

# **Inferring Intentions from Consequences**

## **Online Appendix**

### Contents

1. The “Knobe effect” and Related Studies
2. Survey Questionnaire
3. Manipulation Check Results
4. Results from University Sample (Survey #1, not presented in main paper)
5. Distribution of the Dependent Variable (Survey #2, presented in the main paper)
6. Auxiliary analysis from Mechanical Turk Sample (Survey #2)
7. Sample Demographics
8. Balance test
9. References

## The “Knobe Effect” and Related Studies

Many people believe that intended harms are a deeper moral problem than incidental harms (Borg et al. 2006; Greene 2013; Mikhail 2011; Rai & Fiske 2011; Traven 2015). This belief is not only reflected in just war doctrine (Aquinas 2002, 170), but it also forms a major part of international humanitarian law (IHL). Recent findings in moral psychology suggest that the following may also be true: moral intuitions about the consequences of an action shape the degree to which people think the action was intended (e.g., Knobe 2003). Similar to research on framing effects in economics (Kahneman and Tversky 1979), when company policies are described as “harming” the environment, people are more likely to believe that harm was intentional, compared to policies described as “helping” the environment. This is known as the Knobe effect (KE) in moral psychology.

What explains the KE? In the KE literature, this question is broken up into two main issues. First, what kind of concept is the concept of intentionality? Is it fundamentally *descriptive*, or is it fundamentally *evaluative*? In other words, when people use the concept of intentionality, are they using it to *describe or explain* what other people do, or are they using it to *cast blame* on the actions of others? One line of scholarship holds that the idea of intentional action is essentially descriptive. On this view, when subjects in the canonical KE experiment answer that the chairman intentionally harmed the environment, cognitive or emotional biases are causing them to misapply the concept (Nado 2008; Nadelhoffer 2006). Other scholars, including Knobe himself (Knobe 2006), argue that the concept of intentionality is fundamentally evaluative. When subjects respond that the chairman intentionally harmed the environment, they are not misapplying the concept. Instead, they are using it appropriately, but to cast blame, which is what the concept of intentionality is all about.

Second, what lies behind this asymmetry in intentionality assessments? Is it merely a result of the so-called negativity bias (Hibbing, Smith, and Alford 2014), is it a result of specific emotions like anger or disgust (Nadelhoffer 2004, 2006; Zucchelli, Starita, Bertini, Guisberti, and Ciaramelli 2019), or is it a result of moral intuitions focused on blame or justifications for harm (Knobe 2006; Vonasch & Baumeister 2016)? Is it sensitive to moral and political ideology (Ditto et al. 2009)? In this study, we take no position on the conceptual nature of intentionality (whether it is a descriptive or evaluative concept). However, we defend two key claims concerning the source of the KE. First, we argue that *moral intuitions*, specifically intuitions about the rightness or wrongness of a course of action, play a key role in shaping the KE. Second, we suggest that even though the KE might be sensitive to differences in moral and political ideologies, the KE largely holds across demographic subgroups, including those based on political ideology. These findings indicate that the KE is not just a figment of ideology, but rather that it is specifically related to moral intuitions regarding the rightness or wrongness of a specific course of action.

In the following survey experiments, we apply the main findings of the KE to a domain of scholarship relevant to political scientists: citizen evaluations of wartime conduct. There is already a burgeoning literature that seeks to explain how mass publics react to wartime decisions, just war theory, and the laws of armed conflict (e.g., Sagan & Valentino 2018; Chu 2018; Wallace 2013). However, this literature has yet to explore how people assess the intentionality of wartime behavior, specifically decisions to target civilian persons and civilian objects. This gap in the literature is not trivial. First, perceptions of intent directly affect judgments of legal culpability, and they also affect how domestic and international audiences react to wartime behavior. Even if a state tries to avoid killing civilians, the collateral damage deaths that result from their targeting decisions may cause people to see the actions as intentional anyways—as the Belgrade example in the main text shows. Even if states are not technically culpable for targeting civilians or targeting cultural heritage sites,

their interest in upholding their reputation requires that they *not be perceived* to be killing civilians or targeting cultural heritage sites on purpose.

