Here’s the link to the published version: http://psycnet.apa.org/doi/10.1037/a0024861

Please cite this article as:


Acknowledgments

This work was supported by a research grant from the John Templeton Foundation, the Neubauer Family Faculty Fellows program, and research funds from the University of Chicago Booth School of Business. We gratefully acknowledge Kaushal Addanki, Brittani Baxter, Logan Berg, Andrew Gallucci, Eric Guo, Jasmine Kwong, David Levari, Matt Liebert, and Paul Thomas for assistance in conducting these experiments, and Nicholas Epley, Steve Guglielmo, Reid Hastie, Ann McGill, Jane Risen, Oleg Urminsky, and Adam Waytz for insightful comments on this work.
Abstract

People’s intuitions about the underlying causes of past and future actions might not be the same. In three studies, we demonstrate that people judge the same behavior as more intentional when it will be performed in the future than when it has been performed in the past. We found this temporal asymmetry in perceptions of both the strength of an individual’s intention and the overall prevalence of intentional behavior in a population. Because of its heightened intentionality, people thought the same transgression deserved more severe punishment when it would occur in the future than when it did occur in the past. The difference in judgments of both intentionality and punishment were partly explained by the stronger emotional reactions that were elicited in response to future actions than past actions. We consider the implications of this temporal asymmetry for legal decision making and theories of attribution more generally.
Predicting Premeditation:
Future Behavior Is Seen As More Intentional Than Past Behavior

WITWER: Let’s not kid ourselves, we are arresting individuals who’ve broken no law.

JAD: But they will.


In the 2002 film Minority Report, a specialized Precrime police department relies on the psychic foreknowledge of three clairvoyant “precognitives” to apprehend criminals who will commit future crimes. Although such a world would seem to fall squarely in the realm of science fiction, beliefs about the likelihood of a criminal’s future actions are routinely factored into important legal rulings. For instance, the Supreme Court recently established that federal officials have the legal right to hold certain prisoners beyond their sentence length to prevent them “from causing reasonably foreseeable ‘bodily harm to others’” (United States vs. Comstock, 2010). In this paper, we argue that people exhibit a systematic difference in their judgments of an agent’s future, as compared to past, behavior. Specifically, we demonstrate that people ascribe a greater degree of intentionality to actors for their future actions, and that these ascriptions of intentionality predict people’s willingness to punish future transgressions more extremely than equivalent past transgressions.

We base our predictions on two important features that differentiate the past from the future; namely, the future is typically more uncertain and more controllable than the past.
Feelings of uncertainty intensify people’s emotional reactions to situations (Bar-Anan, Wilson, & Gilbert, 2009), and events that one can control arouse more preparatory emotions than events that one cannot control (e.g., Frijda, 1988). Direct empirical comparisons of past- and future-oriented thought support the proposition that prospection arouses more intense emotional responses than retrospection (Caruso, 2010; Caruso, Gilbert, & Wilson, 2008; D’Argembeau & van der Linden, 2004; Van Boven & Ashworth, 2007).

These emotional reactions are an important input to assessments of an actor’s intentions. For instance, the emotional response of anger implies an appraisal based on an actor’s intentional agency (Tiedens, 2001; Weiner, 2001). Because the experience of negative emotion itself may lead to heightened assessments of intentionality (Alicke, 2000), actors or actions that evoke the most negative affect elicit the most blame for harmful events (Kahneman, Schkade, & Sunstein, 1998; Knobe, 2003), and negative behavior elicits more blame when it is seen as intentional rather than unintentional (e.g., Guglielmo, Monroe, & Malle, 2009; Heider, 1958).

Here, we explore the possibility that people will interpret the same action with ambiguous intent as being more intentional if it is about to happen in the future than if it has already happened in the past. Because conceptions of intentionality can be tightly bound with moral meaning (Knobe, 2003), we focus on moral actions for which both intentions and emotions affect people’s assessments of an actor’s behavior (e.g., Nichols & Mallon, 2006; Pizarro, 2000). In doing so, we integrate and extend previous research by demonstrating that moral actions that will occur in the future are judged as being more intentional than those same actions that did occur in the past. We further demonstrate that this asymmetry in ascriptions of intentionality 1) is in part explained by the stronger emotional reactions that accompany future actions, and 2) predicts people’s willingness to punish an actor’s future behavior more than past behavior.
Experiment 1: A Game of Chance

People routinely make inferences about an agent’s intentions from the consequences of that agent’s actions (e.g., Gilbert & Malone, 1995). Therefore, in Experiment 1 we provided participants with the same information about an agent’s action and tested whether inferences of intentionality are stronger when this action will happen in the future than when it did happen in the past.

