Title
Expectation and Evaluation in Moral and Non-Moral Contexts

Permalink
https://escholarship.org/uc/item/9k11f8j5

Author
Powell, Derek

Publication Date
2016

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA
Los Angeles

Expectation and Evaluation in Moral and Non-Moral Contexts

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of
Philosophy in Psychology

by

Derek Powell

2016
ABSTRACT OF THE DISSERTATION

Expectation and Evaluation in Moral and Non-Moral Contexts

By

Derek Powell

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2016

Professor Keith Holyoak, Chair

People’s evaluations of events often depend on their expectations about those 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 this dissertation I propose a theory of event evaluation based on contrasts between global representations of states of the world that accounts for the role of expectations in evaluations. I then present four studies showing evidence for this model in moral and non-moral domains. In Study 1, I generated quantitative predictions from the model to fit existing data describing people’s emotional reactions to risky monetary gambles. Study 2 is a naturalistic study examining football fans’ reactions after wins and losses based on their Twitter activity. Fans’ reactions are influenced by their prior expectations of winning or losing each game. The evaluation of events is intimately linked to the evaluation of actions, in that one mode for evaluating action is simply to evaluate the events brought about by those actions. Consequently,
Studies 3 and 4 shift focus to judgments of moral actions. In Study 3, participants were asked to contrast pairs of identical actions taken against victims in different circumstances. Participants’ expectations about victims’ prior risk affected their moral judgments of the actions so that a victim’s greater prior risk mitigated the severity of their moral judgments. Study 4 examined a naturalistic dataset of global Twitter activity in response to terrorism events. Terrorism events evoked larger public reactions when those events were surprising, with weaker reactions resulting when terrorism occurred in countries that suffer from high rates of terrorism. Finally, I discuss the implications of these findings for the normative status of human moral judgments. I discuss how expectations may play a role in the omission-commission distinction, moral luck, and victim blaming. Although it seems justifiable that expectations influence evaluations of non-moral events, in judgments of moral actions this tendency may lead to pervasive moral error.
The dissertation of Derek Powell is approved.

Hongjing Lu

Matthew D. Lieberman

Alan Page Fiske

Keith Holyoak, Committee Chair

University of California, Los Angeles

2016
# Table of Contents

1. Introduction .................................................................................................................. 1
1.1 Utility .......................................................................................................................... 2
1.2 Utility and Morality ..................................................................................................... 4
1.3 Expectations and Moral Judgment ............................................................................. 8
1.4 Overview ..................................................................................................................... 9

2.1 The Role of Expectations in Utility Evaluation: Existing Theories ......................... 11
2.2. Accuracy and Error in Event Evaluation .................................................................... 13
2.2.1 Evaluation of Events by Contrasting Uncertain States of the World ..................... 16
2.2.2 Evaluating the State of the World .......................................................................... 19
2.2.2.1 Numerical Example ......................................................................................... 27
2.2.3 Estimating Utilities and Probabilities ...................................................................... 29
2.3 Extending Event Evaluations to Moral Judgments of Actions .................................. 31

3. Studies .......................................................................................................................... 35
3.1 Study 1: Modeling of Data from Mellers et al. (1997) ................................................ 35
3.1.1. Mellers et al. (1997) Experiment and Results ....................................................... 36
3.1.2 Generating Quantitative Predictions from the Contrastive Theory of Evaluation .... 38
3.1.3 Results and Model Comparisons ............................................................................ 40
3.1.4 Comparisons with Decision Affect Theory .............................................................. 41
3.1.5 Discussion of Study 1 ........................................................................................... 43
3.2 Study 2: NFL Fan Reactions in Response to Wins and Losses as Measured by Twitter Activity ......................................................................................................................... 43
3.2.1 Constructing the Dataset ....................................................................................... 44
3.2.2 Results .................................................................................................................... 46
3.2.3 Discussion of Study 2 ........................................................................................... 49
3.3 Study 3: Victims’ Prior Risk Affects Judgments of Immoral Actions ....................... 49
3.3.1 Experiment: Methods .......................................................................................... 49
3.3.1.1 Participants ....................................................................................................... 50
3.3.1.2 Materials ......................................................................................................... 50
3.3.1.3 Procedure ...................................................................................................... 51
3.3.2 Experiment: Results ............................................................................................ 52
3.3.3 Discussion of Study 3 .................................................................................................................. 54
3.4 Study 4: Twitter Activity Following Terrorism Events ................................................................. 56
3.4.1 Method .................................................................................................................................... 57
3.4.2 Results .................................................................................................................................... 60
3.4.3 Discussion of Study 4 ............................................................................................................... 70
4. Implications for Ethical Theory and Human Moral Judgments ....................................................... 72
4.1 Implications for normativity of human moral judgments .......................................................... 72
4.2 Omission-Commission Distinction ............................................................................................. 75
4.3 Moral Condemnation of Risky Actions ....................................................................................... 76
4.5 Victim Blaming .......................................................................................................................... 78
5. Conclusions .................................................................................................................................... 79
Acknowledgements

I must first thank my friend and collaborator Zachary Horne, both for his friendship and for his innumerable intellectual contributions to the work described here. We have worked together so closely and for so long now that there is likely some of his influence in any idea I would call my own. More specifically, he should be credited for collaborating alongside me in conceiving of the theory advanced here and in the execution of Study 3. I am also grateful for his helpful comments on earlier drafts and I am greatly indebted to his knowledge of ethics and moral psychology, of which I frequently availed myself in writing this dissertation.

I must also acknowledge and thank my advisor, Keith Holyoak, for his help in my entire graduate career. Keith has always been immensely supportive of me and my ideas, even when he has been skeptical of the wisdom behind them. I’ve been proud to work under him and to follow as best I could the example set by his research and his writing. I have greatly valued his advice and his sense of humor and hope I can continue to benefit from both after departing from UCLA.

Next, I am grateful to my committee members, Matthew Lieberman, Alan Fiske, and Hongjing Lu, for their comments at my preliminary orals and throughout the dissertation process. In particular, I would like to thank Alan Fiske for several excellent and challenging conversations and Hongjing Lu for her help with the formal mathematical aspects of Section 2.

Finally, though I hope they already know they have my love and appreciation, I want to thank my friends and family, and in particular my parents and my brother for their love and their support. I also want to acknowledge and thank my loving and incredibly supportive partner Cassidy. I love her, am continually amazed by her, and am grateful to be with her every day. Recently, I am particularly grateful for her patience as a sounding board, for reminding me of the rules of both algebra and the English language, and for helping to imagine misfortunes of
varying likelihoods. Whatever I have accomplished in my dissertation or elsewhere, it is hard to imagine how I could enjoy it without her, my friends, or my family.
Derek Powell

Education

<table>
<thead>
<tr>
<th>Year</th>
<th>Degree</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011–2016</td>
<td>PhD (expected)</td>
<td>University of California, Los Angeles, Psychology</td>
</tr>
<tr>
<td>2011–2012</td>
<td>MA</td>
<td>University of California, Los Angeles</td>
</tr>
<tr>
<td>2006–2010</td>
<td>BS</td>
<td>Arizona State University, Tempe, AZ</td>
</tr>
</tbody>
</table>

Peer-Reviewed Articles ( * authors contributed equally)


Book Chapters and Conference Proceedings


**Talks**


**Powell, D.** (2014, Sept.). Parsimony and competition in causal learning. Psychology department developmental division brown bag, University of Illinois at Urbana-Champaign.

**Powell, D.** (2014, Sept.). What is knowledge? Evidence from semantic integration. Psychology department cognitive division brown bag, University of Illinois at Urbana-Champaign.


**Posters**


**Honors and Fellowships**

UCLA Graduate Summer Research Mentorship 2014
UCLA Graduate Research Mentorship 2013-2014
National Science Foundation: Graduate Fellowship Honorable Mention 2013
UCLA Graduate Summer Research Mentorship 2012
Caton Award for Graduate Writing
Edwin W. Pauley Fellowship
1. Introduction

People’s evaluations of events often depend on their expectations about those events. This point seems evident in a variety of behaviors (Bell, 1985) such as rooting for the underdog, holding surprise parties, trying not to get one’s hopes up, giving bad news by “letting people down gently,” and the strategic setting of expectations in politics and business. There is often disappointment when events fail to meet expectations, and there is a special thrill to having one’s expectations exceeded. Sometimes we can be upset even by positive events if they fail to meet expectations, and we can be relieved by negative events when they do not turn out as badly as we feared.

Consistent with these intuitions, Mellers et al. (1997) found that expectations influenced affective responses during a gambling task. People had the strongest affective reactions when they won a gamble after expecting to lose or when they lost after expecting to win. Expectations sometimes played a greater role than the amount won—people were often more excited to unexpectedly win a small dollar amount than they were to win a much larger amount when the win was anticipated.

People’s evaluations of events are also influenced by expectations in more realistic contexts. Shepperd and McNulty (2002) reported a study in which experimenters posed as medical professionals administering a test for thiaomine acetylase (TAA) deficiency, a fictitious medical condition. Depending upon the condition to which they were assigned, participants were told that, as college students, they were at either high or low risk of TAA deficiency. Then, a mock test was administered and participants were given a “diagnosis” – either positive or negative for the condition. Participants given a positive diagnosis had more strongly negative self-reported emotional reactions to the news when they thought they were at low prior risk than
when they thought they were at high risk. Likewise, participants given a negative diagnosis had little reaction when they thought they were at low risk but had a stronger positive reaction when they thought they had been at greater risk. Another study showed that people are consciously aware of the role of expectations in shaping reactions. When asked to assess how a person would react to a $1000 raise, participants’ recognized that they would be happier if they had been expecting a $500 raise rather than expecting a $1500 raise.

1.1 Utility

The evaluation of events is typically expressed as an evaluation of their utility. Utility serves, in some form or another, as a foundational concept in virtually all theories of evaluation and choice (for a recent and concise review see Johnson & Busemeyer, 2010). Early economic theorists argued that rational agents should seek to maximize their total wealth or the total expectation of their wealth if they face risk or uncertainty (Stigler, 1986). However, this theory of choice lead to seemingly irrational conclusions. For instance, it would preclude the purchase of insurance.

In response to these issues, Bernoulli (1738/1954) introduced the concept of utility as a measure of psychological satisfaction. Bernoulli proposed to redefine rational economic behavior as that which maximizes utility (i.e., obtains the most utility possible for the individual). Further, where agents face risk or uncertainty, rational behavior is that which maximizes the mathematical expectation of utility. These ideas were formalized further as expected utility theory by Von Neumann and Morgenstern (1947) and later as subjective expected utility theory by Savage (1954). Von Neumann and Morgenstern (1947) laid out a set of axioms describing optimal choice based on maximizing expected utility. In risky decision contexts, this means maximizing the utility of choice outcomes weighted by their probability of obtaining. For
instance, the expected utility of a 25% chance to win $100 is half that of a 50% chance to win $100. Therefore, decision makers should prefer the latter to the former. Von Neumann and Morgenstern’s expected utility theory was expanded by Savage (1954), who argued that probabilities should be thought of as representing subjective degrees of belief (and which may or may not be perfectly accurate). Expected utility theory subsequently became and remains a widely accepted theory of rational choice (Mellers, 2000).

The relationship between psychological utility and some objective quantity in the world is described by a utility function. A utility function $U(x)$ is a mapping between some input $x$ and some real-numbered utility value measuring the psychological satisfaction brought about by $x$. The exact specification of such a function is clearest when the input is itself some real-valued quantity (e.g., money or commodities). Classically, Bernoulli proposed a log function mapping monetary wealth to psychological utility as $U(x) = \log(x)$, as shown in Figure 1. According to this function, an equal change in wealth will be felt more strongly for someone who is poor than for someone who is wealthy. Thus, Bernoulli’s utility function is able to capture the phenomenon of diminishing marginal utility or “diminishing returns” (Gossen, 1854/1983). Diminishing marginal utility also explains why a sure gain of $1 million might be preferable to a risky chance at $1.25 million.
In principle, the input to a utility function can be anything at all. Where the inputs cannot be clearly defined, the utility function remains an abstraction. A utility function represents utility from the subjective perspective of the evaluator. In many cases, people may share fundamentally similar utility functions (e.g., a concave utility function over wealth). In other cases, they may differ dramatically. For myself, a perfectly cooked steak has a good deal of positive utility; for my vegetarian partner, not so much. Prescriptions for rational choice can only be made under the assumption of some specific mapping of real-world quantities to utility. Thus divergent choices might each be rational for different individuals: it could be rational for me to seek to maximize my steak eating (perhaps up to some point) and for my partner to seek to minimize hers.

1.2 Utility and Morality

The assignment of utility to events is fundamental both to ethics and to people’s moral judgments. Ethics is chiefly concerned with determining what is good and what is right (Ross, 1930; Rawls, 1971). Much debate in ethics concerns the relative priority of these quantities.

Figure 1. Natural log function relating wealth to utility as proposed by Bernoulli.
Ethicists who take a teleological approach define the right in terms of the good. This leads to various form of consequentialist ethics (Mill, 1863/2004; Singer, 1979; 2005). According to consequentialist theories, actions are morally right to the extent that they produce or lead to an increase in some measure of the good (e.g., pleasure). Consequentialist moral reasoning involves anticipating the possible outcomes of an action and evaluating the moral utility produced by these outcomes.

In contrast, a non-teleological approach (Rawls, 1971) assumes that the right is prior to the good, generally leading to forms of deontology. Deontological ethics are systems of normative ethical rules that describe permissions and obligations to perform or refrain from performing different actions (e.g., Hurd, 1994; Kant, 1780; Kamm, 2007). Generally, these rules are derived from principles concerning others’ moral rights and duties. For instance, people have a right not to be physically harmed, leading to the duty to avoid harmful actions. This suggests that moral reasoning involves the application of moral rules. Of course, deontologists also acknowledge that there might be some measure of the good; they simply deny that what is right ought to be defined by this measure.

One form of consequentialism that has been hugely influential within moral psychology is utilitarianism, which originated with Bentham (1823/2009), and was subsequently advanced by John Stuart Mill (1863/2004), Henry Sidgwick (1907/1981), and many other more recent philosophers (e.g., Portmore, 2011; Singer, 1979, 2005). At its simplest, utilitarianism defines the right act as that which maximizes some utility measure summed across an entire group of moral patients. Theories differ in their assumptions about the relevant consequences (e.g., pleasure, happiness, or simply the satisfaction of preferences) and how they are best distributed (e.g., ethical egoism seeks the greatest good for me). Bentham (1823/2009) viewed utility as
happiness, or the balance of pleasure and pain. Mill (1863/2004) accepted this basic idea, but
drew a distinction between higher and lower pleasures, which he argued were qualitatively
different. Perhaps the purest form of utilitarianism is act utilitarianism, whereby the moral status
of an action is determined solely by the utility it produces. Thus, under act utilitarianism,
evaluations of actions and evaluations of events are one and the same: a moral judgment of an
action is made just by evaluating the utility of the events produced by that action.

A number of influential moral psychologists have treated act utilitarianism as the
normative moral standard against which human moral reasoning ought to be evaluated. (e.g.,
Baron, 1994; Greene, 2008; Sunstein, 2005). However, the status of act utilitarianism, and
consequentialism in general, is a major point of debate within ethics—many ethicists endorse
deontology rather than consequentialism (and for alternative views within psychology, see e.g.,
Bennis, Medin & Bartels, 2010; Mikhail, 2011; Rai & Fiske, 2011). Meanwhile, there is very
little reason to suspect that many people (save ethicists themselves) wholeheartedly endorse
either of these views. People often make both consequentialist and non-consequentialist
judgments (Baron, 1994; Greene et al., 2008). Moreover, when directly queried about their
beliefs, most people do not strongly endorse consequentialist or deontological viewpoints over
one another (Lombozo, 2009).