Beyond research on just war doctrine and public opinion, research on the Knobe effect may also be relevant for research on cooperation over wartime conduct. Rationalist theories often model social interaction as iterated prisoner's dilemmas where rational players use tit-for-tat reciprocity to respond to cooperation with cooperation and defection with defection (Axelrod 1984). In these strategic interactions, punishing perceived defections can be risky, especially in information scarce environments where players are likely to misunderstand the intentions of one another. For example, if one state punishes another state for a "transgression" that was actually an accident, the other state may regard such punishment as unwarranted and then respond with mutual defection. In short, an action that is falsely seen as intentional could generate a costly cycle of retaliation and a breakdown in the order that international institutions are designed to uphold. Morrow (2014) describes this as the problem of cooperation under noise. The inverse is also relevant: when side-effects are positive, states may be less likely to receive "credit" or "praise" for the action, as the effects may well be viewed as unintentional. Research on the KE, including the findings that we present here, suggests that to mitigate this problem, we have to understand how people ascribe intentions to others.

Drawing from the KE literature, in the following survey, we test the following hypotheses. Hypothesis 1 holds that *wartime behaviors that cause negative, as opposed to positive, side-effects are more likely to be perceived as intentional*. This hypothesis reflects Knobe's original hypothesis. However, who decides what is morally right or wrong, especially during armed conflict? To figure out whether the KE is grounded in *moral* intuitions, Hypothesis 2 holds that *wartime actions that are judged as morally wrong are more likely to be seen as intentional*.

## Survey Questionnaire

The experiment has three independent randomizations: the consequence (*POSITIVE/NEGATIVE*), the scenario (*CIVILIAN/UNESCO*), and the severity (*LOW/HIGH*).<sup>1</sup> This generates a 2x2x2 design. The randomizations of severity are placed in brackets in the following table.

### Overview

	<i>CIVILIAN</i>	UNESCO
<i>POSITIVE</i>	[13/132] civilians saved	UNESCO heritage site [and priceless historical artifacts] protected
<i>NEGATIVE</i>	[13/132] civilians killed	UNESCO heritage site [and priceless historical artifacts] destroyed

---

<sup>1</sup> This manipulation of SEVERITY was not included in the Survey #2, the University Sample.

## **POSITIVE\_CIVILIAN\_[LOW/HIGH SEVERITY]**

Next, you will read about a hypothetical situation relating to the U.S. war in Iraq. Please read the story carefully, and then answer the following questions.

### **[page break]**

The United States fought a war against insurgent groups in Iraq. During the war, Alex, a commanding officer, discovered a major weapons facility controlled by the insurgents. Alex's command team concluded that a military operation to destroy the weapons facility would also save about [15/150] civilians held captive by the insurgents.

Alex responded, "Look, I know that the civilians will be saved, but I don't care. I just want to get rid of the weapons." Alex ordered the operation. As expected, the weapons facility was eliminated, and [13/132] civilians were saved.

In the story above, do you think that Alex intentionally saved the [13/132] civilians?

1. Yes
2. No

[If Yes]

How strongly do you feel that Alex intentionally saved the [13/132] civilians?

1. Not strongly at all
2. Somewhat strongly
3. Very strongly

[If No]

How strongly do you feel that Alex did not intentionally save the [13/132] civilians?

1. Not strongly at all
2. Somewhat strongly
3. Very strongly

## **NEGATIVE\_CIVILIAN\_[LOW/HIGH SEVERITY]**

Next, you will read about a hypothetical situation relating to the U.S. war in Iraq. Please read the story carefully, and then answer the following questions.

### **[page break]**

The United States fought a war against insurgent groups in Iraq. During the war, Alex, a commanding officer, discovered a major weapons facility controlled by the insurgents. Alex's command team concluded that a military operation to destroy the weapons facility would also kill about [15/150] civilians held captive by the insurgents.

Alex responded, "Look, I know that the civilians will be killed, but I don't care. I just want to get rid of the weapons." Alex ordered the operation. As expected, the weapons facility was eliminated, and [13/132] civilians were killed.

In the story above, do you think that Alex intentionally killed the [13/132] civilians?

1. Yes
2. No

[If Yes]

How strongly do you feel that Alex intentionally killed the [13/132] civilians?

1. Not strongly at all
2. Somewhat strongly
3. Very strongly

[If No]

How strongly do you feel that Alex did not intentionally killed the [13/132] civilians?

