Expectations bias judgments of harm against others

Derek Powell
derekpowell@stanford.edu
Department of Psychology
Stanford University

Zachary Horne
zachary.horne@asu.edu
Department of Psychology
Arizona State University

Abstract

People's expectations play an important role in their evaluations and reactions to events. There is often disappointment when events fail to meet expectations—sometimes even when the events are still positive overall—and there is a special thrill to having one's expectations exceeded. In four studies, we examined how expectations influence people's judgments of events where another person or people were harmed. Participants judged pairs of events where a victim experienced a similar harm, but where victims were at different prior risk of being harmed. We found that people judged these events as being worse when they were less expected–that is, when the victims were initially at lower risk of being harmed. We argue that this bias has pernicious moral consequences.

Keywords: Judgment and decision-making; Moral judgments; Bias

Introduction

On the evening of November 13th, 2015, a terrorist attack in Paris left 130 people dead and injured over 300 more. In the aftermath, millions took to Twitter to express their shock, horror, and outrage at this tragedy under hashtags like #parisattacks and #jesuisparis. Yet, most of those mourning had little to say 15 hours earlier, when another tragic attack killed at least 43 people in Beirut. Several factors are surely at play in these different reactions (e.g. group affiliations, Brewer, 1999), yet one potentially fundamental factor has gone unmentioned: the fact that the Paris attack was more surprising than the attack in Beirut. In contrast to France, Lebanon had experienced dozens of terrorist bombings and attacks in recent years. Consequently, Beirut may seem to many like the sort of place where "these things happen," whereas Paris is perceived as being stable and safe.

We can see in everyday experience that people's evaluations of events often depend on their expectations about those events. There is often disappointment when events fail to meet expectations, and there is a special thrill to having one's expectations exceeded. Anecdotally, these forces seem to drive people's tendencies to root for the underdog, hold surprise parties, and foreshadow bad news to ease its delivery (Bell, 1985).

Indeed, laboratory studies suggest that expectations play an important role in people's evaluations of the utility of an event. For instance, Mellers and colleagues (1997) found that expectations influenced affective reactions during a gambling task. Given a gamble with a 10% chance to win $30 and 90% chance to win $0, little disappointment is felt upon resolving with the $0 outcome, but considerable elation is experienced upon winning the $30. Conversely, given a gamble with a 90% chance of winning $60 and 10% chance of winning $0, the $0 outcome engenders considerable disappointment and the $60 relatively muted enjoyment. In fact, in gambles similar to these, Mellers and colleagues (1997) found that people were happier with the smaller unexpected gain than with the larger but more expected gain (also see Shepperd & Mcnulty, 2002).

Considering the important roles of both utility and affect in moral judgment (e.g., Greene, Sommerville, Nystrom, Darley, & Cohen, 2001), it is plausible that expectations might shape how people react to morally harmful events, such as acts of terrorism. However, unlike in the context of gambles, in these contexts the effects of expectations on evaluations may have harmful consequences. When events are shocking, people may perceive them as more severe and consequently be roused to action. In contrast, when events harm victims who are generally considered to be at greater risk–the poor, sick, or those living in unstable regions of the world, reactions may be more muted. If so, observers who learn of these events may experience reduced moral concern, and thus be less likely to donate time or money to aid victims, to take political action, and so forth.

Here, we examined whether people's evaluations of morally harmful events are affected by their expectations about those events. We asked people to compare pairs of simple events where a victim suffered an identical harm, but where the events differed in their prior probability. For each pair of events, participants were asked to judge which event was worse. Across four studies, we found that people tended to view unexpected negative events as worse, even when the harm to victims was identical.

General Methods

Here we present four studies examining the role of expectations in moral evaluations. In Studies 1a and 1b, using a forced-choice task, we tested the hypothesis that people would exhibit stronger reactions toward unexpected as compared to expected negative events. Studies 2a and 2b provided more stringent tests of our hypotheses by using new items and a more conservative judgment task to increase the generalizability of our results.

Materials

In all four studies, participants were presented with a series of trials where they read brief (one sentence) descriptions of two different events and were asked to indicate which of the two events seemed worse. In "experimental" trials, the two events were highly similar, but differed in their prior probabilities: one event was more expected and one more unexpected. (The perceived likelihood of a given event was confirmed in prior norming studies). These prior expectations were manipulated by changing the context in which the events occurred. For example, participants considered the following stimulus:

- “A 30 year old man in California dies in an earthquake” [Expected]

- “A 30 year old man in Oklahoma dies in an earthquake” [Unexpected]

In each event, the harm to the victim is the same (here, death) but one event is should be more expected than the other, given the different likelihoods of earthquakes occurring in California versus Oklahoma.