However, though few can be said to be truly utilitarian, people’s moral judgments are
clearly influenced by utility considerations. Much of moral psychology has focused on people’s
judgments of moral dilemmas pitting the violation of moral rules (e.g., “do not kill”) against
utility considerations (e.g., saving a greater number of lives). The classic “trolley” or “switch”
dilemma offers a prototypical case (Foot, 1967; Thomson, 1985). Imagine a runaway trolley is
headed toward five workmen who will be killed unless the trolley is switched down a side track
where only one workman is standing. Switching the track will kill this one workman but will save the five. In hypothetical dilemmas like this, people often judge that it is morally acceptable to sacrifice one to save the many (Greene et al., 2001, 2004, 2008; Moore et al., 2012; Kahane et al., 2011; Koenigs et al., 2007; Knutson et al., 2010; Hauser et al., 2007; Shenhav & Greene, 2014). Of course, this is not always the case. In the very similar “push” dilemma, a runaway trolley is again headed toward a group of five workmen, but this time the only way to stop it is to shove a large man into the path of the trolley, using his body to stop it and save the five workmen. People overwhelmingly disapprove of taking this action to save five lives. However, if the number of lives at stake is increased sufficiently, most people can be convinced that pushing the man is the right thing to do (Trémolière & Bonnefon, 2014).

People’s utility evaluations in moral judgment contexts appear largely consistent with their utility evaluations in economic and other non-moral decision making contexts. For instance, people exhibit similar patterns of behavior in situations where lives can be saved or lost as they do for situations where money can be gained or lost. People appear to aggregate lives using a concave utility function for gains and a convex utility function for losses, just as they do for monetary gains and losses (e.g., Kahneman & Tversky, 1984; Cropper, Aydede, & Portney, 1992; also see Trémolière & Bonnefon, 2014). People also exhibit similar biases in their selections of moral actions as they do when acting as consumers (Rai & Holyoak, 2010; Fetherstonhaugh et al., 1997). Research on mental accounting has shown that people tend to measure savings relative to a commodity’s cost rather than in absolute terms. For instance, people are more likely to go out of their way to save $5 on a $15 calculator than to save $5 on a $125 jacket (Kahneman & Tversky, 1984). Likewise, people are more willing to send money to save 4,500 lives from a refugee camp with 11,000 total refugees than to do so from a camp with
250,000 total refugees (Fetherstonhaugh et al., 1997). This mental accounting bias is also present in judgments of moral dilemmas: people are more willing to sacrifice two lives in order to save 8 of 10 total lives than to save 8 of 40 total lives (Rai & Holyoak, 2010).

1.3 Expectations and Moral Judgment

A study by Olivola and Sagara (2009) provides some initial evidence that expectations can shape evaluations and decision-making in the moral domain. These researchers conducted a cross-national study examining people’s decision making when lives rather than dollars were at stake. They asked participants in the U.S., Japan, Indonesia, and India to imagine themselves in the role of a public health official facing a deadly outbreak threatening the lives of 40 people. Participants were asked to choose between two programs aimed at addressing this outbreak—one that would save 20 lives but allow 20 to die with certainty, and another giving a 50% chance that zero people would die and a 50% chance that all 40 would die. Participants’ responses to this judgment task give an indication of how they value people’s lives. If each life is seen as equally important, then they should be indifferent between the two options. When participants prefer the risky option over the sure option, this implies a reduced sensitivity to loss of life as the number of lives lost increases.

Based on news reports in each country, Olivola and Sagara (2009) generated estimates of participants’ expectations surrounding mass loss of life. As might be intuitive, events with large death tolls are relatively more common in India and Indonesia than they are in the U.S. and Japan. Therefore, people’s expectations about the likelihood of mass loss of life in these countries should be expected to differ. Consistent with these different expectations, these researchers found greater risk seeking among U.S. and Japanese participants than among Indian and Indonesian participants. For those in the U.S. and Japan, the death of 20 people is already
such an extreme event that the death of 40 people is not seen as being that much worse—they are not very able to discriminate between the two events. Therefore, these participants saw no reason not to take a chance with the risky program. In contrast, participants in Indian and Indonesia, where larger death tolls are more expected, were better able to discriminate between events killing 20 and 40 people in their utility evaluations.

1.4 Overview

People’s evaluations of events are influenced by their expectations about those events. The first purpose of this paper is to propose a theory that explains this phenomenon. Section 2 describes this theory, under which event evaluations result from contrasts between the states of the world preceding and resulting from those events.

As discussed, event evaluations are an important component of moral judgments of actions. Consequently, the second purpose of this paper is to examine how expectations might affect moral judgments of people’s actions. Section 3 describes four studies providing evidence that expectations influence evaluations in moral and non-moral contexts. These phenomena are demonstrated both in the lab (Studies 1 and 3) as well as in naturalistic settings using a “big data” approach—looking at Twitter activity following real-world events (Studies 2 and 4).

Finally, Section 4 considers the implications of these findings for ethics and for the normativity of human moral judgments. The application of domain-general evaluative mechanisms to moral judgment can sometimes lead to moral error. For instance, expectations offer one potential explanation for people’s tendency to blame victims in certain circumstances. In addition, understanding the role of expectations in moral judgment may help to resolve certain disputes in ethics: differences in expectations may explain ethicists’ intuitive defense of the omission-commission distinction. Lastly, this account begins to offer an explanation for
judgments of risky but ultimately harmless behavior, an issue sometimes addressed under the heading of “moral luck” (e.g., Nagel, 1979; Williams, 1981).

2. A Global Contrast Theory of Event Evaluation

Evaluation is most often considered in the context of decision-making, as decision-makers must evaluate their options and select among them. However, people also make evaluations for their own sake. For instance, sometimes we need to take stock: to appraise the events that transpired in the wake of our actions so that we can learn from our mistakes or successes. We also engage in social evaluation, appraising the behavior of others to determine how we should relate with them (Fiske & Rai, 2015; Iacoboni et al., 2004; Lieberman, 2013; Rai & Fiske, 2011). Perhaps the most important evaluations of this sort are moral judgments about our own or others’ actions. One component of moral judgments consists in evaluating the events that actions brought about or are likely to bring about. These evaluations are best understood from a theory of evaluation independent from decision-making.

The following section presents such a theory of event evaluation. Similar to many theories of choice, this model is an idealization. Typically, behavioral economic theories of choice operate within idealized game-theoretic models (following Von Neumann & Morgenstern, 1947). These are meant to explain behavior in simple scenarios where options are limited and complete information is available to the decision-maker\(^1\). Often these are situations where choices are binary and expressed as monetary payouts. However, a satisfying theory of event evaluation cannot be so constrained if it is to capture moral evaluations and the evaluation of events generally.

---

\(^1\) This is the case for decision-making under risk. Other theories concern decision-making under uncertainty, where complete knowledge is not assumed. In either case, the situations under consideration are typically highly idealized (Johnson & Busemeyer, 2010).
The present theory is intended to offer a maximally general theory of event evaluation. Von Neumann and Morgenstern (1947) noted that examining decision-making and rational choice from the perspective of simple games limited the realism of their models, but argued this was a justifiable sacrifice in the name of quantitative accuracy. As the reader might expect, the theory proposed here will not escape trade-offs between generality and quantitative precision. Whereas game-theoretic models make quantitative predictions about highly idealized situations, the present theory can generate predictions for realistic events, with the limitation that these predictions are only qualitative. However, this general theory can also be extended to the simple scenarios of economic games. Given some additional quantitative treatment, the general theory can be used to make quantitative predictions in these situations. This will be illustrated in Section 3.1.

2.1 The Role of Expectations in Utility Evaluation: Existing Theories

Theories of choice commonly hold that evaluations of choice options are a result of comparison between a decision-maker’s present reference state and the imagined future states corresponding to different choice options. This idea (owing to Markowitz, 1952) is a crucial element of many descriptive theories of decision-making and choice, including prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), regret theory (Loomes & Sugden, 1982), and certain theories of intertemporal choice (e.g., Loewenstein & Prelec, 1992).

To illustrate, norm theory (Kahneman & Miller, 1986) emphasizes the role of counterfactual comparisons in evaluations (with counterfactuals based on expectations potentially among these). Famously, the salience of different counterfactuals has been invoked to explain the curious finding that Olympic Bronze medalists are often happier than Silver medalists (Medvec, Madey, & Gilovich, 1995; Medvec & Savitsky, 1997). It is thought that
Bronze medalists tend to compare their place against failing to medal and thereby feel victorious, whereas silver medalists tend to compare themselves to Gold, thereby feeling defeated.

Comparison offers a clear path for expectations to influence evaluations, as evaluators compare the obtained outcome to their prior expectation. Indeed, McGraw, Mellers, & Tetlock (2004) argue that Olympic Silver medalists’ disappointment relative to Bronze medalists is actually a function of comparison against expectations, rather than resulting from a direct comparison against the others on the podium. More generally, a number of theories have been proposed to explain how expectations shape evaluations, sometimes referred to as “disappointment theories” (e.g., Bell, 1985; Gul, 1991; Loomes & Sugden, 1986; Mellers et al., 1997). These theories incorporate a “disappointment function” that augments utility evaluations. Generally speaking, a disappointment function decrements or increments utility evaluations according to some comparison of obtained and expected outcomes. The most ambitious theory of disappointment is perhaps Gul’s (1991) disappointment aversion theory, which sought to supplant expected utility theory (Von Neumann & Morgenstern, 1947; Savage, 1954) as a rational theory of economic choice.

Existing disappointment theories have three major weaknesses. First, these theories are greatly limited in scope, focused only on explaining choice behavior in artificial gambling scenarios. Second, each of these theories rests on an essentially arbitrary and ad hoc augmentation of expected utility theory. Some theoretical advance is needed to account for the influence of expectations in evaluation and choice, yet it would be preferable if this modification is principled rather than arbitrary. Here I attempt to motivate the role of “disappointment” in evaluation through deeper consideration of the problem of event evaluation. Third, although there is strong empirical support for the qualitative prediction that expectations influence
evaluations, there is little quantitative support for any of the specific models proposed under these theories. In fact, Abdellaoui and Bleichrodt (2007) found that both Bell (1985) and Gul’s (1991) theories were unable to account for human decision-making behavior. In Section 3.1, I compare the contrastive theory of event evaluation to another disappointment theory, decision affect theory (Mellers et al., 1997), in more detail.

2.2. Accuracy and Error in Event Evaluation

The approach of the present theory differs somewhat from most behavioral economic theories of evaluation and choice. Researchers proposing theories of choice are often concerned with assessing whether those theories are *rational* (and even where theorists’ goals are merely descriptive, comparisons with rational choice are often important). Attempts to develop rational theories of choice often consist in positing a set of axioms and equations justified on some rough sense of prima facie plausibility. Choices are said to be rational—under the theory—if they obey these axioms. Theorists then engage in a meta-level analysis of the rationality of the axioms. Generally, if obeying the axioms produces intuitively irrational behavior (e.g., choosing a sure $5 payout over a sure $10 payout), then the theory is said to fail to provide a rational account of choice.

However, rationality is not an appropriate criterion to apply to a theory of event evaluation. What is rational depends on one’s purposes, and there is no single purpose inherent in the evaluation of events or actions for their own sake. If there is any goal at all, it is simply to be accurate in assigning value to different events.

Thus criteria for assessing the normativity of evaluations are more closely analogous to the criteria for assessing perceptual accuracy than for assessing the rationality of choice. In perception, the simplest notion of accuracy is just the correspondence between the world and its
perceptual representation. However, this is generally too demanding a criterion. For instance, consider the difficulty inherent in translating a 2D retinal imagine into a 3D object representation: perceptual representations will necessarily fail to accurately correspond to reality anytime a viewer is given limited views of a novel object. Consequently, this strict sense of accuracy is not particularly interesting, as it can be immediately known that human perception will fail to meet this standard. To be of use, a notion of perceptual accuracy would have to take into account the inherent limits of the information available to the perceptual system. This is the goal of ideal observer analysis (for a review see Geisler, 2003).

The central component of ideal observer analysis is the “ideal observer.” This is a theoretical device imagined to perform optimally in a (typically perceptual) task based on the available information. When there is uncertainty in a perceptual or cognitive task, even an ideal observer can err. In a visual task, an ideal observer performs at the statistical or physical limits of performance in a given situation. One use of ideal observer analysis is to identify what types of errors would result purely from uncertainty in the task problem. This allows for an analysis of human errors that differentiates between errors resulting from human perceptual or cognitive processing and those resulting from inherent uncertainty in the task. Ideal observer analyses are a type of computational-level theory (Marr, 1982)—theories that attempt to give formal descriptions of different cognitive or perceptual problems. In addition to their role in perception research, computational-level theories have been used to understand human cognition in many domains, such as memory (e.g., Anderson & Schooler, 1991), learning (e.g., Cheng, 1997; Lu et al., 2008; Powell et al., 2016), and knowledge (Powell et al., 2015).
It should be noted at the outset that this theoretical account will not constitute a true ideal observer analysis of event evaluation.² Still, the goal here is roughly analogous to the goal of ideal observer analysis: to consider how events might be evaluated by an idealized evaluator without memory or processing limits. Furthermore, we can assume this evaluator behaves optimally in whatever cognitive respects are relevant: e.g. that they learn by optimally incorporating new information with prior beliefs, that their prior beliefs are well justified by their previous experiences, and so forth. Essentially, we allow ourselves to imagine an evaluator that is idealized in whatever ways we like, save for the limits of their knowledge of the world. The result of this approach will be similar to a computational-level theory in that it gives one potential description of the cognitive problem faced by an evaluator. However, there is no attempt here to guarantee that this analysis provides a true or best description of the computational problem. Instead the goal at present is more modest: to provide one sensible treatment of this problem.

As will be shown, this analysis leads to the conclusion that the prior probability of an event will affect its evaluation. Although upon reflection we may eventually conclude that the influence of expectations results in evaluative errors, this analysis suggests that these errors are not cognitive in nature (e.g., they do not appear to result from forgetting or inaccurate estimation per se). Instead they seem to be an inherent consequence of performing event evaluation through comparison.