1. Not strongly at all
2. Somewhat strongly
3. Very strongly

## **POSITIVE\_UNESCO\_[LOW/HIGH SEVERITY]**

Next, you will read about a hypothetical situation relating to the U.S. war in Iraq. Please read the story carefully, and then answer the following questions.

### **[page break]**

The United States fought a war against insurgent groups in Iraq. During the war, Alex, a commanding officer, discovered a major weapons facility controlled by the insurgents. Alex's command team concluded that a military operation to destroy the weapons facility would also protect a UNESCO cultural heritage site [and priceless historical artifacts] that is under threat from the insurgents.

Alex responded, "Look, I know that this will help protect the UNESCO site, but I don't care. I just want to get rid of the weapons." Alex ordered the operation. As expected, the weapons facility was destroyed, and the UNESCO cultural heritage site [and historical artifacts] were protected from the insurgents.

In the story above, do you think that Alex intentionally protected the UNESCO site [and historical artifacts]?

3. Yes
4. No

[If Yes]

How strongly do you feel that Alex intentionally protected the UNESCO site [and historical artifacts]?

4. Not strongly at all
5. Somewhat strongly
6. Very strongly

[If No]

How strongly do you feel that Alex did not intentionally protect the UNESCO site [and historical artifacts]?

4. Not strongly at all
5. Somewhat strongly
6. Very strongly

## **NEGATIVE\_UNESCO\_[LOW/HIGH SEVERITY]**

Next, you will read about a hypothetical situation relating to the U.S. war in Iraq. Please read the story carefully, and then answer the following questions.

### **[page break]**

The United States fought a war against insurgent groups in Iraq. During the war, Alex, a commanding officer, discovered a major weapons facility controlled by the insurgents. Alex's command team concluded that a military operation to destroy the weapons facility would also destroy a nearby UNESCO cultural heritage site [and priceless historical artifacts].

Alex responded, "Look, I know that the UNESCO site will be destroyed, but I don't care. I just want to get rid of the weapons." Alex ordered the operation. As expected, the weapons facility was destroyed, and the UNESCO site [and historical artifacts] [was/were] destroyed.

In the story above, do you think that Alex intentionally destroyed the UNESCO site [and historical artifacts]?

5. Yes
6. No

[If Yes]

How strongly do you feel that Alex intentionally destroyed the UNESCO site [and historical artifacts]?

7. Not strongly at all
8. Somewhat strongly
9. Very strongly

[If No]

How strongly do you feel that Alex did not intentionally destroyed the UNESCO site [and historical artifacts]?

7. Not strongly at all
8. Somewhat strongly
9. Very strongly



## **Morality Question**

Do you feel that Alex's decision to order the operation was morally right or morally wrong?

1. Morally right
2. Morally wrong

[If Morally right]

How strongly do you feel that ordering the operation was morally right?

1. Not strongly at all
2. Somewhat strongly
3. Very strongly

[If Morally wrong]

How strongly do you feel that ordering the operation was morally wrong?

1. Not strongly at all
2. Somewhat strongly
3. Very strongly

## Manipulation Check

The manipulation check was asked after the study's dependent variable in Survey #1. It asked the following:

In the study you just read, what happened when Alex ordered the operation?

1. 13 civilians died
2. 13 civilians were saved
3. A UNESCO site was destroyed
4. A UNESCO site was protected
5. None of the above
6. I'm not sure

Choices 1-4 were randomized.

An analysis of the manipulation check scenario yielded the following results:

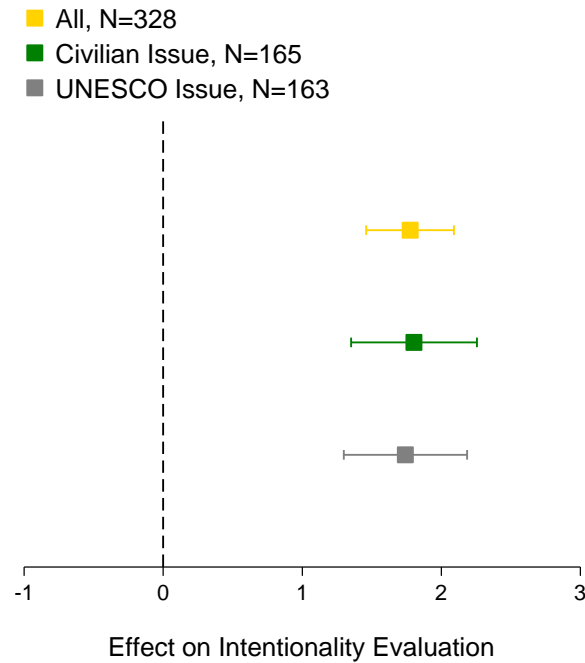
**Table A1. Manipulation Check Results.**