Each study contained between 6 and 12 experimental event-pairs that spanned a variety of different events and contexts. All experimental materials for these studies are available as supplemental online materials at `https://osf.io/a6pbj/`.

Studies 1a and 1b also included “equivalent” filler trials. In these trials, the two events differed in trivial contextual details that we did not expect would affect participants' judgments. For example:

- “A man in Connecticut starts a house fire.” [Equally expected]

- “A man in New Hampshire starts a house fire.” [Equally expected]

These filler trials were meant to prevent participants from becoming explicitly aware of the structure of the experimental trials.

In addition to these filler trials, studies 2a and 2b added “non-equivalent” filler trials, the two events differed substantially in the degree of harm suffered by a victim, so that one event was expected to be seen as considerably worse than the other. For example:

- “An 11-year-old child sets a doll on fire” [Less severe]

- “A 12-year-old child sets a cat on fire” [More severe]

These trials were included to allow participants a chance to use the extremes of the response scale and to reduce any task demands that might drive them to make artificially fine-grained distinctions between the severity of events.

Exclusions

Each study also made use of attention-check items. These questions asked participants to enter a particular response to ensure that they were paying attention and reading the items as they proceeded through the study. A final question asked participants if they had paid attention and taken the study seriously, encouraging them to be honest in their replies.

Data Analysis

We analyzed our data by performing Bayesian estimation using the probabilistic programming language Stan (Carpenter et al., 2017). We tested our predictions by computing Bayes Factors (i.e. BF01) on the intercept term of our regression model using the hypothesis function in the R package brms. As a reminder, Bayes Factors express the ratio of the probability of data under the null hypothesis to the probability of the data under an alternative hypothesis. Therefore, larger Bayes Factors indicate that the data are more likely under the null hypothesis (e.g., that the intercept is not different from zero) than the alternative hypothesis (e.g., that the intercept is different from zero), and vice versa. Bayes Factors can be influenced by prior choices so we also performed prior robustness checks to confirm that the prior alone was not accounting for the effects that we predicted.

Study 1a

Study 1a examined the hypothesis that changes in the prior probability of an event would affect people's evaluation of that event.

Participants

A total of 55 participants were recruited from Amazon's Mechanical Turk work distribution website (mTurk). Of these, 53 passed attention checks and were included in the final analyses (24 male, 29 female, median age = 32 years old). All participants were paid \$1.00 for their participation.

Materials and procedure

Participants judged 12 experimental event-pairs and 12 equivalent filler event-pairs. The events were described in the passive voice, and participants were asked to judge which event seemed worse. For example:

- A 32 year old woman gets food poisoning after eating a hamburger at a fast food restaurant. [Expected]

- A 32 year old woman gets food poisoning after eating a hamburger at a four star restaurant. [Unexpected]

On each trial, participants were presented with the event-pair stimulus and had to judge which outcome was worse in a two-alternative forced choice task. The two events were labeled “Outcome 1” and “Outcome 2” and their order was randomized.

Results and discussion

We predicted that people would think that events where unexpected harm occurred were worse than events that entail similar harm but were comparatively more expected. As indicated in Figure 1, when forced to choose, people judged that unexpected negative events were worse than expected events. To confirm this difference formally, we fit a Bayesian logistic random effects model with participants' responses as the dependent variable (1 = unexpected event is worse; 0 = expected event is worse) and a random intercept for subject. The intercept in this model represents the log-odds of selecting the unexpected event as being worse. Thus, by examining the population-level intercept, we can test whether participants were biased toward selecting the unexpected event ($\beta > 0$), the expected event ($\beta < 0$), or were unbiased ($\beta = 0$). Consistent with our hypothesis, we found that people were much more likely to think that unexpected events were worse than events that were expected, Intercept = 1.141, 95% CI [0.916, 1.386], BF01 < .001. Bayes factors and the estimate of the intercept were similar under different prior choices.

Figure 2 (panel 1) shows participants' responses broken-down by individual items. Participants' bias toward selecting the unexpected event as worse was largely consistent across the 12 experimental items.