² A satisfying description of an ideal observer would need to be capable of precisely (and quantitatively) describing the observer’s behavior in a given situation. This will not be possible here, both because of the difficulty of specifying the evaluator’s knowledge of the world and of specifying a utility function defined over realistic inputs. In addition, the assignment of utility is inherently subjective, so that no normative standard has been rationally established. Therefore, human accuracy and error cannot be judged in the traditional sense, as a comparison between an ideal observer’s and human’s performance.
2.2.1 Evaluation of Events by Contrasting Uncertain States of the World

There are at least two crucial components to this theoretical approach: the information available to the evaluator and the procedure by which the evaluator translates that information into an evaluation of the event. The information available to the evaluator is determined by the event representations over which evaluative process operates. One way to represent events is to represent them globally, as changes in the state of the world. This global representation ensures that the maximum possible information is used in evaluating an event, making the evaluation as comprehensive as possible. Representing events in this fashion, the evaluation of an event can be made through comparison between a prior state of the world (S) and the evaluator’s final event-induced state (S’) (eq. 1).³

\[
V(\text{event}) = E[U(S')] - E[U(S)]
\]

A state S is simply a state of affairs in the world—how things were, are, and will be. This might encompass any or all past, present, and future goings-on in the world. If we assume that the universe is deterministic, there is in truth one eventual set of occurrences that will obtain. However, the evaluator’s knowledge of this true state of the world is limited, so that they must consider the expected utilities E[U(S)] and E[U(S’)]. Depending on the circumstances we wish to model, we can imagine an evaluator with varying degrees of information about these world states. In actuality, people’s knowledge of the state of affairs in the world is extremely limited and uncertain. If we allow that the information available to the evaluator is similar to that experienced by people—even with perfect memory for all learning and all experienced events—

³ Note that the function V(x) is applied to estimate the value of an event, whereas a utility function U(x) is applied to the states of the world themselves. This is analogous to the labeling used by Kahneman & Tversky (1979) in their development of Prospect Theory, where a value function is applied over entire prospects and utility function applied to the potential outcomes within those prospects.
even our highly idealized evaluator will have largely incomplete knowledge of past, present, and future events. So long as we ignore the uninteresting case of an omniscient evaluator, there will minimally be limits to their knowledge of future goings on.

Though other representations may be possible, considering events globally as changes in the state of the world captures some important aspects of event evaluation. First, events occur in context: A person being poisoned to death is typically a very negative event, yet this might not be so in the case of euthanasia. Likewise, a sudden $1000 windfall will be very positive for most graduate students, but quite trivial for a billionaire. Second, the true utility of an event is not always immediately apparent. Imagine that Jim and Bob both receive life-saving liver transplants on Monday. These events would appear to have equal utility, but the true utility of an organ transplant is not in the procedure itself (which is likely quite unpleasant) but instead in the length and quality of life then lived by the recipient. With this in mind, if Jim is run over by a bus on Tuesday it will no longer seem as if the two transplants generated equal utility. Evaluating events by comparing states of the world allows contextual features of the event and uncertainty over the event’s future consequences to be modeled in a natural way—probabilistically as part of those states of the world. It might seem more pragmatic to instead focus on the resulting state of the world produced by an event. In decision-making, it seems rational to be concerned only with the final result of a decision. However, this fails to capture the value of events evaluated as events or of actions evaluated in terms of their consequences. Consider a company with a $1 billion valuation: we would hold the CEO in high regard if the company was a young startup formerly run out of a garage, but not if the company was an aging giant formerly valued at $10 billion.4

---

4 It might also seem more direct to model event evaluation by simply assuming some utility function $U(x)$ that applies over any event in question. Conceiving of event evaluation in this way is likely possible, but this representation does not appear to lend itself to capturing context and uncertainty. For instance, in order to evaluate events in their proper context, the “event” descriptions entering into this utility function will need to be just as (or
Viewing event evaluation in terms of comparison is also supported by behavioral economic theories of choice and descriptively supported by findings in judgment and decision making (for a review see Kahneman & Tversky, 2000). A key component of prospect theory is that options are evaluated as relative changes from a reference state (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). In addition, Loomes and Sugden (1982) proposed a similar interpretation of event evaluation as part of regret theory, though in practice their theory did not make much use of this conceptual framework.

Given their incomplete knowledge, we should expect that evaluators’ subjective evaluations of events are driven by what they learn from those events or how those events change their understanding of the state of the world. Events can be viewed as transmitting information to the evaluator. Thus, an important aspect of an event’s evaluation is the amount of information that it carries. Considered from the perspective of information theory (Shannon, 1948), the informativeness of an event is measured in the degree to which it reduces uncertainty. When an evaluator learns of an event (either through direct experience or through some form of communication), their uncertainty about S’ is reduced relative to S and a comparison of these states can be made to evaluate the utility of the event. Therefore, we can conjecture that the utility of an event in the world will be proportional to the information provided by the event.

One important principle of information theory is the concept of self-information, or “Surprisal.” By definition, the information carried by an event is determined by the event’s prior probability, so that low probability events carry more information than high probability events nearly as) complex as the state of the world representations. This could be a proper way to represent the evaluation of events “in a vacuum”—for instance, the value of $1000 without consideration for who possesses it, or when, or for what purposes. However, it is unclear whether this is a meaningful psychological notion. In any case, it seems clearly distinct from the evaluation of a particular event as considered here.
(Shannon, 1948). To develop the intuition, consider flipping a coin versus rolling a die: before we flip the coin and roll the die we have some uncertainty about what face will result on each. However, if we were to take a guess for the coin we would be right $\frac{1}{2}$ of the time, and only right about the die $\frac{1}{6}$ of the time. So, we stand to gain more information from observing the result of the die roll than of the coin flip.

The principle of self-information, together with our conjecture linking information and utility evaluations, suggests that evaluations of the utility of events may depend on the prior probability of those events: Low probability events will carry more information and will do more to reduce uncertainty about $E[U(S')]$ relative to $E[U(S)]$. Therefore, low probability events should have a greater psychological impact and should be given more extreme evaluations. Positive events should be seen as better when they are low as compared to high probability, and negative events should be seen as worse when they are low as compared to high probability.

### 2.2.2 Evaluating the State of the World

Establishing these predictions beyond conjecture requires defining a representation for these states of the world as well as for the evaluation of these world-states’ utilities. One way of defining a state of the world $S$ is to think of the world as a probability “experiment.” A probability experiment can be any hypothetical or real-world process that generates outcomes with uncertainty. Such an experiment can have many possible outcomes $w_i$ that together form a sample space $\Omega$. The exact nature of these outcomes depends upon the scope over which the sample space is defined or the “world” being considered. If the experiment is a single roll of a die, there are six potential resulting outcomes: one for each face of the die. If the experiment is a series of four rolls of a die, then each outcome might be represented as a vector of length four (with each element representing a roll) and there are $6^4$ total outcomes. The sample space is also
affected by what is known—if we know (for whatever reason) that the third roll of the die will be a one, then only those outcomes with a one in the third position can have non-zero probability, leaving $6^3$ total outcomes in the space. Finally, note that each of these outcomes is mutually exclusive—only one die face (or series of faces) will result once the die is rolled.

The world of the probability experiment depends on the computational problem we wish to model. Since we assume the evaluator faces no limits to memory or processing capacity, a fully idealized “state of the world” representation would encompass all the goings on in the entire universe until its eventual heat death. An outcome in this space might be represented as a (likely infinite) vector describing the occurrence or non-occurrence of all possible goings-on in the universe. Once again, the particular outcomes included in the space are determined by the evaluator’s knowledge. If the evaluator knows that it rained in London on Tuesday, March 24th in 1942, then only those outcomes consistent with this fact should remain in the sample space. In addition, each of these potential outcomes are again mutually exclusive—when we reach the end of the universe there will be some final outcome that actually obtained.

More generally, how we define the state of the world will depend on our purposes. More limited accounts would see “the world” constrained to some set of relevant agents and occurrences: perhaps events involving human beings in the next 1,000 years, those living in Los Angeles in the next 100 years, the life of one individual person, and so forth. In any case, the nature of the outcomes in $\Omega$ is determined by the definition of the world.

For any sample space and any outcome description (e.g., a single value, a vector of values, or something else), I assume there is some utility function $U(w)$ that maps each outcome in the sample space to some real-valued number representing the utility that would result if that
outcome were to obtain.\textsuperscript{5} A state of the world $S$ describes some uncertain probability experiment with potentially many possible outcomes. The expected utility of some state of the world $S$ is just the expected utility of the outcomes in the corresponding probability space $\Omega$.\textsuperscript{6} For simplicity in the description here, I will assume the sample space of $S$ is discrete, although these calculations could be readily extended to continuous sample spaces. For $S$ defined by a discrete $\Omega$ with $n$ outcomes, the expected utility of $S$ can be calculated as eq. 2:

$$E[U(S)] = E[U(\Omega)] = \frac{1}{n} \sum_{i=1}^{n} U(w_i)$$

Recall that the value of an event is computed as the difference between the state of the world before and after the event. Before some occurrence we say the world is in state $S$ and that afterwards it is in state $S'$. The subjective value $V$ of an event or happening in the world $h$ is then evaluated as the difference in the utilities of $S'$ and $S$, as eq. 3:

$$V(h) = E[U(S')] - E[U(S)]$$

The probability space corresponding to $S'$ is a conditioning of the probability space $S$ on an occurrence in the world. Figure 2 provides a visual representation of the result of conditioning a space on an event or on an occurrence in the world. The sample space of $S'$—$\Omega'$—is a subset of $\Omega$, containing all $w_i$ in $\Omega$ consistent with the occurrence in the world (eq. 4):

$$\Omega' = \{w \in \Omega | w \text{ is consistent with event in world}\}$$

\textsuperscript{5} Note also that this utility function is assumed to be continuous (following Von Neumann & Morgenstern, 1947). Therefore, $P(U(w_i) = 0) = 0$.

\textsuperscript{6} Of course it should immediately be clear that utility calculations over infinite (or even very large) probability spaces are not psychologically plausible. If our purpose is to model human evaluations, we will likely have to constrain the world being represented considerably. As will be shown, however, the general principles derived from this account will apply no matter how “the world” is specified. Section 2.2.2.1 illustrates the operation of the model using a toy example centered on a world concerning one agent and one event of interest.
Figure 2. Euler diagram for sample space of probability experiment generating one random letter from English alphabet. Each outcome in the space is a letter and events in the space are sets of letters. Here we imagine we have learned that the letter generated is a vowel. If we condition $\Omega$ on this event we are left with $\Omega'$. We can also define an event $x_h$ within the original space of $\Omega$ that picks out the outcomes consistent with the event in question—namely, all the vowels. Its complement, $\sim x_h$ is the remainder of the space—namely, all the consonants.

This same utility calculation can also be expressed with respect to an event or set of events in the probability space of state $S$. Note that “event” in the language of probability theory refers to a set of outcomes within a sample space. For the remainder of this section I will use “event” exclusively in this fashion. As above, the expected utility of an event $x$ (a set of $n$ outcomes $w_i$) is calculated as eq. 5:

$$E[U(x)] = \frac{1}{n} \sum_{i=1}^{n} U(w_i)$$

The total expected utility of a state $S$ can be expressed in terms of the union of the events $x_1, x_2, \ldots, x_n$ defined in that space along with the complement of this union (note that the second term is not needed if the first set of events are exhaustively defined in the space) (eq. 6):
\[
E[U(S)] = E[U(x_1 \cup x_2 \cup \ldots \cup x_n)] + E[U(\neg(x_1 \cup x_2 \cup \ldots \cup x_n))]
\]

This allows us to express the utility of a state of the world in terms of whatever events within that state may be of interest to us.

Most straightforwardly, we can evaluate the utility of an occurrence in the world by defining a single event \(x\) within the space that somehow corresponds to an occurrence in the world whose value we wish to evaluate. The expected utility of some state of the world \(S\) can then be expressed in terms of an event \(x\) and its complement \(\neg x\), as eq. 7:

\[
E[U(S)] = E[U(x)]P(x) + E[U(\neg x)](1 - P(x))
\]

Similarly, recall that \(S'\) is the result of conditioning \(S\) on some occurrence in the world \(h\). Therefore, the expected utility of \(S'\) can be expressed in terms of these two events after conditioning \(U(x)\) and \(P(x)\) on \(h\), as eq. 8:

\[
E[U(S')] = E[U_h(x|h)]P(x|h) + E[U_h(\neg x|h)](1 - P(x|h))
\]

The general equations expressed above are not particularly useful unless some values are known for each term. What is needed is a particular specification of \(x\) that will allow the equations to be combined in a simple way. Fortunately, there is such a specification that aligns perfectly with our goals of evaluating the utility of some occurrence in the world \(h\): We can define an event \(x_h\) in exact correspondence to that occurrence in the world, as the set of all outcomes consistent with \(h\) or those with non-zero probability given \(h\). Connecting this with our previous definition, the event \(x_h\) contains the same outcomes as were contained in \(\Omega'\). That is, \(x_h = \Omega'\). As a result of its definition, it is ensured that the outcomes in \(x_h\) are the same in \(S\) as in \(S'\), so that \(U(x|h)\) is equal to \(U(x)\). If we further assume that the evaluator’s utility function is unaffected by the occurrence in the world, then the conditional utility evaluation \(U_h(x)\) is equal to
Under this definition of $x_h$ and this assumption, $U_h(x_h|h) = U_h(x_h) = U(x_h)$. Therefore, substituting equations 7 and 8 into equation 3 for $x_h$, we find that the subjective value of some event in the world can be calculated, as eq. 9:

$$V(h) = E[U(S')] - E[U(S)]$$

$$= \left( E[U(x_h)]P(x_h|h) + E[U(\sim x_h)](1 - P(x_h|h)) \right)$$

$$- \left( E[U(x_h)]P(x_h) + E[U(\sim x_h)](1 - P(x_h)) \right)$$

From here, some simple algebra (obscured only by verbose notation) allows this equation to be further simplified, as eq. 10:

$$V(h) = \left( E[U(x_h)]P(x_h|h) + E[U(\sim x_h)] - E[U(\sim x_h)](P(x_h|h)) \right)$$

$$- \left( E[U(x_h)]P(x_h) + E[U(\sim x_h)] - E[U(\sim x_h)](P(x_h)) \right)$$

$$= E[U(\sim x_h)] - E[U(\sim x_h)] + E[U(x_h)]P(x_h|h) - E[U(\sim x_h)](P(x_h|h))$$

$$- E[U(x_h)]P(x_h) + E[U(\sim x_h)](P(x_h))$$

$$= E[U(x_h)]P(x_h|h) - E[U(\sim x_h)](P(x_h|h)) - E[U(x_h)]P(x_h)$$

$$+ E[U(\sim x_h)]P(x_h) = (E[U(x_h)] - E[U(\sim x_h)]) \ast (P(x_h|h) - P(x_h))$$

Finally, we arrive at the theoretical result we set out to demonstrate initially. Recall that the outcomes in $x_h$ are consistent with $h$ by definition, so that $P(x_h|h)$ is necessarily equal to one.

---

7 This assumption clearly seems appropriate when considering the evaluations of some idealized evaluator. Moreover, this assumption appears reasonable in most cases of prospective and retrospective evaluation by a human evaluator: In the case of retrospective evaluation, the evaluator’s utility function is conditioned on the occurrence both when evaluating $S'$ and when looking back to evaluate $S'$. In the case of prospective evaluation, the anticipatory utility function is unconditioned for $S'$ as it is for $S$.

8 The two larger terms are first broken out into their components so that common terms may cancel in step 2 and finally may be factored in step 3. Note also that the resulting in an equation that not at all unlike the risk or loss functions employed in Bayesian decision theory (e.g., Smith, 2010).
Therefore, it can be shown that the value of an occurrence in the world $h$ is inversely proportional to the prior probability of the corresponding event $x_h$ in $S$ (eq. 11):

$$V(h) = (E[U(x_h)] - E[U(\sim x_h)]) \* (P(x_h|h) - P(x_h))$$

$$= (E[U(x_h)] - E[U(\sim x_h)]) \* (1 - P(x_h)) \propto 1 - P(x_h)$$

From this proportionality we can conclude that an evaluator’s evaluation of an event will be affected by the evaluator’s expectations about that event.