<b>Sample</b>	<b>Percent that Passed</b>
Overall	92.7
Consequence = <i>NEGATIVE</i>	94.0
Consequence = <i>POSITIVE</i>	91.3
Scenario = <i>CIVILIAN</i>	95.2
Scenario = <i>UNESCO</i>	90.2

## Results from University Sample (Survey #1)

The figure below shows the treatment effect of negative versus positive consequences from a military action. As discovered from the analysis in the main paper, wartime conduct that produces negative as opposed to positive consequences are more likely to be viewed as intentional acts. The effect is consistent across the laws of war scenarios (civilian lives and cultural heritage sites). Thus, the findings in the main paper replicate these findings with a second sample.

**Figure A1. Actions with Negative Consequences Cause Greater Attribution of Intentionality**



*Note: Each coefficient is generated from a separate difference-in-means test between the negative and positive-consequence experimental group. The outcome is measured on 6-point scale. 95 percent confidence intervals are displayed.*

## Distribution of the Dependent Variable (Survey #2)

Table A2 below gives the distribution of the dependent variable for the overall sample and each of the two main treatment groups (i.e., negative versus positive consequence).

**Table A2. Distribution of the Dependent Variable, %.**

<b>Value of Dependent Variable (Intentionality)</b>	<b>Full Sample</b>	<b>Positive Consequence</b>	<b>Negative Consequence</b>
1 – Strongly believe that the act WAS NOT intentional	13.4	24.4	2.3
2	22.7	30.8	14.6
3	6.4	8.7	4.1
4	3.5	3.3	3.6
5	32.4	23.7	41.2
6 – Strongly believe that the act WAS intentional	21.6	9.2	34.1

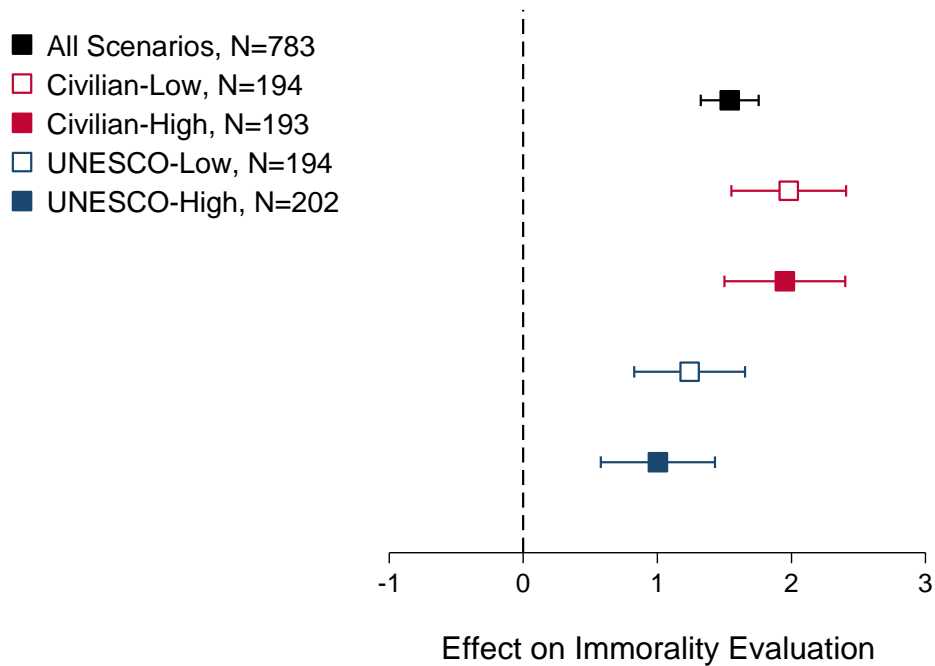
*Notes: Values might not sum to 100 due to rounding.*

## Auxiliary Analysis from MTurk Sample (Survey #2)

*MORALITY.* Figure A2 shows that, across the various scenarios, wartime conduct that produces negative outcomes are seen as more immoral. Interestingly, this effect was discovered even though the operation to destroy an insurgent weapons facility succeeded—one may think that citizens might overridingly care about whether their national military achieved its goals. Instead, the effects highly correlate with citizen evaluations of intentionality as discussed in the main paper.