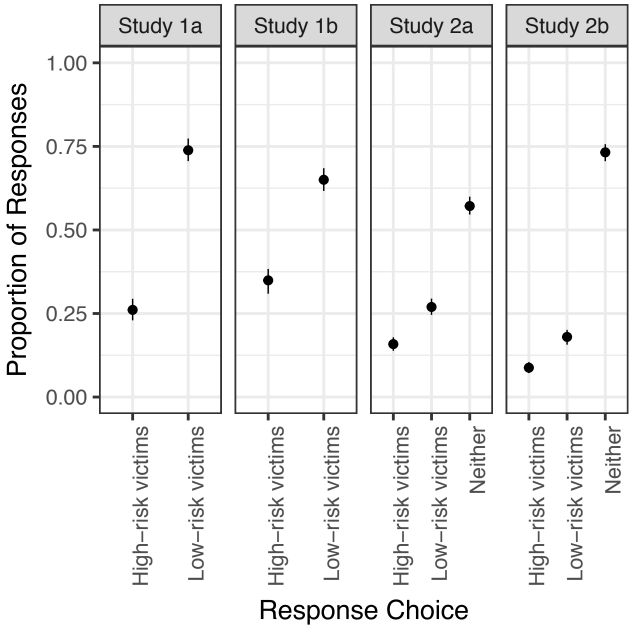

Figure 1: Proportion of response choices across studies 1-4 (pooled across items). Error bars indicate 95 % bootstrapped CI.

Study 1b

Consistent with prior work suggesting that people's utility evaluations are affected by their expectations (e.g., Mellers et al., 1999), Study 1a provided evidence that people view unexpected moral harm as worse than expected moral harm. To expand on these findings, and conceptually replicate the results of Study 1a, in Study 1b we asked people to evaluate other people's actions rather than the outcomes of events.

This study allowed us to test whether expectations might also bias people's judgments of others' actions.

Participants

A total of 112 participants were recruited from Amazon's Mechanical Turk work distribution website (mTurk). Of these, 110 passed attention checks and were included in the final analyses (61 male, 49 female, median age = 30 years old). All participants were paid $1.00 for their participation.

Small sample sizes tend to overestimate effect sizes (Button et al., 2013). Consequently, we also increased our sample size to confirm that the large effect observed in Study 1a was robust.

Materials and procedure

Participants judged 6 experimental event-pairs and 6 equivalent filler event-pairs. These event-pairs were adapted from event-pairs in Studies 1a and 1b in which a victim is harmed by another person's actions. The events were rephrased into the active voice in order to focus on the agent who took the action rather than the victim who was harmed by it. The actions participants selected between were labeled "Action 1" and "Action 2." For example, participants were presented with the following stimulus and had to judge which action was worse:

- "A wanted criminal shoots and wounds a police officer during a drug raid." [Expected]
- "A wanted criminal shoots and wounds a police officer during a traffic stop." [Unexpected]

As in Study 1a, participants were asked to choose which of the two actions seemed worse in a two-alternative forced choice task.

Results and discussion

We predicted that people would think that unexpected actions that caused harm were worse than expected actions that entailed similar harms. Just as we found in Study 1a, people judged that unexpected actions were worse than expected actions (see Figure 1, panel 2). We confirmed this difference formally by again fitting a Bayesian logistic random effects model with participants' responses as the dependent variable (1 = unexpected action is worse; 0 = expected action is worse) and a random intercept for subject. This analysis indicated that people were more likely to think that actions that were unexpected were worse than actions that were expected, Intercept = 0.675, 95% CI [0.493, 0.862], BF01 < .001.

Figure 2 shows participants' responses broken-down by individual items. Participants' bias toward selecting the unexpected action as worse were reasonably consistent across the six experimental items but there appeared to be more variation than we observed in Study 1a.

In summary, Study 1b suggests that people think that unexpected actions are worse than expected actions, again indicating that when comparing to events, people's reactions to negative events are influenced by their expectations.