To illustrate the model’s operations within a realistic example, suppose that $S_a$ is the state of the world June 28th, 1914, moments before Austrian Archduke Franz Ferdinand was assassinated and suppose further that $S_a'$ is the state of the world immediately after his assassination. Each outcome $w_i$ in $\Omega_a$ is a vector representing all relevant goings on in the world, among them, we will assume, the assassination of Ferdinand represented as a binary 1 or 0 at index $k$. In $\Omega_a$, there were outcomes where Ferdinand was assassinated and outcomes where he was not—i.e., in $\Omega_a$ there were vectors $w_i$ with both 1 and 0 at the $k^{th}$ index. After the occurrence, $S_a$ is conditioned on Ferdinand’s assassination, so that $\Omega_a'$ contains only those outcomes where Ferdinand was assassinated (i.e., only those $w_i$ with 1 at the $k^{th}$ index). Likewise, represented as events in $S_a$, $x_h$ is the set of these same outcomes with 1 in the $k^{th}$ index and $\sim x_h$ the set of all other outcomes in $S_a$ (those with 0 in the $k^{th}$ index).

An evaluator’s knowledge not only determines the possible outcomes in $\Omega$, but also the structure of those outcomes, or the conditional dependencies among potential events within the space. Ferdinand’s assassination would eventually lead to the start of World War I and consequently to the deaths of tens of millions of people. This can be represented in the outcomes of $S_a'$, where a greater proportion of these outcomes include the events of WWI (call them $x_{ww}$) than do the outcomes of $S_a$. Therefore, a knowledgeable evaluator would recognize that $P(x_{ww} |$
\( x_h \) > \( P(x_{ww}) \). Of course, if an evaluator were ignorant of this fact, they would represent the outcomes in \( S_a \) and \( S_a' \) differently—perhaps so that Ferdinand’s assassination and the start of WWI were independent events, so that \( P(x_{ww} \mid x_h) = P(x_{ww}) \).

Thus the evaluation of events depends greatly upon the evaluator’s knowledge and their representations of \( S \) and \( S' \). Given simply that it was a case of murder, it seems clear that Ferdinand’s assassination produced negative utility, so that \( E[U(S_a')] \) will be less than \( E[U(S_a)] \). If the assassination is evaluated without awareness of its geopolitical impact, then this might be the end of the story. However, if the assassination is evaluated with the benefit of hindsight, it will be seen to have increased the probability of WWI, and should be evaluated as having much greater disutility.

We can also consider two evaluators with similar understandings of geopolitics but different expectations for Ferdinand’s safety. That is, both evaluators recognize that Ferdinand’s assassination would likely set off a world war, but one thinks his assassination is exceedingly unlikely whereas the other fears it is quite likely. These evaluators hold very different views of \( S \)—to the first evaluator the world seems at peace, to the second evaluator the world rests on a knife’s edge of violence and war. If Ferdinand is assassinated and war breaks out, the first evaluator will experience a massive change in their understanding of the state of the world. Consequently, they will evaluate the event very negatively. Meanwhile, the second evaluator will only see their fears confirmed—the result will be terrible, but in their mind it will be a terror that was always present. Their evaluation, though still negative, might therefore be less extreme.

Finally, it is important to stress that the connection between an event’s prior probability and its evaluation is one of proportionality: a low probability event will be given a more extreme evaluation than a higher probability event only under the assumption of all other factors being
equal. There may be instances where evaluators’ perceptions of probabilities do not lead to these patterns of evaluation, especially when evaluations are compared among different evaluators.

To illustrate, consider a comparison of optimists and pessimists. Optimists might be expected to place high probabilities on positive events and low probabilities on negative events. Conversely, pessimists might place high probabilities on negative events and low probabilities on positive events. If so, then optimists should give more extreme evaluations to negative events and weaker evaluations to positive events as compared with pessimists. This might lead one to question how optimists and pessimists retain their dispositions, if optimists are always more gravely disappointed by negative events and pessimists are always more pleasantly surprised by positive events. Assuming all else is equal between optimists and pessimists, we should predict this counterintuitive pattern of evaluation. However, this assumption seems quite likely to fail in this case, as evaluators who are optimistic or pessimistic might also be expected to assign different utilities to events. Though we expect both optimist’s and pessimist’s evaluations of an event to be proportional to their perception of its prior probability, this point may be outweighed by even greater differences in their utility functions.

2.2.2.1 Numerical Example

Here we consider the numerical calculation of event utilities within a toy example. Imagine Bob works at a company that is undergoing restructuring next week. Bob is hoping that he will be promoted out of his current position to the position of regional manager. In any case, Bob is going to be assigned a new job position title, although at his company any title that doesn’t include “manager” is essentially meaningless. Before the week begins, Bob is in state $S_1$ with possible outcomes as described in Table 1. As before, each of these outcomes is assumed to have equal probability in $\Omega$. 

27
Table 1. Toy example: sample space for hypothetical state of the world $S_1$.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>New Title</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Administrative Manager</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>Administrative Specialist</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Records Coordinator</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>Services Administrator</td>
<td>5</td>
</tr>
</tbody>
</table>

As shown in Table 1, there are four possible outcomes in $\Omega$ and $E[U(\Omega)] = \frac{1}{4}(25 + 5 + 5 + 5) = 10$. Bob will either be promoted to management if his title is changed to “Administrative Manager” or he will essentially remain in his current position if it is changed to anything else. Suppose that outcome 1 obtains so that it is the only outcome in $S_1'$. We can calculate the value of this event as eq. 13:

$$V(h) = E[U(S')] - E[U(S)] = E[U(w_1)] - E[U((w_1, w_2, w_3, w_4))] = 25 - 10 = 15$$

If instead outcome 4 obtains so that it is the only outcome in $S_1'$ we can also calculate this value (eq. 14):

$$V(h) = E[U(S')] - E[U(S)] = E[U(w_4)] - E[U((w_1, w_2, w_3, w_4))] = 5 - 10 = -5$$

Bob will experience +15 utility if he is promoted and −5 utility if he is not promoted.

Next we can consider how Bob’s expectations will affect his evaluations. Suppose that rather than starting in $S_1$, Bob instead begins in state $S_2$ as shown in Table 2.

Table 2. Toy example: sample space for hypothetical state of the world $S_2$.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>New Title</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Administrative Manager</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>Records Manager</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>Coordinating Manager</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>Administrative Specialist</td>
<td>5</td>
</tr>
</tbody>
</table>
Again there are four possible outcomes in $\Omega$ but whereas in $S_1$ being promoted to management was unlikely, in $S_2$ it is expected. Therefore, here $EU(\Omega) = 20$. Again we can calculate the values of obtaining outcomes 1 and 4 (eq. 15 and eq. 16):

\[
V(h) = E[U(S')] - E[U(S)] = E[U(w_1)] - E[U\{(w_2, w_3, w_4)\}] = 25 - 20 = 5
\]

\[
V(h) = E[U(S')] - E[U(S)] = E[U(w_1)] - E[U\{(w_2, w_3, w_4)\}] = 5 - 20 = -15
\]

Here Bob will be much less happy if he is promoted and much more disappointed if he is passed over for the promotion. This same point can be illustrated by calculation over events. For instance, if we define $h_1$ as Bob being promoted, the value of $h_1$ if Bob starts in $S_1$ is (eq. 17):

\[
V(h_1) = V(x_{h_1}) = [EU(x_{h_1}) - EU(\sim x_{h_1})] \ast (1 - P(x_{h_1})) = (25 - 5) \ast .75 = 15
\]

In contrast, if we define $h_2$ as Bob being promoted when he starts in $S_2$, the value of $h_2$ is (eq. 18):

\[
V(h_2) = V(x_{h_2}) = [EU(x_{h_2}) - EU(\sim x_{h_2})] \ast (1 - P(x_{h_2})) = (25 - 5) \ast .25 = 5
\]

### 2.2.3 Estimating Utilities and Probabilities

Computationally, evaluating the utility of a state of the world requires evaluating the utility of all possible outcomes in the sample space representing that state. This will generally not be possible for human evaluators. Computing utilities of any realistic description of a state of the world is likely to lie well beyond the limits of human cognitive capacity due to both memory and processing limitations. To be sure, there is no requirement that human evaluations proceed by considering events globally as changes in the state of the world. Still, we may wish to know whether a psychologically plausible account could resemble or approximate this method of evaluation. More specifically, we may wonder whether differences between the idealized method of global comparison and human evaluations owe simply to human cognitive capacity limits or whether they reflect more fundamental differences in the evaluation processes involved.
Two things are required for an implementation of this account to be psychologically plausible: people must be able to estimate both utilities of states of the world and the prior probabilities of events. More specifically, the utility of an event can be evaluated so long as an evaluator is able to form estimates of the terms in equation 11: \[\text{EU}(x_h) - \text{EU}(-x_h)\] and \(P(x_h)\). Importantly, there is no need for the utilities of states of the world to be calculated exactly. Instead, they can be estimated, with the precision of this estimate determined by the cognitive limits of the evaluator.

Likewise, the psychological plausibility of this model requires that people have some awareness of the probability of the events they are evaluating, but this is a fairly minimal requirement: this awareness need not be explicit, nor need it be perfectly accurate (indeed, it need not be accurate at all, though this will necessarily lead to errors). It is implausible that people have stored representations in their mind for the probability of all or even many of the possible events that might occur to them. Rather, it seems more likely that these probabilities are constructed on-the-fly by some type of generative estimation process.

Where people’s subjective expectations do not accurately reflect the true prior probabilities of events, people’s evaluations of event utilities will be inaccurate or distorted. It has been widely observed that peoples’ representations of probability are distorted (e.g., Gonzalez & Wu, 1999), a point that is typically modeled according to various probability weighting functions. Such a function can easily be incorporated into a psychological extension of this idealized theory of evaluation, as discussed in Section 3.1.

Nevertheless, there is also good reason to believe that people’s prior expectations are a largely accurate reflection of true prior probabilities in many domains. Prior expectations are an essential component of Bayesian models of human cognition (for a review see Griffiths, Kemp,
Although the plausibility of these models is debated (e.g., Jones & Love, 2011), there is good evidence that people are able to estimate prior probabilities in a variety of domains. For instance, there is evidence that perception (e.g., Yuille & Kersten, 2006) and memory (e.g., Anderson & Schooler, 1991) processes are aided by largely accurate probabilistic models of the environment, allowing human cognition to approximate optimal statistical inference in these domains. In addition, although errors are often produced by “base rate neglect” when people reason explicitly about probabilities (e.g., Kahneman & Tversky, 1973; Tversky & Kahneman, 1974), people show great sensitivity to base rates in more natural reasoning contexts (e.g., Griffiths & Tenenbaum, 2006; Gigerenzer & Hoffrage, 1995), and even very young children use base-rate information to aid in inference (e.g., Kushnir, Xu, & Wellman, 2010; Téglás et al., 2011) and language learning (e.g., Xu & Tenenbaum, 2007).

To illustrate, Griffiths and Tenenbaum (2006) asked participants to make estimates about a variety of topics, such as how much longer a 60 year-old man would live, how much a movie would gross in total if it had grossed $40 million after one week, and how much longer a cake baking for 35 minutes should be left in the oven. People’s estimates of these quantities were well explained by a Bayesian model incorporating accurate prior expectations. That is, people appear to have good knowledge of or be able to accurately construct the true distributions of base-rates for a variety of everyday domains. Even in cases where people’s estimates were inaccurate, they appeared to rely on priors with the same functional form as the true distribution of base-rates.

2.3 Extending Event Evaluations to Moral Judgments of Actions

Recall that under utilitarianism the moral status of an action is determined by the utility of its consequences. In particular, under act utilitarian rendering a moral judgment about an action just means evaluating the utility of the events that action produced. However, human
moral judgments are not so simple. Instead, they appear to incorporate a variety of non-utilitarian considerations.

Often these non-utilitarian considerations reflect important concerns within deontological ethics. For instance, people’s moral judgments seem to be based partly on considerations of people’s rights and corresponding duties. Deontological ethics typically assign positive rights (e.g., to life) that imply corresponding negative duties (e.g., a duty not to actively kill). In accord with these rights and duties, people view actively harming another (violating a duty) as morally worse than passively allowing harms to occur, even when the harm was foreseen (Borg et al., 2006; Cushman et al., 2006, 2012; DeScioli et al., 2011; Shultz et al., 1981; Spranca et al., 1991; Baron & Ritov, 2004; 2009). In addition, people are often more concerned with the intention behind an act than its ultimate consequences. In fact, intentional but unsuccessful attempts to harm are often judged more harshly than accidental harms that actually occur (Moran et al., 2011; Young & Saxe, 2009; Young et al., 2007). Likewise, people generally feel it is worse to do harm intentionally than as an unintended but foreseen consequence (Borg et al., 2006; Greene et al., 2009; Hauser et al., 2007; Moore et al., 2008; Young et al., 2007; Young & Saxe, 2009, 2011) and an agent’s ill intentions can magnify the perceived harmfulness of their actions (Ames & Fiske, 2013; 2015).

Moral judgments are also concerned with and affected by interpersonal relationships. For instance, fairness is an important moral value (e.g., Crockett et al., 2013; Haidt, 2007; Graham, Haidt, & Nosek, 2009; Turiel, 1983) that is recognized from an early age (e.g., Bloom, 2013; Sloan, Baillargeon & Premack, 2012). However, the evaluation of fairness can depend upon the relationships between agents: rules about fairness and equality are relaxed when one person is in a role of authority (Hoffman et al., 1994; Rai & Fiske, 2011). Similarly, when asked to consider
sacrificial actions, people are much less likely to approve when the victim to be sacrificed is a family member, friend, or other ingroup member (Bleske-Rechek et al., 2010; Cikara et al., 2010; O’Neill & Petrinovich, 1998; Petrinovich, O’Neill, & Jorgenson, 1993; Swann et al., 2010; Uhlmann et al., 2009).

The assignment of utility to events is sure to depend on people’s culture and their individual preferences. However, in some cases people seem to hold “protected” or “sacred values” for which they are unwilling to make any kind of consequential trade-offs (Fiske & Tetlock, 1997; Tetlock, 2002; Baron & Ritov, 2004; 2009). Baron & Spranca (1997) found that certain protected values display “quantity insensitivity.” They found that participants were indifferent to the consequences produced by actions sacrificing these values, suggesting that people are unwilling to properly assign these values utility or perhaps even assign them infinite utility (but see Bartels & Medin, 2007; Baron & Leschner, 2000, for evidence of sensitivity to tradeoffs).

With the above considerations in mind, caution must be exercised in applying a theory of event evaluation to the moral judgment of actions. Still, as reviewed in Section 1.2, there is ample evidence that utilities affect moral judgments in many if not most situations. I propose that, all else being equal, the moral judgment of an action $M(a)$ will be proportional to the evaluation of the events $h$ brought about by that action $a_h$ (eq. 19):

$$M(a_h) \propto V(h) = [EU(x_h) - EU(\sim x_h)] * (1 - P(x_h))$$

For example, it is clear that physically assaulting another person is morally wrong due to the violation of a moral rule and the rights of the victim. Beyond that however, its severity should also be judged in proportion to the disutility it creates, as reflecting the physical and
emotional harm that it causes to its victim. Generally speaking, an assault that bruises someone is wrong, but an assault that breaks their bones is worse.