**Figure A2. Negative Consequences Affect Morality Evaluations.**

---



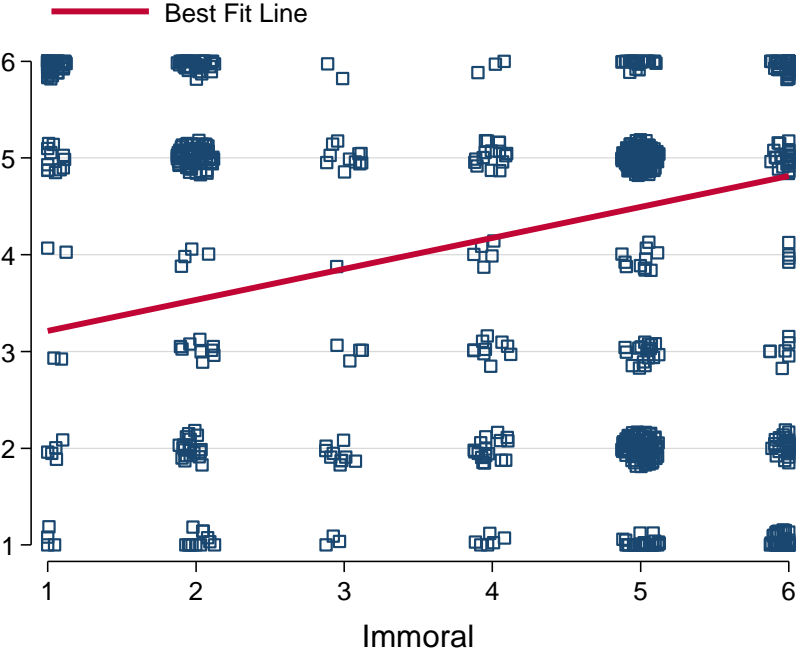
---

*Note: Each coefficient is generated from a separate difference-in-means test between the negative and positive-consequence experimental group. The outcome is measured on 6 point scale. 95 percent confidence intervals are displayed.*

Next, Figure A3 shows that people’s evaluation of (im)morality are correlated with their beliefs about intentionality (i.e., the more immoral, the more intentional). These two findings taken together are consistent with interpreting the “Knobe Effect” as resulting from a moral intuition.

**Figure A3. Respondents who believed that the act was immoral were more likely to believe that the act was intentional.**

---



---

*Note: Individual squares are jittered and represent each individual respondent. The red line is a linear fit showing the positive correlation between immorality and intentionality assessments.*

## Sample Demographics

**Table A3: Survey Respondent Characteristics**

Respondent Attribute	Survey #1 University	Survey #2 MTurk
<b>Gender</b>		
Female	56.4%	41.9%
Male	43.6%	58.1%
<b>Age</b>		
25 %-tile	19 years old	28 years old
50 %-tile	20 years old	32 years old
75 %-tile	21 years old	40 years old
<b>Race</b>		
White	73.5%	78.7%
Non-White	26.5%	21.3%
<b>Education</b>		
No college degree	n/a	42.0%
Bachelor's or higher	n/a	58.0%
<b>Party Identification</b>		
Republican	12.5%	26.6%
Democrat	61.9%	45.7%
Independent/Other	25.3%	27.7%

## Balance Test

Table A3 reports the p-values from a two-tailed difference in means test between treatment and control groups for several covariates. It shows that there is balance across groups—there is no significant difference even at the 0.10 level the standard demographic items.