Method

Four hundred seventy-two participants were presented with a hypothetical game (adapted from Cushman, Dreber, Wang, & Costa, 2009) and imagined they were playing with a stranger for real money. In the game, Player 1 chooses one of three dice that each produce different monetary payoffs for Player 1 and Player 2 depending on the number it lands on. All participants imagined that they were Player 2, who received money based on the outcome of the roll of the die chosen by Player 1. Die A was “selfish” (most outcomes favored Player 1); Die B was “fair” (most outcomes favored Player 1 and Player 2 equally); and Die C was “generous” (most outcomes favored Player 2). Participants were shown the full payoff matrix in Figure 1.

The experiment was a 2 (temporal perspective: past vs. future) X 2 (outcome: good vs. bad) between-participants design. All participants were told that Player 1 selects Die B (the fair die). Some imagined that this game happened yesterday, and others imagined that it would happen tomorrow. In addition, some were told to imagine that the outcome of the die is 6 (a good outcome for the participant), and others that the outcome of the die is 5 (a bad outcome for the participant). After reading the description of the game, participants were then asked the extent to which they thought that Player 1 intentionally tried [will try] to roll the specific number that the die did [will] land on (0=definitely not intentional; 8=definitely intentional).
Results and Discussion

Consistent with our hypothesis, participants thought the die roll was more intentional when it would happen in the future than when it did happen in the past, $F(1, 468)=7.97, p=.005, \eta^2=.02$ (Table 1). Consistent with previous research (Cushman et al., 2009; Knobe, 2003; Morewedge, 2009), participants also thought the die roll was more intentional when it led to a negative result (for the participant) than when it led to a positive result, $F(1, 468)=138.12, p<.001, \eta^2=.23$ (Table 1). There was no significant temporal perspective X outcome interaction, $F(1, 468)=1.64, p=.202, \eta^2=.003$. Planned comparisons revealed that the difference in judgments of intent was significant for bad outcomes, $F(1, 468)=8.41, p=.004, \eta^2=.02$, but not for good outcomes, $F(1, 468)=1.20, p=.275, \eta^2=.003$. These results demonstrate that even when holding an agent’s action constant, those who evaluated a future outcome saw the behavior as more intentional than those who evaluated an identical past outcome.

Experiment 2: Misreporting Taxes

To assess the robustness of this phenomenon, Experiment 2 used a different time manipulation and a different measure of intentionality. In addition, we measured participants’ affective reactions and their beliefs about the appropriate punishment for an ambiguously-intentioned behavior.

Method

Four hundred sixteen participants completed a short questionnaire assessing their opinions about taxes. To manipulate whether people were thinking about the past or the future, we used a natural time manipulation. Participants took the survey either 8-10 days prior to (future condition) or after (past condition) April 15th, the due date for income tax filings in the United States. Participants were informed that reliable analyses of tax returns have estimated
that, in any given year, approximately 21 million tax returns contain some form of misreporting, and then read that there are a number of reasons why misreporting happens. Some misreporting was described as unintentional (honest error due to ignorance of tax law, poor math skills, etc.), whereas some misreporting was described as intentional (purposeful failure to report earned money, overstating deductible income, etc.).

Participants were then asked to estimate the percentage of misreporters that they thought did [will] intentionally misreport. Following this estimate, participants indicated 1) how angry the thought of all forms of misreporting in general made them feel (0=not angry at all; 8=extremely angry), and 2) how severe the punishment should be for misreporting on taxes in general (0=not severe at all; 8=as severe as possible). Finally, participants reported whether they or someone else prepared their taxes.

Results and Discussion

Consistent with our hypothesis, participants 1) estimated that a larger percentage of tax misreporters were intentional when thinking about future misreporting compared to past misreporting, $t(414)=2.33, p=.020, d=0.23$; 2) reported that the act of misreporting on taxes made them more angry in the future than in the past, $t(412)=2.93, p=.004, d=0.29$; and 3) thought that future misreporting in general should be punished more severely than past misreporting, $t(412)=2.26, p=.025, d=0.22$ (Table 2). None of these effects differed as a function of whether participants prepared their own taxes or not, all $Fs<1$.