Finally, it is also worth considering how the utility function applied by an evaluator is subjective to the perspective of that evaluator. In the tradition of egalitarianism, or in accord with Rawls’ (1971) proposal that moral judgments be made from behind “the veil of ignorance”, we might imagine a Rawlsian evaluator who assigns utility to outcomes as an independent and impartial observer. Identifying the particular utilities entailed by this stance would require us to cast ourselves in this same role as the evaluator as best we can. However, we can also consider evaluators as moral agents with their own culturally-determined and subjective preferences and with feelings toward and relationships with other agents. In this case, we might be able to begin to account for some of the seemingly non-utilitarian tendencies evinced by human moral judgments. For instance, an agent could assign greater utility to the lives of friends, family, and other in-group members (e.g., Portmore, 2011). In addition, this perspective could potentially provide an account of sacred or protected values if these quantities are given extreme utility.
3. Studies

The preceding section advanced a theory of event evaluation according to which events are evaluated globally as changes in the state of the world. Here, the predictions of this theory are examined in four different studies, in both controlled experiments and in naturalistic settings. Study 1 demonstrates a quantitative extension and fitting of the model predictions to existing data describing participants’ emotional reactions in a monetary gambling task. The three remaining studies present novel empirical findings. Study 2 examines the role of endogenous expectations in event evaluation in a naturalistic setting. I examined football fans’ reactions to wins and losses using Twitter activity following games in the 2015-2016 National Football League season. Studies 3 and 4 turn to evaluations in the moral domain. Study 3 examines participants’ explicit judgments of immoral actions against victims at different levels of risk. Finally, Study 4 is a naturalistic study examining how prior expectations of risk affects people’s moral reactions to terrorism events. Global Twitter activity was measured immediately before and after 122 distinct terrorism events. Twitter activity following terrorism events corresponded to the frequency of terrorism events in the affected countries, supporting the role of expectations in these reactions.

3.1 Study 1: Modeling of Data from Mellers et al. (1997)

Mellers et al. (1997) used a simple gambling task to show that people’s affective responses to monetary gains and losses are influenced by their expectations. In this study, the global contrast theory of evaluation was extended to generate quantitative predictions in a simple gambling task utilized by Mellers et al. (1997). This was accomplished by defining states of the world corresponding to the experimental task and by incorporating quantitative assumptions about the utility function and the representation of probability. The result of this extension will
be referred to as the “global contrast model of evaluation” although the quantitative specifics of this model are independent from the theoretical perspective advanced above.

3.1.1. Mellers et al. (1997) Experiment and Results

Mellers et al. (1997) conducted an experiment examining the effects of people’s expectations on their emotional reactions to events. Participants engaged in a simple gambling task, where they had a probability to win or lose a certain amount of money on each trial of the experiment. For all trials the alternative outcome was a gain of zero dollars. For instance, one trial might give participants a 17% chance to win $17.50 or else to win nothing with a probability of 83%. To aid participants’ comprehension, probabilities were represented visually with a pie chart. An example trial is shown in Figure 3. The task was realistic: participants were told they would actually keep the money that they won (or owe the money that they lost). Following the result of each gamble, participants rated how they felt on a scale from -50 (extremely disappointed) to +50 (extremely elated).

![Example of experimental trial display from Mellers et al. (1997). Reprinted with permission.](image)

Mellers et al. (1997) found that participants’ emotional responses were affected not only by the outcomes of each trial, but also by their initial probability of winning or losing. For
instance, participants had a stronger positive reaction to winning $31.50 on a trial for which they initially had a 9% chance of winning than they had to winning $56.70 on a trial for which they initially had a 94% chance to win. Figure 4 shows participants’ average emotional responses to trials where they won or lost (left) and where the zero outcome obtained (right).

![Win/Loss obtained outcomes](image1)

![$0 outcomes](image2)

**Figure 4:** Results of Experiment 1 from Mellers et al. (1997). Reprinted with permission.

Unfortunately, the original data reported in Mellers et al. (1997) could not be made available to me as these records no longer exist (Mellers, personal communication). To remedy this, data were transcribed from a scanned image of the figures reported in the Mellers et al. paper. Some data points in these figures are difficult to read. To avoid bias, Figures 2 and 3 (here shown together as Figure 4) from Mellers et al. (1997) were stripped of axis and unit labels and transcribed by two hypothesis-blind coders. The coders used WebPlotDigitizer, a web-based tool for extracting numerical data from images of graphs using human input (http://arohatgi.info/WebPlotDigitizer/). If two points could not be distinguished clearly from the graphs, the coders were instructed to use the average of the two points for each. The two coders
transcribed the figures together and resolved any disagreements through discussion between themselves.

3.1.2 Generating Quantitative Predictions from the Contrastive Theory of Evaluation

Recall equation 11 from Section 2.2.2:

$$ V(h) = V(x_h) = [EU(x_h) - EU(\sim x_h)] \times (1 - P(x_h)) $$

This general equation can be adapted to generate predictions for the task used by Mellers et al. (1997). To do so, I first assume that the emotional reactions obtained by Mellers et al. reflect participants’ underlying utility evaluations. Next, I assume a simple world representation where \( \Omega \) contains two outcomes: one outcome corresponding to winning or losing the gamble’s stake occurring with the gamble’s assigned probability \( p \), and another outcome of zero dollars occurring with probability \( q = (1 - p) \). Within this sample space I define the event \( x_h \) as corresponding to the result of the gamble, and its complement \( \sim x_h \) as the unobtained outcome. The expected utility of these events is just the subjective utility of each outcome, and its probability is \( p \) in the case of the winning or losing stake and \( q \) in the case of the zero outcome.

In order to generate quantitative predictions, it is necessary to specify the exact nature of the utility function and of the representation of probabilities. First, for simplicity, I assume the power functions \( U(x) = x^\frac{1}{2} \) for positive outcomes and \( U(x) = -|x|^\frac{1}{2} \) for negative outcomes. For gains, this produces a concave utility function similar to the log function classically proposed by Bernoulli (1738/1954), and for losses this produces a convex utility function. Together, these reflect the form of the utility function commonly observed in human judgments (e.g., Kahneman & Tversky, 1979). Here, a power function is preferred to a log function because it is properly defined for the zero outcome. Second, human probability representations have widely been shown to be distorted, typically following an s-shaped curve overweighting low probabilities and
underweighting high probabilities (e.g., Kahneman & Tversky, 1979; Gonzalez & Wu, 1999). To account for this pattern, a probability weighting function is added to the model. Here, I assume probabilities are weighted according to the one-parameter probability weighting function \( \pi(p) \) with free parameter \( \beta \) between 0 and 1 (Gonzalez & Wu, 1999), as eq. 20:

\[
\pi(p) = \left( \frac{p^\beta}{(p^\beta + (1 - p^\beta)^\beta)} \right)
\]

This yields the final model equation, eq. 21:

\[
V(h) = (E[U(x_h)] - E[U(\sim x_h)]) \ast (1 - \pi(x_h))
\]

The \( \beta \) parameter of \( \pi(p) \) was fit using a grid search to maximize the correlation between model estimates and predicted values, with the resulting parameter \( \beta = .55. \) Incidentally, this parameter is similar to the estimate of .56 obtained in a study by Camerer and Ho (1994). These estimates were then fit to the actual values obtained from the figures using linear regression (resulting in the addition of two parameters). This procedure allowed the residual sum of squares (RSS) of the model to be computed. From this information, the model’s Bayesian Information Criterion value (BIC; Schwarz, 1978) was also calculated.

To evaluate the fit of a linear model, the BIC can be calculated according to eq. 22:

\[
BIC = n \ast \ln \left( \frac{RSS}{n} \right) + k \ast \ln (n)
\]

In this equation, BIC is calculated based on the number of observations (\( n \)) and the model RSS, along with a penalty term based on the number of model parameters (\( k \)) and observations (\( n \)). Lower BIC values indicate better fit.
3.1.3 Results and Model Comparisons

Figure 5. Scatterplot of model predictions and human emotion ratings for all trials in experiment reported by Mellers et al. (1997).

The resulting model (shown in Figure 5) provides an extremely strong fit to the data from Mellers et al. (1997), $r = .959$, $R^2 = .919$, $BIC = 394.9$. This model also captures important qualitative aspects of the data, such as the rank-orderings of certain expected/unexpected win and loss pairs. For instance, it predicts that winning $31.50 with .09 probability will be more elating than winning $56.70 with .94 probability—accurately reflecting the ordering in the human data.
It is valuable to contextualize this model fit by comparing it to the fits provided by alternative models. First, I consider a simple model focused only on obtained outcomes. Specifying $U(x)$ as before, this simple model estimates the value of a gamble outcome by eq. 23:

$$V_g(h) = U(x_h)$$

This model provides a reasonable but comparatively weaker fit, $r = .863$, $R^2 = .745$, $BIC = 504.9$. Although it fits the data acceptably for wins and losses, the simple model does very poorly when estimating people’s reactions to zero outcomes. A comparison of BIC clearly favors the global contrast model over this simple model, $\Delta BIC = -110.0$. (Any reduction in BIC greater than 10 is conventionally considered very strong evidence in favor of the better-fitting model; Robert & Rafferty, 1995.)

Next I consider a model based on counterfactual comparison. Counterfactual comparison has been invoked as an important component of utility evaluation by a number of decision-making theories (e.g., Kahneman & Miller, 1986; Loomes & Sudgen, 1982). The counterfactual model assumes that evaluators compare the obtained outcome with the unobtained outcome, as eq. 24:

$$V_c(h) = U(x_h) - U(\sim x_h)$$

model provides a better account of data from zero outcome trials. Overall, the fit is quite good and significantly improved over the simple model, $r = .934$, $R^2 = .873$, $BIC = 435.1$. However, the global contrast model again offers a stronger fit, $\Delta BIC = -40.2$. This model also offers a superior qualitative fit, as discussed above.

### 3.1.4 Comparisons with Decision Affect Theory

Mellers et al. (1997) use their findings to provide evidence for their proposed decision affect theory (DAT), a theory of post-decision affect. One fundamental difference between these
two theories is that DAT conceives of affective reactions as distinct from utility evaluations of event, whereas the global contrast model of evaluation does not draw this distinction.

For a binary gamble with outcomes $a$ and $b$, the emotional reaction associated with outcome $a$, $R_a$ is defined in the DAT model as eq. 25:

$$R_a = a[u_a + g(u_a - u_b)(1 - s_a)] + b$$

In this equation, $u_a$ and $u_b$ are utility estimates of outcomes $a$ and $b$, $s_a$ is the subjective probability of $a$, and $g(x)$ the “disappointment function”, which reflects a comparison between the two outcomes. The parameters $a$ and $b$ are linear coefficients that are not of theoretical interest to the model, but can be fit using OLS regression. Thus the theoretically substantive portion of the equation can be expressed as the proportion (eq. 26):

$$R_a \propto u_a + g(u_a - u_b)(1 - s_a)$$

Mellers et al. (1997) report that the DAT model accounts for over 99% of the variance in their data. However, they also report their model as using 20 free-parameters to fit 100 data points. It is unfortunately not possible to compare their model’s fit with my own, as the data being described are not identical. Regardless, the number of parameters in their model suggest a serious threat to parsimony. In fairness, it is likely that a strong fit could be obtained while fitting fewer parameters. For instance, Mellers et al. (1997) estimated parameters for each utility value independently. Rather than being estimated individually, these values could almost surely be fit quite well using a single parameterized function (e.g., a power function). In any case, it remains somewhat unclear what would result might be obtained in a direct comparison.

Though they are not identical, the DAT proportionality expressed in 26 and equation 11 describing the global contrast model of evaluation clearly bear some resemblance. Despite this, the underlying theories differ quite dramatically. First, DAT is very limited in scope: DAT
specifically concerns affective responses and is defined only for binary and ternary gambling scenarios. Second, Mellers et al. (1997) never articulate the motivation behind the definition of DAT. Thus, although reasonably plausible, DAT suffers from the arbitrariness of other existing disappointment theories. In contrast, the model proposed here is derived from a highly-general analysis of event evaluation and was developed independently from the data being fit.

3.1.5 Discussion of Study 1

In Study 1, an extension of the global contrast theory of evaluation was used to quantitatively model participants’ emotional reactions to monetary gains and losses based on their exogenously determined expectations. Thus, Study 1 demonstrates that the general theory proposed in Section 2.2.2 can be readily applied to make quantitative predictions in simple game-theoretic situations. The idealized theory can also be augmented to account for various descriptive features of human cognition, such as distorted representations of probability. This highlights a virtue of this approach: the exact specification of the model can be made to trade-off between generality and quantitative specificity, as well as between idealization and psychological description.

3.2 Study 2: NFL Fan Reactions in Response to Wins and Losses as Measured by Twitter Activity

The behavior of sports fans is a classic case in which expectations seemingly influence evaluations. Intuitively, fans of a team are likely to be especially happy when their team wins as the underdog and especially disappointed when their team loses after being favored. This may also explain why people often choose to root for the underdog in contests where they do not otherwise have any personal stake, perhaps anticipating that it will be more exciting if they win in an upset (Bell, 1985).
This study investigates how fan reactions to wins and losses are influenced by their expectations by examining Twitter activity following every game in the 2015–2016 National Football League (NFL) regular season. Fan expectations were modeled based on the closing Las Vegas bookmaker’s money odds for each game. In prior studies examining evaluations of expected and unexpected events, researchers have established participants’ prior expectations exogenously, for instance by informing them of the probability of winning a gamble (Mellers et al., 1997) or of their risks of having an undesirable medical condition (Shepperd & McNulty, 2002). Here, fan’s expectations for winning or losing are determined endogenously, based on their understanding of the relative ability or performance of their team and their opponents. Although highly intuitive, the role of fan expectations in their reactions to wins and losses has not been empirically examined. This study constitutes a naturalistic test of the intuitive prediction.

3.2.1 Constructing the Dataset

There were 32 teams competing in the 2015–2016 NFL regular season, each playing a total of 16 games. A dataset was constructed so that each row of the dataset represents one team’s win or loss in a single week, resulting in a total of 512 data points. Final results for each game were acquired from a football statistics website (www.pro-football-reference.com).

For each of the 512 data points, the prior probability of that team’s winning in that week was calculated using the closing Las Vegas bookmaker’s money odds. These were recorded from a sports betting website (www.footballlocks.com). Bookmakers make their money by keeping a percentage of each bet (the vigorish, or vig). Consequently, they work to ensure that their financial success does not depend on the outcome of the game—they do not ever wish to be left owing more money to the winning betters than they are taking in from the losing betters.
Therefore, they attempt to estimate each team’s likelihood of winning a certain game in order to balance the amount of money being bet for each team. The opening odds represent this estimate, but these odds can be adjusted based on actual betting. Thus, the final closing odds represent a mix of expert opinion and crowd wisdom.

Moneyline odds are expressed in terms of positive or negative dollar amounts. For instance, in week one the Carolina Panthers were -150 favorites to beat the Jacksonville Jaguars, who were +130 underdogs. Due to Carolina’s superior odds of winning, a gambler would have to bet $150 on Carolina in order to win $100, but could win $130 while betting only $100 on Jacksonville. An alternative (and perhaps more familiar) form of betting is based not just on winning or losing but on a point spread. Here, teams must win or lose by a certain amount in order for a bet to pay out, with equal payouts on either side of the point spread. Point spread and moneyline odds are equivalent following appropriate transformations, leading to the same implied probabilities. The implied probability of winning can be calculated from the moneyline odds z using the following equation (eq. 27):

\[
P(\text{win} \mid z) = \begin{cases} 
-\frac{z}{z + 100}, & z < 0 \\
\frac{100}{z + 100}, & z \geq 0
\end{cases}
\]

Twitter activity was measured by querying the IBM Bluemix Twitter Insights database. Twitter Insights keeps a two-year running historical archive of the Twitter “Decahose”—a 10% random sampling of all tweets. The database was queried for @replies (tweets directed at an

---

9 The Las Vegas bookmakers’ moneyline odds include money spent toward the vig, which is not part of the wager itself, leading to a bias in the probability estimates. However, since this applies equally to both teams’ lines, this factor can be accounted for by normalizing each team’s probability of winning the game.