Table A4

Covariate	P-value from a difference in means test	
	S1: College Sample	S2: MTurk Sample
Age	0.691	0.868
Gender	0.966	0.843
White	0.284	0.124
Education	n/a	0.825
Political Ideology	0.407	0.809
Party ID	0.369	0.277



## References

- Aquinas, St. Thomas. 2002. *On Law, Morality, and Politics (Second Edition)*. Indianapolis: Hackett Publishing Company.
- Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.
- Borg, Jana, Catherine Hynes, John Van Horn, Scott Grafton, and Walter Sinnott-Armstrong. 2006. "Consequences, Action, and Intention as Factors in Moral Judgments: An fMRI Investigation." *Journal of Cognitive Neuroscience* 18 (5): 803-817.
- Chu, Jonathan. 2018. "A Clash of Norms? How Reciprocity and International Humanitarian Law Affect American Public Opinion." *Journal of Conflict Resolution* 63(5): 1140-1164.
- Cova, Florian, Anthony Lantian, and Jordane Boudesseul. 2016. "Can the Knobe Effect Be Explained Away? Methodological Controversies in the Study of the Relationship Between Intentionality and Morality." *Personality and Social Psychology Bulletin* 42 (10): 1295-1308.
- Dill, Janina. 2015. *Legitimate Targets? Social Construction, International Law, and U.S. Bombing*. Cambridge: Cambridge University Press.
- Ditto, Peter, David Pizarro, and David Tannenbaum. 2009. *Motivated Moral Reasoning*. Vol. 50, in *Psychology of Learning and Motivation*, by Daniel, Christopher Bauman, Linda Skitka, and Douglas Medin Bartels, 307-338. Amsterdam: Academic Press.
- Downes, Alexander. 2008. *Targeting Civilians in War*. Ithaca: Cornell University Press.
- Fazal, Tanisha and Brooke Greene. 2015. "A Particular Difference: European Identity and Civilian Targeting." *British Journal of Political Science* 45 (1): 829-851.
- Greene, Joshua. 2013. *Moral Tribes*. New York: Penguin Press.
- Hibbing, John, Kevin Smith, and John Alford. 2014. "Differences in Negativity Bias Underlie Variations in Political Ideology." *Behavioral and Brain Sciences* 37: 297-350.
- Kahneman, Daniel and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263-292.
- Knobe, Joshua. 2003. "Intentional Action and Side-Effects in Ordinary Language." *Analysis* 63 (3): 190-194.
- Knobe, Joshua. 2006. "The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology." *Philosophical Studies* 130 (2): 203-231.
- Mikhail, John. 2011. *The Elements of Moral Cognition*. Cambridge: Cambridge University Press.
- Morrow, James. 2014. *Order Within Anarchy*. Cambridge: Cambridge University Press.
- Nadelhoffer, Thomas. 2006. "Bad Acts, Blameworthy Agents, and Intentional Actions." *Philosophical Explanations* 9 (2): 203-219.
- Nadelhoffer, Thomas. 2004. "On Praise, Side Effects, and Folk Ascriptions of Intentionality." *Journal of Theoretical and Philosophical Psychology* 24 (2): 196-213.
- Nado, Jennifer. 2008. "Effects of Moral Cognition on Judgments of Intentionality." *British Society for the Philosophy of Science* 59: 709-731.
- Nichols, Shaun and Joseph Ulatowski. 2007. "Intuitions and Individual Differences: The Knobe Effect Revisited ." *Mind & Language* 22 (4): 346-365.
- Pechenkina, Anna, Andrew Bausch, and Kiron Skinner. 2019. "How Do Civilians Ascribe Blame for State Indiscriminate Violence?" *Journal of Peace Research* 56 (4): 545-558.

- Rai, Tase and Alan Fiske. 2011. "Moral Psychology is Relationship Regulation: Moral Motives for Unity, Hierarchy, Equality, and Proportionality." *Psychological Review* 118 (1): 57-75.
- Sagan, Scott and Benjamin Valentino. 2018. "Not Just a War Theory: American Public Opinion on Ethics in Combat." *International Studies Quarterly* 62 (3): 548-561.
- Traven, David. "Moral Cognition and the Law and Ethics of Armed Conflict." *International Studies Review* 17(4): 556-587.
- Wallace, Geoffrey. 2012. "Welcome Guests, or Inescapable Victims? The Causes of Prisoner Abuse in War." *Journal of Conflict Resolution* 56 (2): 955-981.