Study 2a

In Studies 1a and 1b we found that people's judgments of events were biased by their expectations about those events. When forced to choose between two events, participants decided that unexpected events were worse than expected

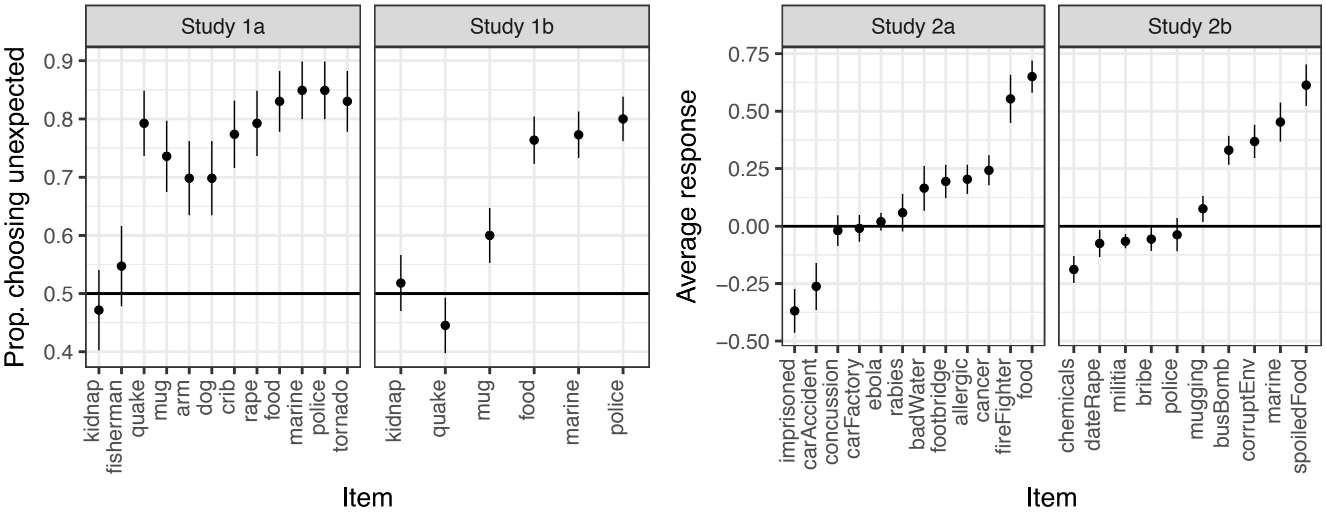

Figure 2: Responses by item for studies 1-4. Error bars indicate standard errors. Responses in studies 2a and 2b are represented using scale-means for visualization purposes only (higher scores indicate greater bias toward unexpected event).

events. In Study 2a, we sought to test our hypothesis using a more conservative method. Accordingly we made two changes in Study 2a: First, we introduced a new type of filler item "non-equivalent" filler trials and 2) we provided participants with a more expressive response scale so that if they viewed the events under consideration as equally harmful, their responses could reflect their attitude.

Participants

A total of 108 participants were recruited from Amazon's Mechanical Turk work distribution website (mTurk). Of these, 103 passed attention checks and were included in the final analyses (59 male, 44 female, median age = 31 years old). All participants were paid $1.00 for their participation.

Materials and procedure

Participants judged 12 experimental event-pairs. In all of these event-pairs, a victim suffers a negative outcome due to misfortune, rather than another person's intentional actions. Some event-pairs were reused from Study 1a without modification, others were revised or novel to improve the generalizability of our findings. These materials were created by (1) eliminating materials that may have confounded expectations with, for instance, an out-group bias and (2) creating novel items to again increase the generalizability of our findings. See supplemental online materials for a full list of items used in each study `https://osf.io/a6pbj/`.

In addition, participants judged 12 filler event-pairs. We introduced a new type of filler event-pair: "non-equivalent" event-pairs. As described previously, these are event-pairs that clearly differ in the degree of harm suffered or committed. For example, participants were presented with this stimulus and had to judge which was worse:

- "A man in Washington carjacks someone at gunpoint." [More severe action]
- "A man in Oregon steals a parked car." [Less severe action]

These items were introduced to address the concern that the high similarity within all event-pairs may drive participants to make overly-fine distinctions in their judgments. Such a task demand might inflate the effect sizes we observed in Studies 1a and 1b. Of the 12 filler events, six were "equivalent" event-pairs like those used previous studies and six were non-equivalent event pairs.

As in Study 1a, on each trial of the study, participants were presented with a pair of actions labeled "Outcome 1" and "Outcome 2" and were asked, "Which outcome seems worse?" However, unlike previous studies, in Study 2a participants made their rating on a five-point scale (Outcome 1 seems worse, Outcome 1 seems a little worse, neither seems worse, Outcome 2 seems a little worse, Outcome 2 seems worse). By forcing a choice between the two events, experiments 1a and 1b may have inflated the degree of bias participants exhibited. This more expressive response-scale was used in Studies 2a and 2b to address this concern.