Using a natural time manipulation, Experiment 2 provides converging evidence that people not only see the same behavior as more intentional (Experiment 1), but also see intentional behavior itself as more prevalent in the future than in the past. However, these data do not afford an accurate specification of the relationship between intentions, emotions, and
punishment because such reactions are typically based on the behavior of a specific actor (e.g., Alicke, 2000; Malle & Nelson, 2003); it is therefore not clear how beliefs about the overall prevalence of negative intentions in a population and general punishment decisions (as we measured in Experiment 2) should translate to a specific actor within that sample. Experiment 3 addresses this issue.

**Experiment 3: A Spouse’s (Accidental?) Death**

Thinking about an event in the future tends to arouse more intense affect than thinking about that same event in the past (Van Boven & Ashworth, 2007), and such emotional responses have been shown to explain in part why people’s judgments of moral behavior can be more extreme in prospect than in retrospect (Caruso, 2010). However, this past research does not address inferences of an actor’s intentions. In addition, there is some debate over whether emotion contributes to the link between intentional attribution and moral judgment, with some researchers suggesting that emotion plays a key role (Malle & Nelson, 2003; Nadelhoffer, 2004) and others suggesting that it does not (Young, Cushman, Adolphs, Tranel, & Hauser, 2006). To shed light on this question, Experiment 3 examined the relationship between intentional ascriptions, emotions, and punishment decisions as a function of temporal perspective.

**Method**

One hundred nineteen participants read a scenario (adapted from Lagnado & Channon, 2008) in which they were asked to imagine an elderly woman named Gertrude who gives her husband the wrong medication that causes him to have a fatal heart attack. Participants either imagined that Gertrude’s actions did take place last month or will take place next month. The scenario provided reasons why Gertrude could provide the wrong medication unintentionally (her poor eyesight and the small text of the medication label) and intentionally (her rocky
relationship with her husband and a sizeable life insurance policy she would receive upon his death).

After reading the scenario, participants completed two measures of intentionality. First, they rated Gertrude’s intentions when acting on a scale ranging from 0 (absolutely did [will] not intend to kill him) to 100 (absolutely did [will] intend to kill him); second, they rated how intentional Gertrude’s action was [will be] on a scale ranging from 0 (completely unintentional) to 8 (completely intentional). Participants then rated how angry, upset, and bad the thought of Gertrude’s actions made them feel on separate scales (0=not at all; 6=extremely), and how severe a punishment they thought Gertrude should receive for her actions (0=no punishment at all; 8=extremely severe punishment).

Results and Discussion

We formed composite indices of intentionality ($r=.77, p<.001$) and negative affect ($\alpha=0.80$). Compared to participants who evaluated Gertrude’s action in the past, those who evaluated her action in the future 1) thought it was more likely that she intended to kill her husband, $t(117)=2.06, p=.042, d=0.38$; 2) had stronger negative affective reactions to her behavior, $t(117)=2.24, p=.027, d=0.41$; and 3) felt that she deserved a more severe punishment for her actions, $t(111)=2.02, p=.045, d=0.38$.

We next examined whether negative affect mediated the effect of temporal perspective on judgments of intentions (Baron & Kenny, 1986). When we included ratings of negative affect in the model, they produced a significant effect on intentions ratings, $\beta=0.39, t=4.55, p<.001$, and the effect of temporal perspective dropped to nonsignificance, $\beta=0.11, t=1.25, p=.212$ (Sobel $z=2.24, p=.025$). We also examined the opposite path; namely, whether judgments of intentions mediated the effect of temporal perspective on negative affect. When we included intentions
ratings in the model, they produced a significant effect on ratings of negative affect, $\beta=0.39$, $t=4.55, p<.001$, and the effect of temporal perspective dropped to nonsignificance, $\beta=0.13$, $t=1.53, p=.129$ (Sobel $z=1.88, p=.061$). Although the correlational nature of these analyses precludes causal inference, the results provide evidence for a bidirectional relationship whereby thoughts of future actions are associated with both stronger emotional reactions and heightened assessments of intentionality than thoughts of past actions.

We next examined the effect of affect and intentionality on punishment ratings. To do so, we conducted a series of regression analyses in which we treated temporal perspective as the independent variable and severity of punishment as the dependent variable. When we included ratings of negative affect in the model, they produced a significant effect on punishment ratings, $\beta=0.54$, $t=6.73, p<.001$, and the effect of temporal perspective dropped to nonsignificance, $\beta=0.07$, $t<1$. When we further included perceived intentions in the model, they produced a significant effect on punishment ratings, $\beta=0.58$, $t=8.28, p<.001$, and the effect of negative affect was significantly reduced (to $\beta=0.29$; Sobel $z=3.47, p<.001$). In the full model, the effect of intentions on punishment ratings was of significantly greater magnitude than the effect of negative affect, $z=3.66, p<.001$.

