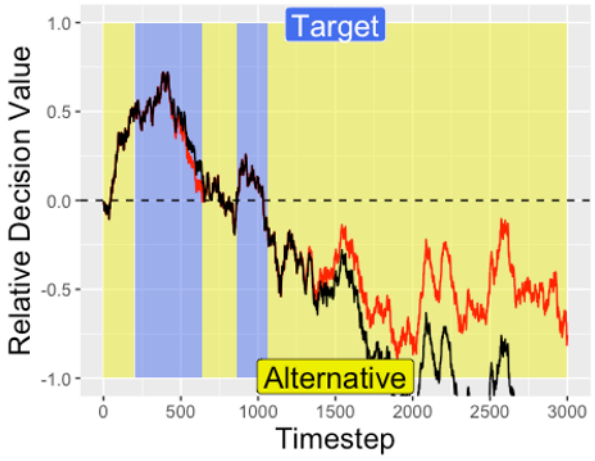

Figure 5: An example simulated trial on which the DDM "timed-out" after 3s (because cumulative target fixation time < 750 timesteps), then selected the non-target alternative. Periods of fixating the target are marked in blue, with fixations to the non-target alternative in yellow. The black line represents the relative decision value, while the red line represents attention.

Results

This simple model was not intended to precisely characterize these psychometric variables in our population, but rather to show that drift diffusion models straightforwardly predict the pattern of results obtained when using biased or unbiased stimuli. Thus, to avoid overfitting, no parameter tuning was done. The gaze-bias parameter (θ) was set to .5, such that the currently-unfixated option was discounted by half. The rate of information accumulation (d) was set to .001 and the gaussian noise (W) was set to a mean of 0 and a standard deviation of .01. The attentional fatigue parameter (f) was set to .0005. To simulate our normed stimuli, we set the values of both options .5. To simulate biased stimuli, we set the values of one option to .8 and the other to .2, randomly determined on each trial. 10000 simulated trials were run for each set of values. For each trial, we recorded the choice made by the model (determined by the sign of x when the trial terminated) as well as whether or not the trial "timed-out" by reaching 3000ms without meeting the gaze-time thresholds.

The general behavior of this simulation approximates our data remarkably well, especially given that we have not systematically explored the parameter space with this model. Beginning with our simulated moral stimuli (when values were set to .5 for both options), the model resulted in an average of of 42.28% time-out trials (compared to 23.59% in the human data). The model chose the target option 79.89% of the time when time-outs were excluded (compared to 62.67% for the humans), and selected the target 27.53% of the time on time-out trials (compared to 20.18% for humans). When time-outs were included, the model selected the target 57.75% of the time (compared to 52.3% for the humans).

For our simulated filler stimuli, where values were set to .8 and .2, the model timed-out 55.78% of the time (compared to 64.87% for the humans). The model chose the target option 82.2% of the time when time-outs were excluded (compared to 65.67% for the humans), and selected the target 34.31% of the time on time-out trials (compared to 40.16% for humans). When time-outs were included, the model selected the target 55.49% of the time (compared to 49% for the humans).

Discussion

The results of this very simple drift diffusion model provide a close approximate match to the results of our experiment. The relatively uncertain moral stimuli, exhibiting minimal intrinsic cognitive bias toward either of the response options, produced time-out trials less than half of the time, whereas the intrinsically biased filler stimuli produced time-out trials more than half of the time. When these time-out trials were excluded from analysis, both moral and filler stimuli exhibited strong choice preferences for the pre-chosen target response in the 65% range, just as that seen in our human data. However, when filler trials were included in the analysis, only the moral items showed a preference for the pre-chosen target response – again, just as that seen in our human data.

General Discussion

Our results provide converging evidence that even seemingly "high-level" cognition such as moral reasoning is a product of a highly interactive dynamical system – not the product of an informationally-encapsulated cognitive module. First, our experiment replicated the designs used in Pärnamets et al. (2015) and Newell and Le Pelley (2018) with an adjusted stimulus set that was normed to generate uncertainty in the experimental population. With this adjustment, which was not used in Newell and Le Pelley (2018), the effect of the gaze-based timing manipulation remained for moral decisions even when including time-out trials. This lends support to the original finding and indicates that it is not merely a methodological artifact. Although the magnitude of the effect is reduced when time-out trials are included, it is still significant (see also Ghaffari & Fiedler, 2018). Second, our drift diffusion model simulation provides some insight into the potential mechanisms underlying this effect. When the cognitive bias for a query is about equally balanced for the two response options, the sensory tendency to fixate both options is strong, thus promoting the likelihood that both response options will indeed be fixated. When that fixation pattern just happens to adventitiously exhibit a bias toward the option that the software has pre-chosen as the "target", interrupting the deliberative process at that point has a good chance of triggering a response based on the option that was most recently being fixated – which is likely to be the "target" response. Thus, perhaps even something as humanly precious as our moral decision making is not ushered forth solely from some internal "moral compass," but is also influenced by the subtle timing of our sensorimotor interactions with the world.

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433–436.

Ditzinger, T., & Haken, H. (1995). A synergetic model of multistability in perception. In *Ambiguity in mind and nature* (pp. 255–274). Springer.

Falandays, J. B., Batzloff, B. J., Spevack, S. C., & Spivey, M. J. (2018). Interactionism in language: From neural networks to bodies to dyads. *Language, Cognition and Neuroscience*, 1–16.

Fürstenau, N. (2007). A computational model of bistable perception-attention dynamics with long range correlations. In *Annual conference on artificial intelligence* (pp. 251–263).

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

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. MIT press.

Gu, J., Zhong, C.-B., & Page-Gould, E. (2013). Listen to your heart: When false somatic feedback shapes moral behavior. *Journal of Experimental Psychology: General*, *142*(2), 307.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature neuroscience*, *13*(10), 1292.

Newell, B. R., & Le Pelley, M. E. (2018). Perceptual but not complex moral judgments can be biased by exploiting the dynamics of eye-gaze. *Journal of Experimental Psychology: General*, *147*(3), 409.

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

Parnamets, P., Richardson, D., & Balkenius, C. (2014). Modelling moral choice as a diffusion process dependent on visual fixations. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 36).

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, *17*(3), 145–171.

Prinz, J. (2007). *The emotional construction of morals*. Oxford University Press.

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

Salthouse, T. A., & Ellis, C. L. (1980). Determinants of eye-fixation duration. *The American journal of psychology*, 207–234.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature neuroscience*, *6*(12), 1317–1322.

Spivey, M. (2008). *The continuity of mind*. Oxford University Press.

Tversky, A., & Kahneman, D. (1989). Rational choice and the framing of decisions. In *Multiple criteria decision making and risk analysis using microcomputers* (pp. 81–126). Springer.