Results and discussion

The events participants were asked to compare were, by design, highly similar. Consequently, we expected that participants' would typically indicate that neither event seemed worse. This was by far the choice participants most frequently made (see Figure 1, panel 3). However, we also observed that when participants did perceive one event was worse than the other, they were biased to perceive unexpected negative events as worse than more-expected negative events.

To examine these findings formally, we performed cumulative (ordinal) regression using a Bayesian random effects model with participants' scale responses as the dependent variable (1 to 5) and a random intercept for subject.

This model produces four intercept coefficients, representing the cumulative log-odds of responses at each scale point or higher. For instance, the second coefficient represents the log-odds participants chose a 2 ("outcome 2 seems slightly worse") or lower on the scale. Similarly, the third intercept coefficient represents the log-odds participants chose a 3 or lower on the scale. By comparing the second intercept coefficient to the inverse of the third intercept coefficient (thereby representing the log-odds *not* choosing a 3 or lower–i.e., choosing a 4 or 5), we can test whether participants were more likely to choose the expected or unexpected event as being worse in cases where they did not choose the neither option. This analysis indicated that people were more likely to think that events that were unexpected were worse than events that were expected, BF01 $<$ .001 (see supplemental online materials for full model results) – when participants did exhibit a bias in their responses about which event was worse, they reliably chose the unexpected event was worse than the expected event.

However, these findings should be qualified by acknowledging the considerable inter-item variability across the 12 items. Figure 2, panel 3 shows participants' responses across individual items. For visualization purposes only, we display these results using the mean response across the 5-point scale. Participants were strongly biased to perceive the unexpected event as worse for approximately half of the items, but were less strongly-biased for others, and slightly biased in the reverse direction for two items.

Study 2b

Participants

A total of 114 participants were recruited from Amazon's Mechanical Turk work distribution website (mTurk). Of these, 106 passed attention checks and were included in the final analyses (48 male, 58 female, median age = 31 years old). All participants were paid $1.00 for their participation.

Materials and procedure

Participants judged ten experimental event-pairs, five "equivalent" filler event-pairs, and five "non-equivalent" filler event-pairs. We created additional items in this study to improve and expand upon the event-pairs used in Study 1b.

As in Study 1b, these events all involved an action that harmed a victim. On each trial of the study, participants were presented with a pair of actions labeled "Action 1" and "Action 2" and were asked, "Which action seems worse?" Using the same procedure as Study 2a, participants made their rating on a five-point scale (Action 1 seems worse, Action 1 seems a little worse, neither seems worse, Action 2 seems a little worse, Action 2 seems worse).

Results and discussion

Participants pattern of responses were similar to those observed in Study 2a. We found that participants chose the "neither" option in the majority of trials, but when participants did perceive one action as worse than the other, they were biased to perceive unexpected negative actions as worse than more-expected negative actions (Figure 1, panel 4). To examine these findings formally, we again performed cumulative (ordinal) regression using a Bayesian random effects model with participants' scale responses as the dependent variable (1 to 5) and a random intercept for subject. To test our hypothesis, we compared the Bayes Factor for intercept coefficients representing the log-odds of choosing the expected and unexpected actions as worse or slightly worse. As predicted and suggested by Figure 1, this analysis indicated that people were more likely to think that actions that were unexpected were worse than actions that were expected, BF01 $<$ .001 (see supplemental online materials for full model results).

Here too, our findings should be qualified by acknowledging the considerable variability across the 10 items of Study 2b (see Figure 2, panel 4). As shown in the plot, participants were strongly biased to perceive the unexpected event as worse for four of the items, but showed almost no bias for the other six items.

Discussion

The results of four studies suggest that people view unexpected harmful events more negatively than expected harmful events. Just as people react more strongly to unexpected monetary gains and losses (Mellers et al., 1997), people similarly react more severely to unexpected moral harm than expected moral harm–judging those unexpected events as "worse".

Why should our expectations influence our reactions to events? A number of researchers have sought to develop theories of disappointment–the psychological reactions that result when experiences fail to meet expectations–and its role in evaluation and decision-making (e.g., Bell, 1985; Gul, 1991; Loomes & Sugden, 1986). These theories posit that decisions and evaluations are affected by the objective (e.g., economic) utilities of options and events, as well as disappointment individual people experience as a function of their expectations. Alternately, numerous theories of decision-making, including Prospect Theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), have emphasized the role of relative comparisons in evaluation and decision-making. In this vein, expectations might help set the reference points against which people compare potential future outcomes.