Taken together, the results from Experiment 3 demonstrate that thinking about an action in the future leads to stronger emotional reactions and heightened assessments of the actor’s intentions, and that both emotion and perceived intentions uniquely contribute to the relatively harsher punishment judgments that people think future transgressions warrant. Analyses of these data suggest that the effect of the past or future framing on punishment is mediated by the two pathways of negative affect and intentionality. By manipulating affect and intentionality directly, future research could attempt to pinpoint a more specific process model to determine, for
instance, whether emotional reactions precede intentionality judgments or intentionality judgments precede emotional reactions.

**General Discussion**

Oliver Wendell Holmes once suggested that “even a dog distinguishes between being stumbled over and being kicked” (Holmes, 1881). This sentiment neatly encompasses the idea that an evaluator’s interpretation of an action depends on the perceived intent of the actor. Indeed, empirical research has confirmed that human beings routinely distinguish between intentional and unintentional action (Gibbs, 2001; Malle & Knobe, 1997), and incorporate both intentions and outcomes into assessments of responsibility and blame (Cushman, 2008; Walster, 1966). In three studies, we have shown that such assessments are systematically affected by the temporal perspective that the evaluator adopts. Relative to evaluations of past behavior, evaluations of future behavior were accompanied by stronger current affect and heightened ascriptions of intentionality, both of which independently affected the extent to which actors were held accountable for the outcomes of their behavior.

There are a number of factors that could help explain why such a temporal asymmetry exists. When preparing for any type of social interaction, understanding another agent’s intentions is often necessary to make accurate predictions of its future behavior (Dennett, 1987; Gergely & Csibra, 2003; Heider, 1958). Intentional explanations for behavior are particularly helpful in understanding entities that are unpredictable, and perceiving intentions in another’s behavior – like perceiving patterns more generally – helps to fulfill a fundamental motivation to reduce uncertainty and increase feelings of personal control (Waytz, Morewedge, Epley, Monteleone, Gao, & Cacioppo, 2010; Whitson & Galinsky, 2008). Because future events are typically more uncertain and seemingly more unpredictable than past events (e.g., Fischhoff,
predicting future behavior may naturally trigger thoughts of the intrinsic intentions of an actor simply because less is likely to be known about the situational causes of another’s actions.

Our studies do not directly test people’s spontaneous attributions because we purposefully directed participants’ attention to both intrinsic intentions (e.g., Gertrude’s greed) and situational determinants (e.g., the small print on the medicine label) of an agent’s behavior. Previous research has shown that when people are not provided with any specific information about the possible reasons underlying someone’s behavior, they tend to judge all actions as intentional by default (Rosset, 2008). Only when people are motivated to think harder about a behavior do they revise their initial inference of intentional action to incorporate situational information into their judgments. It is therefore possible that the heightened affect associated with future events may constrain people’s ability to process multiple sources of information as inputs into their behavioral explanations. Alternatively, some feature of the past – such as the motivation to make sense of emotional events (e.g., Wilson & Gilbert, 2008) – may lead people to process information about an event that has already happened more extensively than an event that has yet to happen. This enhanced processing could lead them to place more weight on the situational causes that we provided, thereby moderating their automatic assumptions of intentionality.

Such an account would be consistent with research demonstrating that people are more likely to make spontaneous reference to intentions when talking about their future prospects than their past experiences (Buehler, Griffin, & Ross, 1994; Okuda et al., 2003). Indeed, because we found that affect alone was not sufficient to explain why future transgressions warranted harsher punishment than past transgressions, our results suggest that the attributional asymmetry we documented may extend to a far broader range of behaviors than emotion-backed judgments of
moral or legal transgressions. In fact, we have found in a separate line of work that even mundane behaviors (e.g., watering plants; see Malle & Knobe, 1997) are rated as more intentional in the future than the past, and that the temporal asymmetry is larger for relatively more intentional actions than relatively less intentional ones (Burns, Caruso, & Bartels, 2011). We believe that future research on the nature and scope of such temporal asymmetries may be incorporated into existing models of attribution to expand our understanding of the specific ways in which people explain the same behavior at different points in time (see, e.g., Malle & Tate, 2006).