10 The creators of Twitter are partial to water metaphors: tweets can be streamed (collected in real-time), lists of tweet IDs can be hydrated (populated with full tweet information), and the full 100% stream of tweets is called the “firehose.”
individual user) directed at the official Twitter handle for each NFL team (e.g., @Packers, @Chargers, @STLouisRams) in the 24 hours following the end of each game. This 24-hour time period was estimated based on the kickoff time of each game and the average NFL game length, approximately 3 hours 12 minutes (Flint, 2012). The resulting dataset includes data from 139,840 tweets.

### 3.2.2 Results

Fans’ reactions were examined based on the number of tweets following each game. The number of tweets is taken to represent a non-valenced magnitude measure of reaction and corresponding psychological impact. The valence of the reaction must be inferred from context: in response to wins, greater tweeting represents stronger positive reactions; in response to losses, greater negative reactions. The number of tweets following each game was highly skewed, as shown in figure 6 (upper left). Similarly, the average number of tweets per game varied considerably among teams (Figure 6, upper right), likely reflecting differences in their popularity, the Twitter behavior of their fanbases, and possibly the management of their official Twitter account. To transform the number of tweets per game into a dependent variable that was not affected by these extraneous factors and was appropriate for OLS regression, a Twitter Reaction Score (TRS) was calculated as eq. 28:

\[ TRS = \ln \left( \frac{\text{tweets after game}}{\text{mean team tweets}} \right) \]

The resulting values are approximately normally distributed, as shown in figure 6 (lower).

These data were fit with a pair of linear regression models, as summarized in Table 3. The first model shows that winning or losing had no significant effect on the amount of tweeting. The second model represents the addition of the fan’s expectations in the form of variables for the prior probability of winning and the interaction between winning or losing and the prior
probability. Recall that prior probability is expected to have opposite effects in wins and losses, being negatively correlated with reactions to wins and positively correlated with reactions to losses. Inclusion of these factors leads to a significant improvement in prediction, $\Delta R^2 = .054$, $F(2, 508) = 14.70, p < .001$.

![Histograms](image)

*Figure 6.* Histograms for number of tweets per game (upper left), average number of tweets per game for each team (upper right), and Twitter reaction score following each game (lower).

<table>
<thead>
<tr>
<th>Model</th>
<th>Variables</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
<th>Model $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Win</td>
<td>.063</td>
<td>1.423</td>
<td>.155</td>
<td>.004</td>
</tr>
<tr>
<td>2</td>
<td>Win</td>
<td>.730</td>
<td>5.594</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P(win)</td>
<td>.204</td>
<td>3.192</td>
<td>.002</td>
<td>.058</td>
</tr>
<tr>
<td></td>
<td>Win $\times$ P(win)</td>
<td>-.797</td>
<td>-5.356</td>
<td>&lt; .001</td>
<td></td>
</tr>
</tbody>
</table>
Fan’s reactions to wins and losses were affected by their prior expectations of winning. Figure 7 shows a pair of scatterplots plotting Twitter reactions against win probabilities separately for wins and losses. As can be seen, the pattern of results is as predicted: prior win probabilities are negatively correlated with reactions for wins, \( r(254) = -0.250, p < .001 \), and positively correlated for losses, \( r(254) = 0.211, p = .001 \). Visual inspection of the scatterplot suggests there are no worrisome outliers. Subsequent regression diagnostics identified one potentially influential outlier (all other Studentized-deleted residuals \( t < 3 \)), but removing this outlier did not meaningfully affect the results.

![Scatterplots showing Twitter reaction scores separately for wins and losses.](image)

As this is a naturalistic dataset and there are many factors not accounted for in this model, the overall model fit is relatively modest \( (r = .242, R^2 = .058) \). One likely reason for this is the difficulty inherent in predicting the outcomes of football games. Even the bookmakers’ predictions correlate only modestly with the actual results of each game \( (r = .289) \). Although we must speculate somewhat as to the confidence of the sports fans, an idealized observer could be expected to have a rather large deal of uncertainty in their probability estimates.
3.2.3 Discussion of Study 2

As measured by their Twitter activity, NFL fans reactions are affected by their expectations of winning and losing. Study 2 demonstrates that people’s endogenous expectations about the prior probability of events affect their subsequent evaluations of those events. Study 2 thus provides a large naturalistic demonstration of the qualitative predictions of the global contrast theory of evaluation.

3.3 Study 3: Victims’ Prior Risk Affects Judgments of Immoral Actions

Thus far I have examined how prior probabilities affect event evaluations in the domain of monetary gains and losses and in the less tangible domain of sports fandom. These studies have added to existing evidence showing the influence of expectations in utility evaluations (e.g., Mellers et al., 1997; Shepperd & McNulty, 2002). This study shifts focus to the moral domain. There is good evidence that domain-general utility evaluation processes are employed in assessing moral utility (e.g., Rai & Holyoak, 2010) and that utility is an important consideration in people’s moral judgments (e.g., Greene et al., 2001; Hauser et al., 2007). This raises the question, how will people’s expectations about the victims of immoral actions influence their judgments of those actions? Will people judge it less wrong to bring about negative consequences for another person if the victim was independently at greater prior risk of suffering those consequences?

3.3.1 Experiment: Methods

This experiment examined how victims’ prior risk affects people’s judgments of immoral actions in a controlled experiment. Participants were asked to consider pairs of immoral actions taken against victims with high and low prior risk. For each pair, they were asked to judge which
of the two actions seemed worse using a five-point scale. Participants were given the option of indicating that neither action seemed worse using the midpoint of the scale.

3.3.1.1 Participants

A total of 111 participants (48 male, 59 female, median age = 31 years old) were recruited from Amazon’s Mechanical Turk work distribution website (mTurk). All participants were paid $1.00 for their participation.

3.3.1.2 Materials

Ten experimental items were created for this study (see Appendix). Each of these items describes a pair of identical actions taken against victims who should be expected to differ in their prior risk. One victim was at greater prior risk of suffering the negative consequence produced by the action. Victims’ prior risk was manipulated by changing the context in which the action occurred. For instance, one pair of items was: “Islamic terrorists set off a bomb on a bus in London that kills 10 and injures 24” (low probability) and “Islamic terrorists set off a bomb on a bus in Israel that kills 10 and injures 24” (high probability). As illustrated, the risk of the victim was never explicitly stated, but instead can be inferred by participants based on their prior knowledge (e.g., that there is more terrorism in Israel than in London). Thus participants’ prior expectations were again formed endogenously. Although the experimental context is necessarily artificial, these items are quite realistic. Furthermore, there is considerable research suggesting that people are better able to integrate prior expectations into their reasoning and evaluations when these prior probabilities are experienced as natural frequencies in the world, rather than explicitly stated as part of a decision problem (e.g., Gigerenzer & Hoffrage, 1995). The manipulation was validated in a norming study (n = 59 after 3 exclusions) that confirmed that people perceive differences in the relative risk of the victims. Participants chose the “high
risk” victim as being at greater risk in 61.9% of trials and chose the low risk victim in only 3.6% of trials.

In addition to the ten experimental items, I also created ten filler items divided into two types: equivalent filler items and non-equivalent filler items. Equivalent filler items were pairs of actions that differed in trivial contextual details that did not affect victims’ prior risk and that were consequently expected to be seen as equally wrong. For instance, “A man in Connecticut starts a house fire” and “A man in New Hampshire starts a house fire.” Non-equivalent filler items were pairs of actions that differed substantially in the degree of harm suffered by a victim, where one action was expected to be seen as considerably more wrong than the other. For instance, “An 11-year-old child sets a doll on fire” and “A 12-year-old child sets a cat on fire”. These filler items were intended 1) to prevent participants from becoming explicitly aware of the structure of the experimental items, and 2) to potentially help diagnose participants’ use of the response scale.

3.3.1.3 Procedure

On each trial of the experiment, participants were presented with a pair of actions labeled “action 1” and “action 2” and were asked, “Which action seems worse?” They 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). For half of the experimental items, the high victim risk action was labeled action 1 and for the other items it was labeled action 2, with the low victim risk action labeled accordingly. The order of all trials was randomized for all participants.

Two attention check questions were also included. One was randomly intermingled with the experimental trials and another occurred at the end of the trials. 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. Participants filled out demographic information at the beginning of the study.

3.3.2 Experiment: Results

Responses from seven participants were discarded after they failed attention check questions (5) or indicated that they were not paying attention (2), resulting in a final sample of 104 participants.

Likert scales are often treated as if they provide interval scaled data. In this case, however, this assumption seems almost certain to be violated. That is, differences between points on the scale used here are not equally informative. Instead, the critical differences are deviations from the midpoint: when participants indicate that one of the actions was worse as compared with indicating that neither were worse. For this reason, the “Action X seems worse” and “Action X seems a little worse” responses were collapsed into a single category.
Pooling across all responses, participants chose the neutral option on 73.2% of trials (762 of 1040). It was expected that this option would garner the greatest portion of responses, as there is a high degree of similarity among the actions in each pair—arguably, they are identical in all
relevant respects. More importantly, participants judged that actions against low risk victims were worse in 21.3% of trials (222 of 1040) and only judged that actions against high risk victims were worse in 5.4% of trials (56 of 1040). Thus, participants were biased in their evaluations of which actions were worse, being approximately 4 time as likely to select actions against low risk victims than they were to select actions against high risk victims (sign test $p < .001$). Considering items individually, participants’ responses were biased in the predicted pattern for all 10 items. A series of sign tests revealed that this bias was statistically significant for six of the items individually (all $p$’s < .05). Figure 8 shows a histogram of responses across all items and for each of the 10 items individually.

A second set of analyses was conducted across participants using bias-scores calculated for each participant. For each participant, trials were scored as +1 when the low probability action was chosen, and -1 when the high probability action was chosen. When participants indicated that neither action was worse, the trial received as score of zero. From these values a total bias-score was computed by summing across trials for each participant, ($M = 1.57$, $SD = 2.27$). A one-sample t-test revealed that these bias-scores were significantly greater than zero, $t(105) = 7.11$, $p < .001$, indicating that participants were generally biased toward view the actions against a low-risk victims as worse.

3.3.3 Discussion of Study 3

When asked to explicitly compare two immoral actions, people displayed a bias whereby they treated an action against a lower-risk victim as more severe than an action against a higher-risk victim. These findings suggest that people’s expectations not only influence their evaluations of events, but also influence their moral judgments.
In this experiment, the prior risk of each action’s victims was manipulated endogenously by the context in which the action occurred. Although this technique has the virtue of affording these items some degree of realism, manipulating context may affect other aspects of participant’s interpretation of these actions, potentially introducing confounds. I guarded against this possibility by including a variety of different items and contextual manipulations. Even if individual items are affected by confounds (and this suggestion is purely speculative), the only consistent manipulation across the items is in the risk of the victims. Thus, given that all of the items showed the predicted bias, it is most parsimonious to conclude that the manipulation of victims’ prior risk was responsible for the observed pattern of ratings.

Of course, the predicted effect was stronger for some items than for others. Of the ten items, six showed a statistically significant bias toward harsher judgments of actions against lower-risk victims. Descriptively, two of the non-significant items suggested some degree of bias and the other two showed very little bias. This variation among items suggests that factors beyond utility play a role in participant’s moral judgments of these actions. This is no surprise, as human moral judgments seem to be based on a host of factors beyond utility evaluations (see Section 2.3).

Finally, it is worth mentioning the relatively modest strength of the manipulation examined here. That is, it was always unlikely that these victims would suffer the misfortune that resulted from the immoral actions under consideration. Thus the differences in victims’ prior risk were likely quite subtle. This assumption is supported by the preliminary norming study. Participants assessed that higher-risk victims were at greater risk in the majority of instances, but failed to do so approximately one third of the time.
3.4 Study 4: Twitter Activity Following Terrorism Events

On the evening of November 13th, 2015, a coordinated terrorist attack in the city of Paris left 130 people dead and injured over 300 more. All across the world, people were shocked, horrified, and outraged by this terrible tragedy. In the aftermath, Facebook introduced a feature overlaying the colors of the French flag on users’ profile pictures, and millions took to Twitter to express their condolences under trending hashtags like #parisattacks and #jesuisparis. However, most of those mourning in the aftermath of the Paris attacks seemed unaware of yet another set of attacks, which killed at least 43 people in Beirut just 15 hours earlier. The outpouring of support for the victim of the Paris attacks and for all Parisians was surely warranted, yet where was this sense of moral outrage and support for the victims of the earlier attacks in Beirut?

To all but the terrorists themselves, attacks that kill and injure civilians are senseless and reprehensible immoral acts. Few would disagree that this is true no matter where these attacks occur. Yet public response to violent acts of terrorism, especially among those living in European and Western nations, seems to depend heavily on where and to whom these attacks occur. The November Paris attacks were shocking to all the world. Meanwhile, many people may expect terrorist attacks in Lebanon (and other Middle Eastern countries). In contrast to France, Lebanon has experienced dozens of terrorist bombings and attacks over the last three years.

To be sure, there are a wide confluence of factors that may influence people’s reactions to terrorism events: e.g., the greater death toll in Paris, the extent and immediacy of media coverage, and the perception of victims as ingroup or outgroup members (Brewer, 1999). Still, as the previous studies have shown, the prior probability of events affects their evaluation both in and outside of the lab, and that this tendency can even affect people’s moral judgments,
mitigating moral condemnation when victims are at higher initial risk. The following study examines whether prior probabilities affect reactions to immoral actions outside the laboratory, by examining Twitter activity in response to terrorism events.

3.4.1 Method

I identified 141 instances of Islamic terrorism events occurring globally between January and June of 2015 based on a listing kept on Wikipedia. I then conducted an internet search to determine whether each event was given international news coverage by international wire services (AP, Reuters, AFP) and/or by news organizations with wide international readership (BBC, Guardian, CNN, Telegraph, Dailymail, New York Times, Al-Jazeera). Of the original 141 events, international news coverage was given to 122 events. These 122 events formed the basis of the dataset. Figure 9 shows a world map marked with the locations of these events.