On these accounts, the influence of expectations on evaluation is simply a human quirk, a result of the way we evaluate events and decisions. In contrast, we suspect that expectations may influence evaluation through more principled means. The surprise of unexpected events may seem irrelevant to moral evaluations, but it is vital to learning. In Information Theory, the information carried by an event is a direct function of its prior probability, such that low probability events carry more information than high probability events (Shannon, 1948). Likewise, the violation of expectations has long been recognized as fundamental to associative and animal learning models (e.g., Rescorla & Wagner, 1972).

People may learn more about the state of the world when their expectations are violated by shocking world events, as compared to when they are affirmed by less surprising events. In this light, it seems intuitive that people would have stronger reactions to those surprising events. Still, the consequence of this dynamic is unquestionably suboptimal moral behavior.

Limitations

Although we consistently observed a bias to judge unexpected events and actions as worse than expected events across four studies, in Studies 2a and 2b, we also observed that the extent of the bias was quite dependent on the specific content of the items. This is perhaps an unsurprising consequence of our decision to use relatively naturalistic items and to manipulate expectations about these events implicitly by manipulating the context in which those events occurred. This technique has the obvious virtue of affording these items some degree of realism (as compared to artificial gambling tasks using explicitly stated probabilities that participants may or may not believe), but manipulating context may also affect other aspects of participant's interpretation of these actions, potentially introducing confounds. For example, despite our care in creating them, our items may have subtly confounded the likelihood of an event with the perceived race or socioeconomic status of the victims that were affected in the "likely" and "unlikely" events. Thus, it is possible that the expectations effect we observed was, in fact, driven by differences between the victims that participants imagined suffering the event. We did guard against this possibility by including a variety of different items and contextual manipulations. Still, future research is needed both to broaden these findings and to establish converging evidence through methods that are not subject to these concerns.

Conclusions

The bias to view unexpected harm as worse than more expected harm threatens to impose a vicious and morally pernicious cycle: For instance, people living in geo-politically unstable regions or in the developing world are often those who are most affected by terrorism, famine, and natural disasters, and are the very people in greatest need of assistance and concern from the world at-large. However, for these very reasons, it is often unsurprising when harm befalls people living in these circumstances. Our findings suggest a bias whereby the people most likely to suffer and be victimized are the very people that others are least likely to be moved to help. Future research should aim to understand the processes by which this bias arises and to identify how it might be counteracted.

Acknowledgements

We would like to acknowledge the help of Keith Holyoak, Alan Fiske, Matthew Lieberman, Hongjing Lu, John Hummel, and Ellen Markman for their comments and support for the project.

References

Bell, D. (1985). Disappointment In Decision Making Under Uncertainty. *Operations Research*, *33*(1), 1–27. http://doi.org/10.1287/opre.33.1.1

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*, *55*(3), 429–444. http://doi.org/10.1111/0022-4537.00126

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munaf, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience*, *14*(5), 365–76. http://doi.org/10.1038/nrn3475

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). http://doi.org/10.18637/jss.v076.i01

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science (New York, N.Y.)*, *293*(5537), 2105–8. http://doi.org/10.1126/science.1062872

Gul, F. (1991). A theory of disappointment aversion. *Econometrica*, *59*(3), 667–686. http://doi.org/10.2307/2938223

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292.

Loomes, G., & Sugden, R. (1986). Disappointment and Dynamic Consistency in Choice under Uncertainty. *Review of Economic Studies*, *53*(2), 271–282. http://doi.org/10.2307/2297651

Mellers, B. a, Schwartz, a., Ho, K., & Ritov, I. (1997). Decision Affect Theory: Emotional Reactions to the Outcomes of Risky Options. *Psychological Science*, *8*(6), 423–429. http://doi.org/10.1111/j.1467-9280.1997.tb00455.x

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning ii current research and theory* (Vol. 21, pp. 64–99). http://doi.org/10.1101/gr.110528.110

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. http://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shepperd, J. a, & Mcnulty, J. K. (2002). The affective consequences of expected and unexpected outcomes. *Psychological Science*, *13*(1), 85–88. http://doi.org/10.1111/1467-9280.00416

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.