In addition, our findings contribute to a body of empirical research designed to specify the ways in which legal views of human behavior differ from the layperson’s (Malle & Nelson, 2003). The present research also speaks to a related concern for public policy. Policy makers are generally concerned with future action, and hence are forward-looking, whereas policy enforcers are generally concerned with past action, and hence are backward-looking. This difference in temporal orientation could potentially lead to policies that are more draconian, and enforcement that is more lenient, than society would adopt if people were aware of the asymmetry we have documented here.
Temporal Inconsistency in Perceived Intentionality, 14

References


*doi:10.1371/journal.pone.0006699*


Footnotes

1 We believe that the effects of temporal perspective and valence could arise for similar reasons. We argue that emotion is one key mechanism that explains why the future seems more intentional than the past, and emotion has been implicated (in various ways) in explaining why negative actions are seen as more intentional than positive actions (e.g., Malle & Nelson, 2003; Morewedge, 2009). Although our data do not speak directly to this issue, we speculate that the two dimensions may share an underlying similarity in that it is more functional both to attend to bad things (relative to good things; Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001) and to attend to the future (relative to the past; Caruso, 2010).

2 The analyses reported are consistent for each of the two separate measures of intentionality.

3 Six participants who failed to complete this measure are excluded from this analysis.
Tables

Table 1. Means (and standard deviations) of intentionality ratings as a function of temporal perspective (Experiment 1).

<table>
<thead>
<tr>
<th></th>
<th>Intentionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad outcome</td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>3.17 (2.79)</td>
</tr>
<tr>
<td>Future</td>
<td>4.07 (2.79)</td>
</tr>
<tr>
<td>Good outcome</td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>0.87 (1.64)</td>
</tr>
<tr>
<td>Future</td>
<td>1.21 (2.11)</td>
</tr>
</tbody>
</table>
Table 2. Means (and standard deviations) of intentionality ratings, anger, and severity of punishment as a function of temporal perspective (Experiments 2 and 3).

<table>
<thead>
<tr>
<th></th>
<th>Intentionality</th>
<th>Negative Affect</th>
<th>Severity of Punishment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>30.46 (21.20)</td>
<td>2.81 (2.00)</td>
<td>3.81 (1.78)</td>
</tr>
<tr>
<td>Future</td>
<td>35.81 (22.44)</td>
<td>3.47 (2.23)</td>
<td>4.25 (1.86)</td>
</tr>
<tr>
<td>Experiment 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>-0.20 (0.87)</td>
<td>5.19 (4.37)</td>
<td>2.14 (2.06)</td>
</tr>
<tr>
<td>Future</td>
<td>0.15 (1.00)</td>
<td>7.03 (4.63)</td>
<td>3.00 (2.44)</td>
</tr>
</tbody>
</table>
**Figures**

**Figure 1.** Payoff matrix as a function of die choice and die outcome (Experiment 1; adapted from Cushman et al., 2009).

<table>
<thead>
<tr>
<th>Player 1 chooses die:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>You: $0</td>
<td>You: $0</td>
<td>You: $0</td>
<td>You: $0</td>
<td>You: $5</td>
<td>You: $10</td>
</tr>
<tr>
<td></td>
<td>Player 1: $10</td>
<td>Player 1: $10</td>
<td>Player 1: $10</td>
<td>Player 1: $10</td>
<td>Player 1: $5</td>
<td>Player 1: $0</td>
</tr>
<tr>
<td>B</td>
<td>You: $5</td>
<td>You: $5</td>
<td>You: $5</td>
<td>You: $5</td>
<td>You: $0</td>
<td>You: $10</td>
</tr>
<tr>
<td></td>
<td>Player 1: $5</td>
<td>Player 1: $5</td>
<td>Player 1: $5</td>
<td>Player 1: $5</td>
<td>Player 1: $0</td>
<td>Player 1: $0</td>
</tr>
<tr>
<td>C</td>
<td>You: $10</td>
<td>You: $10</td>
<td>You: $10</td>
<td>You: $10</td>
<td>You: $5</td>
<td>You: $0</td>
</tr>
<tr>
<td></td>
<td>Player 1: $0</td>
<td>Player 1: $0</td>
<td>Player 1: $0</td>
<td>Player 1: $0</td>
<td>Player 1: $5</td>
<td>Player 1: $10</td>
</tr>
</tbody>
</table>