The Institute for Economics and Peace (IEP) publishes an annual Global Terrorism Index report (IEP, 2015) summarizing their research tracking terrorism events across the globe. The IEP estimates a terrorism impact score (TIS) that summarizes the prevalence and impact of terrorism in each nation of the world in a single value. TIS is computed based on the number of attacks and the human suffering and loss of life in those attacks. The TIS for each nation is computed based on data collected by the National Consortium for the Study of Terrorism and Responses to Terrorism at the University of Maryland. To confirm that people have accurate expectations about terrorism in a country, a small behavioral study was conducted. In an online study, U.S. participants (recruited via mTurk, n = 34) were asked to estimate how many terrorist attacks occur in different countries. As shown in Figure 10, participants’ average estimates reliably tracked the veridical TIS for each country, r(30) = .68.
Figure 9. World map marked with locations of 122 terrorism events examined in Study 4.
Twitter activity was measured using a uniform set of criteria for all events. The IBM Bluemix Twitter Insights database (a 10% random sampling of tweets) was queried to establish a count of tweets mentioning the city or country (or state/territory where applicable) where the event occurred. Tweets were queried in 13 of the 15 most popular languages among Twitter users (Mocan et al., 2013; the Bluemix platform does not label tweets in Indonesian and Malay), as shown in Table 4. Google Translate APIs were used to translate city, state, and country names into each language, and IBM’s Bluemix language codings were used to determine tweet language. A count of tweets was made for the 72 hours prior and 72 hours following the day of the event (beginning at 12:00am local time at each events’ location).
3.4.2 Results

These queries resulted in a dataset based on 7.8 million tweets. A breakdown of tweets by language is shown in Table 4. Raw counts of tweets following these events were both highly variable and strongly skewed (see Figure 11, top). Rather than perform analyses directly on these counts, a twitter reaction score (TRS) was calculated as the log odds of tweeting before and after the event were calculated. This score provides a measure of the proportional increase in tweets mentioning the relevant locations after versus before the event, controlling for the baseline popularity of a city or location and reducing skew (eq. 29).

\[
TRS = \ln(\text{tweets after event}) - \ln(\text{tweets before event})
\]

Figure 11 (bottom) shows the distribution of twitter reaction scores for the 122 events. As can be seen in the figure, there was a significant increase in tweeting following terrorism events across the 122 events, \(t(121) = 6.22, p < .001\). Although TRS remains rather skewed, it is improved over the raw counts.
Table 4. *Total number of tweets in each language and language group.*

<table>
<thead>
<tr>
<th>Language Group</th>
<th>Language</th>
<th>Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
<td>3,512,702</td>
</tr>
<tr>
<td>Non-English</td>
<td></td>
<td>4,311,145</td>
</tr>
<tr>
<td>Iberian</td>
<td>Spanish</td>
<td>613,469</td>
</tr>
<tr>
<td></td>
<td>Portuguese</td>
<td>103,507</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>716,967</td>
</tr>
<tr>
<td>European</td>
<td>Dutch</td>
<td>39,886</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>602,744</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>49,370</td>
</tr>
<tr>
<td></td>
<td>Italian</td>
<td>73,794</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>765,794</td>
</tr>
<tr>
<td>Asian</td>
<td>Japanese</td>
<td>913,003</td>
</tr>
<tr>
<td></td>
<td>Korean</td>
<td>30,330</td>
</tr>
<tr>
<td></td>
<td>Thai</td>
<td>65,289</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1,008,622</td>
</tr>
<tr>
<td>Arabic</td>
<td></td>
<td>1,643,363</td>
</tr>
<tr>
<td>Other Non-English</td>
<td>Russian</td>
<td>176,390</td>
</tr>
<tr>
<td>All Combined</td>
<td>Total</td>
<td>7,823,847</td>
</tr>
</tbody>
</table>
If prior expectations influence reactions to terrorism events, then we should expect a negative correlation between TIS and Twitter reactions. This is precisely what was observed, $r(120) = -.433$, $p < .001$. Next, the relationship between TIS and TRS was examined within a larger model (see Table 5). Though it is of little theoretical interest, one potentially predictive
variable is the popularity of the location in tweets prior to the attack. The more widely mentioned or discussed a term is, the less we might expect discussion of this term to be proportionally impacted by a single event. Consistent with this hypothesis, location popularity (measured by Twitter mentions in the time period prior to the attack, log-transformed) was negatively correlated with TRS (Table 5, Model 1). More substantively, the number of victims affected by an attack might also be expected to influence the public response. We should anticipate larger public responses to attacks that kill and injure more people. Model 2 adds predictors coding the number of victims killed and injured (both variables were highly skewed and consequently were power-transformed using a square-root function). Together, predictors for the number of victims killed and injured significantly improved model fit, $\Delta R^2 = .051$, $F(2, 117) = 3.586$, $p = .031$, although only the number of victims’ killed was significant as an individual predictor. Finally, TIS was added to the model to examine whether it was uniquely predictive over and above these other factors (Model 3). As shown in Table 5, TIS remained a significant predictor in the larger model, $\Delta R^2 = .192$, $F(1, 116) = 34.80$, $p < .001$, indicating that the influence of expectations on Twitter reactions remained after controlling for these other factors. Figure 12 presents a scatterplot comparing TIS and TRS after residualizing on the variables in model 2 ($r$TRS).

Regression diagnostics identified four potentially troublesome outliers (Studentized-deleted residual $t > 3$), but subsequent reanalysis after removing these outliers did not meaningfully change the model specification. Based on these results, people seem to react more strongly to terrorism events in countries where these events are less expected. Indeed, in terms of variance accounted for, expectations seem to play a greater role in shaping reactions than even the number of victims harmed in an attack.
Table 5. Summary of linear regression models for tweets in all languages.

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
<th>Model $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tweets prior</td>
<td>-0.343</td>
<td>-3.988</td>
<td>&lt; .001</td>
<td>.118</td>
</tr>
<tr>
<td>2</td>
<td>Tweets prior</td>
<td>-0.319</td>
<td>-3.760</td>
<td>&lt; .001</td>
<td>.169</td>
</tr>
<tr>
<td></td>
<td>$\sqrt{\text{injured}}$</td>
<td>0.096</td>
<td>0.959</td>
<td>0.339</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sqrt{\text{killed}}$</td>
<td>0.160</td>
<td>1.592</td>
<td>0.114</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Tweets prior</td>
<td>-0.264</td>
<td>-3.510</td>
<td>&lt; .001</td>
<td>.361</td>
</tr>
<tr>
<td></td>
<td>$\sqrt{\text{injured}}$</td>
<td>0.049</td>
<td>0.553</td>
<td>0.581</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sqrt{\text{killed}}$</td>
<td>0.262</td>
<td>2.900</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TIS</td>
<td>-0.449</td>
<td>-5.898</td>
<td>&lt; .001</td>
<td></td>
</tr>
</tbody>
</table>

Figure 12. Scatterplots of rTRS and TIS for tweets in all languages. Linear trend line is shown in red along with shading for 95% confidence region.
Given the naturalistic and correlational nature of this data, it is necessary to consider how other unobserved factors may have led to these results. One potential influence is media coverage: events that garner more coverage will almost certainly be discussed more than events getting little media attention. That said, the media’s coverage of events is not independent from people’s concern for those events, as media outlets tailor their coverage to the interests of consumers. Thus the media’s effect on Twitter reactions seems most likely to be one of amplification. This may even be especially true for breaking news stories. A study of Twitter activity and media coverage following the November 2015 Paris attacks found evidence that media coverage followed in response to increased Twitter activity (Roy, 2015). Therefore, media reactions are not some independent quantity but instead a component of the Twitter reactions we sought to measure here. To be sure, media reactions are likely to have outsized influence on Twitter reactions. Potential un-modeled differences in media coverage are thus likely to increase the model’s error, but they should not be expected to undermine any qualitative conclusions drawn from these results.

Another potentially influential factor are people’s feelings toward victims as ingroup or outgroup members or the degree to which the victims are members of the evaluators’ moral circle (Singer, 1981). There is strong evidence that people have stronger reactions to harms against ingroup versus outgroup members (Brewer, 1999; Lieberman & Linke, 2007) and are more willing to commit harmful violations against outgroup versus ingroup members (Bleske-Recheck et al., 2010; Cikara et al., 2010; O’Neill & Petrinovich, 1998; Petrinovich, O’Neill, & Jorgenson, 1993; Swann et al., 2010; Uhlmann et al., 2009). Thus, the demographics of Twitter users may be crucial. The worldwide Twitter sample is highly diverse, yet also strongly over-representative of certain groups and strongly under-representative of others. The reasons for this
are socioeconomic, political, and likely historical. In general, per capita Twitter usage is positively correlated with a country’s GDP (Mocanu et al., 2013). However, more specific country-level differences also affect the representation of users. For instance, Chinese users are wildly under-represented, likely because internet users are banned from accessing Twitter in China. In contrast, Twitter is an American company, perhaps explaining why there are more Twitter users in the U.S. than in any other nation (although the U.S. actually ranks 5th in per-capita Twitter users).

Crudely, the overall demographic distribution in the sample here would be expected to over-represent U.S. and Western European nations and to under-represent China, India, and lower SES nations. Thus, one potential concern is whether the observed relationship between TIS and Twitter reactions might be a byproduct of Western attitudes toward victims of different nationalities. Citizens of nations with greater levels of terrorism (e.g., Iraq, Afghanistan, Nigeria, Syria) may be seen as outgroup members by many Westerners, and considering their greater Twitter use, Westerner prejudices (especially those of people in the U.S.) are likely to have an outsized influence on the Twitter reaction score.

To resolve this issue, I analyze tweets from different regional and cultural groups separately. People from different regions of the world all have their own cultural identity and ingroup/outgroup associations. In contrast, expectations about terrorism should generally be shared across people of different cultural backgrounds and nationalities. Unfortunately, geographical information is sparse in the Twitter sample, as only about 1% of tweets are geotagged with GPS coordinates, making it essentially impossible to identify the nationality of the individual behind most tweets. However, one cultural identifier that is present in all tweets is the language in which the tweet is written. The language in which a tweet is written cannot be
expected to pinpoint the nationality of its author, but can be used to roughly differentiate the cultural backgrounds of Twitter users. By examining the relationship between TIS and Twitter reactions for tweets in different languages, we can determine whether this relationship is culturally-dependent. If this relationship is present across a diverse range of cultural groups, this would provide evidence that expectations influence reactions to terrorism events independently from the ingroup/outgroup standing of the victims of those events.

A series of cross-linguistic analyses were conducted to examine the relationship between expectations and reactions to terrorist events among different cultural groups on Twitter (differentiated by language use). First, separate analyses were conducted on tweets in English and on tweets in all other languages. English is the dominant language in use on Twitter. This owes in part to relatively wide use of Twitter in English-speaking countries, as well as to the use of English among Twitter users in non-English-speaking countries. For instance, nearly all tweets from India (92%) and Nigeria (96.6%) are written in English, despite the diversity of languages spoken in these countries (Mocanu et al., 2013). In addition, English is used in over 20% of tweets originating from France, despite being a second language for virtually all French citizens (European Commission, 2012). Still, it seems fair to assume that examining non-English tweets will significantly reduce the Western bias among the sample. Minimally, it has the effect of eliminating the influence of people living in the U.S., as nearly all tweets from users in the U.S. are written in English (96.3%).

A series of analyses were conducted using different language groupings based on a combination of linguistic, geographic, and sample size considerations. Iberian languages (Spanish and Portuguese) were grouped together to represent most of South America as well as Spain and Portugal in Europe. French, German, Italian, and Dutch were grouped together to
represent the remainder of Europe. Japanese, Korean, and Thai were grouped together to
represent tweets from Asian users. Finally, Arabic tweets were examined to reflect reactions
within the Arab world. All told, additional analyses were performed separately for six different
language groups: English, non-English, Iberian, European, Asian, and Arabic. Table 4 shows the
number of tweets recorded in each language group.

Table 6. Correlations between TIS and TRS as well as TIS and rTRS. All correlations are
significant at p < .001.

<table>
<thead>
<tr>
<th>Language Group</th>
<th>All</th>
<th>English</th>
<th>Non-English</th>
<th>Iberian</th>
<th>European</th>
<th>Asian</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>rTRS</td>
<td>-.469</td>
<td>-.399</td>
<td>-.494</td>
<td>-.407</td>
<td>-.499</td>
<td>-.394</td>
<td>-.447</td>
</tr>
<tr>
<td>n</td>
<td>121</td>
<td>121</td>
<td>121</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>TRS</td>
<td>-.433</td>
<td>-.449</td>
<td>-.373</td>
<td>-.386</td>
<td>-.417</td>
<td>-.274</td>
<td>-.439</td>
</tr>
<tr>
<td>n</td>
<td>122</td>
<td>122</td>
<td>122</td>
<td>121</td>
<td>121</td>
<td>121</td>
<td>121</td>
</tr>
</tbody>
</table>

For each language, standardized residuals were saved from a regression model predicting
TRS in that language from pre-attack Twitter activity and the number of victims killed and
injured as were conducted in the analyses reported in Table 5. Correlations were then computed
between TIS and these residualized Twitter Reaction Scores (rTRS). In addition, correlations
were computed between TIS and the unresidualized TRS. The results of each of these analyses
are shown in Table 6. TIS was negatively correlated with Twitter reactions in all analyses,
replicating the global findings within every language group examined. Figure 13 presents
scatterplots showing the relationship between TIS and rTRS for each language group, along with
a linear trend line indicating the strength of correlation.
Figure 13. Scatterplots of rTRS and TIS for individual language groups. Linear trend lines are shown in red along with shading for 95% confidence regions.
3.4.3 Discussion of Study 4

Reactions on Twitter were stronger following terrorism events in countries with lower rates of terrorism than they were following events in countries with higher rates of terrorism. This was true among users tweeting in all languages examined. The pervasiveness of this phenomenon suggests that these different reactions owe to people’s expectations about terrorism, rather than differences in their perception of or connections to victims in different countries. The expectation that terrorist attacks will be perpetrated against people living in politically unstable countries appears to mitigate people’s moral condemnation of these attacks. In contrast, the shock and surprise engendered by terrorist attacks in countries thought to be safe appears to inspire outrage and an outpouring of support for the victims.

This conclusion rests on the interpretation of Twitter activity as an indirect measure of the degree of moral outrage experienced in response to different terrorism events. It should be acknowledged that this is imperfect in a number of respects. First, this approach can only capture the public reactions of Twitter users. Second, it cannot distinguish between tweets related to the terrorism events and unrelated mentions of the search terms associated with those attacks (although this is controlled for by examining mentions both before and after the event). Lastly, it does not account for the possibility of tweets from terrorist sympathizers, who may tweet about these events in celebration rather than condemnation. It is assumed that these tweets represent a sufficiently small group of users so as to be inconsequential to the analyses reported here.

The media’s treatment of these events is also largely unaccounted for in the present study. It was confirmed that all events examined here received some amount of international media attention so that information about these at events was available shortly after their
occurrence. However, some of these events received far more attention than others. Whereas some were reported in a short blurb by the AP or another wire service, others were covered by multiple news organizations and in multiple forms of media, possibly even as breaking news. Moreover, different levels of attention were likely given to these events in different regions of the world.

Prior expectations would be expected to shape the moral reactions of members of the media just as they shape the reactions of individual Twitter users. Therefore, as suggested above, the extent of media coverage for terrorism events might also be affected by the prior probability of these events. It could be also argued that the public’s expectations about these events affects their newsworthiness—stories that violate the public’s expectations can be considered more informative than stories that simply confirm their existing expectations.

In this study, an effort was made to rule out effects of intra- and intergroup relationships in people’s reactions to terrorist attacks against different victims. The concern was that people’s risk of victimhood might be correlated with their ingroup/outgroup status toward Western observers. That is, people in countries like Iraq, Afghanistan, and Nigeria might be seen as members of the outgroup by Twitter users living in the U.S. and Europe. Consequently, Twitter reactions might be muted for actions against these outgroup victims.

Having demonstrated consistently the role of expectations in reactions to terrorism events even among different cultural groups, another perspective is suggested. People’s expectations and outgroup stereotypes may not be independent forces. Instead, expectations may be a mechanism by which certain outgroup stereotypes lead to reduced concern for the victimization of those outgroup members. Victims who are members of certain groups may engender reduced sympathy simply because evaluators are less surprised when members of that group are harmed.
On this hypothesis, it isn’t that people intrinsically do not care for Iraqis or Nigerians, for instance, but instead that they expect Iraqis and Nigerians to be victimized and so have weaker responses to their suffering. A similar mechanism may also contribute to the widespread indifference to acts of genocide in the global community (e.g., Slovic, 2010), as the groups targeted by these actions tend to be those who are already at greater risk of suffering.

4. Implications for Ethical Theory and Human Moral Judgments

The theory and findings here have potentially important implications for ethics and for the normativity of human moral judgments. In this section, I consider some of these implications. First, I discuss how the global contrast theory of event evaluation and the empirical findings reported in Studies 3 and 4 suggest a pervasive pattern of error in human moral judgments. Then, I consider how the influence of expectations on utility evaluations might explain the distinction between omission and commission, a crucial distinction in a number of ethical theories. Finally, I consider how the evaluation of events as changes in probabilistic states of the world can explain blame for risky behaviors and victim blaming.

4.1 Implications for normativity of human moral judgments

Is it rational that human utility evaluations are sensitive to the prior probability of events, and is this morally normative in the case of moral judgments? As discussed, rationality is not the most apt criterion for assessing event evaluations. It is true that the global contrast theory imagines an evaluator who evaluates events by comparing states of the world optimally according to rational principles—those of probability theory and in accord with the sense of utility captured by utility theory. This analysis suggests that sensitivity to prior probabilities is not an error driven by human limitations, but is instead an inherent consequence of comparing
uncertain states of the world. The feeling that it is better to get an unexpected promotion or to win as the underdog is not a uniquely human bias, but instead is a product of limited information.

These consequences seem most troubling when applied to moral judgment. In the moral domain, the influence of expectations on evaluations seems to produce pervasive moral error. To be sure, comparing human moral judgment behavior against a normative standard is not so straightforward as for some other behaviors. For instance, in the case of vision, a normative standard is provided either by the objective facts or by a rationally-established standard such as an ideal observer. In contrast, ethics is roiled in debate over what is truly right and what is truly good. Elsewhere, my collaborators and I have advocated a sort of methodological atheism, whereby researchers should largely avoid normative comparisons (Holyoak & Powell, under review).

Nevertheless, though there is no unified or rationally-established standard of moral behavior, we can compare human behavior against proposed moral principles in a more piecemeal fashion. One principle common to many variants of utilitarianism is that all moral agents should be given equal consideration in calculating utility (Singer, 1979). Bentham is said to have proposed the dictum “Everybody to count for one, nobody for more than one” (Mill, 1863/2004). This dictum seems to be violated when, for instance, a greater loss of utility is perceived in the loss of lives for victims at low versus high risk. In this case, these lives are not given equal consideration. Unless some principled moral justification can be given for privileging one victim over the other, then this appears to be a violation of the principle of equal consideration and a case of moral error. Likewise, unless some justification can be given for why these victims should possess different rights, this appears to be a moral error from a deontological perspective as well. It seems unlikely that these justifications can be made in many
cases. For instance, intuitively it seems plainly wrong to condemn a terrorist act against a group of Londoners more harshly than an identical act against a group of Israelis.

This is not the only case where the application of a domain-general evaluation process to moral issues leads to moral errors. If each life is equally important, as the principle of equal consideration stipulates, then the utility function over lives should be linear (Slovic, 2010). Yet, the utility function over lives has widely been shown to resemble the utility function over wealth (Kahneman & Tversky, 1984; Rai & Holyoak, 2011; Olivola & Sagara, 2009). Although a curvilinear utility function might be quite reasonable when applied over monetary gains, it is decidedly counter-normative when applied over people’s lives, as it seems there can be no principled justification for why the life of the 10th person to be saved should matter more than the 20th.

Yet the moral error resulting from the influence of expectations seems more fundamental than the moral error introduced by the domain-general utility function. It is by no means required that the same utility function be applied over wealth as over lives. Thus, errors in aggregating over lives are straightforwardly resolved by selectively applying utility functions appropriate to the context: applying a curvilinear utility function in economic contexts and a linear utility function in moral contexts. In contrast, the global contrast theory of evaluation suggests that the influence of expectations is inherent to problem of evaluating events under uncertainty.

We can of course trace the source of these errors back to the representation of events as changes in states of the world or to their evaluation as comparisons of these states. If this approach leads even highly idealized evaluators (who optimally update beliefs according to the laws of probability, apply meaningful utility functions, and so forth) to err, then we might wish to reject this conception of event evaluation. That is, we may argue that events should not be
evaluated by global comparison of the states of the world before and after an event. In this case, the question is whether there are any alternatives on the horizon. Although this may indeed be possible, these issues present a challenge to utilitarian ethicists to describe a satisfying mode of event evaluation that somehow avoids these undesirable results. Moreover, such an alternative would also need to somehow account for valid contextual influences and uncertainty over future events, as discussed in Section 2.2.1. So long as there are no answers at the ready, this should press against the reliance on utility evaluations in moral judgments and act utilitarianism as an ethical theory.

4.2 Omission-Commission Distinction

Most people feel (Borg et al., 2006; Cushman et al., 2006; 2012; DeScioli et al., 2011; Schultz et al., 1981; Spranca et al., 1991; Baron & Ritov, 2004; 2009), and many ethicists argue (e.g., Foot, 1967; Kagan, 1988; Kamm, 1993; Kant 1780/1965; Thomson, 1986; but for opposing views see e.g., Rachels, 1975; Thomson, 2008), that immoral acts of omission (e.g., failing to save someone’s life) are less severe moral violations than are acts of commission (e.g., actively killing someone) that produce similar outcomes. This principle is fundamental to certain deontological theories of ethics. The theory and findings described here may offer a psychological explanation for the distinction some ethicists make between acts of omission and commission. When a person is in a position to be harmed through another’s inaction, that person tends to be at greater risk of harm than a person who can only be harmed through direct action. Therefore, omission generally leads to a smaller change in moral utility than do acts of commission, potentially explaining the less severe moral implications of the former.

Rachels (1975) considers a pair of cases that embody the omission-commission distinction.
In the first, Smith stands to gain a large inheritance if anything should happen to his six-year-old cousin. One evening while the child is taking his bath, Smith sneaks into the bathroom and drowns the child, and then arranges things so that it will look like an accident.

In the second, Jones also stands to gain if anything should happen to his six-year-old cousin. Like Smith, Jones sneaks in planning to drown the child in his bath. However, just as he enters the bathroom Jones sees the child slip and hit his head, and fall face down in the water. Jones is delighted; he stands by, ready to push the child's head back under if it is necessary, but it is not necessary. With only a little thrashing about, the child drowns all by himself, "accidentally," as Jones watches and does nothing.

Rachels argues that there ought to be no principled distinction between the two cases, yet many people intuitively feel that acts of commission are morally worse than acts of omission. That is, they would see Smith drowning the child as more worthy of condemnation than Jones neglecting to help the child. However, it should be fairly obvious that these two cases also involve different degrees of prior risk for the young cousin. That is, the cousin’s risk of dying from drowning in the tub is far greater when he is already underwater and in the process of drowning. Therefore, Jones’ act of omission produces a smaller change in utility than Smith’s act of commission, and appears intuitively less morally wrong (although still deplorable, of course). Study 3 showed that people judged identical actions differently depending on the prior risk of the victims, offering a potential explanation for the intuition many people have about cases of omission and commission.

4.3 Moral Condemnation of Risky Actions

The present account seems to offer a natural explanation of moral and legal condemnation of risky actions. This includes condemnation for actions that put others at risk yet cause no actual harm, along with the much harsher punishments dispensed when risky behavior does cause harm. For example, consider two people who choose to drive while intoxicated. The first reaches their destination safely and the other is involved in a collision that kills another person. Legally, both are subject to punishment, but the severity of that punishment is very different—typically a monetary fine in the first case, and jail time in the latter.
Within ethics, these different intuitions about appropriate punishment are a subject of research on moral luck (e.g., Nagel, 1979; Williams, 1981). It is generally considered problematic that two agents might be judged differently for identical actions. The drunk driver who reaches their destination safely has in some ways committed just as grievous a moral violation as the driver who kills someone—the only difference was their moral luck. So, from a deontological perspective, their moral violations are the same, yet from a consequentialist perspective it is unclear that we can morally condemn the driver whose actions produced no harm. The finding that moral judgments are influenced by utility as evaluated in terms of probabilistic states of the world seems to offer some explanation for these different moral intuitions and legal responses. Driving while intoxicated increases the probability of a collision that will harm others, producing a change in utility and warranting moral condemnation and monetary fines. However, the increment in the probability of harm may be fairly small. In contrast, when a harmful collision actually occurs, the probability of harm is raised to one. Consequently, this produces a much greater loss of utility that warrants a harsher sentence.

However, some aspects of moral luck remain puzzling. The above account seems adequate for explaining why we punish drunk drivers (and others engaging in presently harmless but potentially harmful behaviors) when they are caught in the act. Yet, even when a driver reaches home safely we are likely to feel that they did something wrong and to desire to hold them accountable legally. This type of judgment in the face of known outcomes is not as easily explained by the global contrast theory account. Judgments and punishments in such cases may be explained by some other (perhaps deontological) factors.
4.5 Victim Blaming

There is a tendency for people to blame victims of certain crimes for their victimization. For instance, many people have a tendency to hold female victims of sexual harassment and assault at least partially responsible for the action of their male aggressors. This is especially true if the victims dressed or acted sexually, or if they willfully became intoxicated prior to their assault (for reviews, see van der Bruggen & Grubb, 2014; Grubb & Turner, 2012). Instances of victim blaming can also be seen in responses to police misconduct, such as the 2014 shooting of unarmed 12 year-old Tamir Rice in Cleveland, Ohio. Rice was shot by police seconds after they arrived on the scene as he was playing with a toy gun in a public park. Representatives for the city of Cleveland, in federal court filings, argued that Tamir’s death was “directly and proximately caused by the failure of [Tamir] to exercise due care to avoid injury” (Muskal & Raab, 2016). Researchers have argued that victim blaming is motivated by people’s desire to maintain “belief in a just world” (Furnham & Gunter, 1984; Lerner & Simmons, 1966; Olson & Hafer, 2001; Kay, Jost & Young, 2005). Derogating victims makes it easier to maintain this belief by viewing them as deserving of their misfortunes.

Those who blame the victim tend not to place as much blame on their aggressor. Indeed, Kay et al. (2005) argued that the purpose of blaming the victim is to mitigate wrongdoing by the person or societal system at fault. Police departments and defense attorneys alike seem to realize the mitigating force of victim blaming. For instance, defense attorneys may seek to highlight a female victims’ intoxication or previous sexual activities.

The role of prior expectations in evaluation and moral judgment offer a potential mechanism for the mitigating force of victim blaming. Victims are especially likely to be blamed for their victimization when their actions are perceived as increasing the likelihood that they will
suffer negative consequences (Janoff-Bulman, Timko, & Carli, 1985). Given that these victims face higher risk, evaluators’ reactions will be muted when these victims suffer negative outcomes. Thus the global contrast theory offers an explanation for the mitigation of offenses in cases of victim blaming.

A person’s risky behavior might sometimes offer a defensible exception to the dictum of equal consideration. That is, mitigated reactions to events where high risk victims are harmed can sometimes seem sensible: we might rightly have less sympathy for a daredevil motorcyclist who breaks his leg performing a stunt than for another motorcyclist injured on his morning commute. However, this tendency appears to be a moral error when this event evaluation is applied to moral judgments of actions. Even if it is true, statistically speaking, that alcohol consumption increases a woman’s chance of being sexually assaulted, a woman’s blood alcohol level should not be a mitigating factor in the evaluation of her assailant.

5. Conclusions

It seems intuitive that our evaluations of some events should be affected by our expectations about those events. Indeed, an idealized analysis of event evaluations as contrasted states of the world suggest that this pattern of behavior is derived not from human biases or cognitive limits but instead is inherent to problem of evaluation under uncertainty. However, it is less intuitive that our evaluations of moral actions should also be so affected. Such a pattern of behavior would seem to be a moral error. Morally, we should not be less affected by the harming of a victim just because that victim was at greater prior risk, nor should we be less harsh in our moral judgments against transgressors who harm high-risk victims.

Yet this is precisely the pattern of moral judgment and moral reactions that was observed in Studies 3 and 4. In Study 3, participants explicitly judged that actions harming high-risk
victims were less severe moral violations than actions harming low-risk victims. In Study 4, the same pattern of moral response was observed in a naturalistic study of online activity following terrorism events.

These counter-normative patterns of judgment suggest that human moral judgments are subject to potentially pervasive error. Hopefully, knowledge of this tendency will allow us to reflect on our reactions to global events, our sympathy toward victims, and our judgments of moral transgressions. At the same time, it may be hoped that awareness of the mechanisms behind our intuitive moral evaluations will help us determine what is right.
**Appendix.** Experimental materials used in Study 3.

<table>
<thead>
<tr>
<th>Item</th>
<th>High-risk victim</th>
<th>Low-risk victim</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>To save money, an automobile factory in China secretly exposes its factory workers to harmful chemicals.</td>
<td>To save money, an automobile factory in Japan secretly exposes its factory workers to harmful chemicals.</td>
</tr>
<tr>
<td>2</td>
<td>Islamic terrorists set off a bomb on a bus in Israel that kills 10 and injures 24.</td>
<td>Islamic terrorists set off a bomb on a bus in London that kills 10 and injures 24.</td>
</tr>
<tr>
<td>3</td>
<td>A male bartender drugs a woman's drink at a bar and later rapes her.</td>
<td>A male barista drugs a woman's drink at a coffee shop and later rapes her.</td>
</tr>
<tr>
<td>4</td>
<td>A corrupt government official in Mexico takes a bribe that allows a company to pollute the environment.</td>
<td>A corrupt government official in Canada takes a bribe that allows a company to pollute the environment.</td>
</tr>
<tr>
<td>5</td>
<td>A militia group kills 300 civilians in Somalia.</td>
<td>A militia group kills 300 civilians in Kenya.</td>
</tr>
<tr>
<td>6</td>
<td>A cook accidentally sells spoiled food to customers at a fast food restaurant.</td>
<td>A cook accidentally sells spoiled food to customers at a four star restaurant.</td>
</tr>
<tr>
<td>7</td>
<td>A police officer in Mexico detains a pair of tourists and refuses to release them until they pay him a $100 bribe.</td>
<td>A police officer in Spain detains a pair of tourists and refuses to release them until they pay him a $100 bribe.</td>
</tr>
<tr>
<td>8</td>
<td>A young US marine accidentally shoots and kills one of his fellow marines on patrol outside Baghdad.</td>
<td>A young US marine accidentally shoots and kills one of his fellow marines during a training exercise.</td>
</tr>
<tr>
<td>9</td>
<td>A criminal shoots and wounds a police officer while the officer is responding to a 911 call reporting a robbery in progress.</td>
<td>A criminal shoots and wounds a police officer while the officer is responding to a 911 call reporting vandalism.</td>
</tr>
<tr>
<td>10</td>
<td>A man in Detroit robs a young couple at gunpoint as they are walking in downtown.</td>
<td>A man in Denver robs a young couple at gunpoint as they are walking in downtown.</td>
</tr>
</tbody>
</table>
References


