

©2015

Sydney Levine

ALL RIGHTS RESERVED

THE GOOD INTENTION PRIOR: IS “THE PRESUMPTION OF INNOCENCE” AN  
ORIGINAL SIN?

By

SYDNEY LEVINE

A thesis submitted to the

Graduate School – New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Psychology

written under the direction of

Alan Leslie, PhD

and approved by

---

---

---

---

New Brunswick, NJ

January, 2015

## ABSTRACT OF THE THESIS

The Good Intention Prior: Is “The Presumption Of Innocence” An Original Sin?

by SYDNEY LEVINE

Thesis Director:

Alan Leslie

Philosophers and psychologists have long been interested in deontic judgments of cases of double effect -- morally charged scenarios in which one action has two effects and only one of those effects is intended (eg: Foot, 1967; Greene, Sommerville, Nystrom, Darley & Cohen, 2001). Cases of double effect have been critical test-cases for the study of moral judgment, giving us insight into how our moral faculty functions.

The Moral Grammar Hypothesis is one attempt to describe the cognitive mechanisms that underwrite our moral competence (Mikhail 2007, 2011). The Moral Grammar Hypothesis suggests that deontic judgments are the output of a modular system that runs a series of computations over highly structured, informationally rich mental representations. For this theory to be viable, a major question looms large: how do these structured, rich mental representations get formed from the impoverished stimuli available in the environment?

For example, determining the intentional structure of the moral agent’s mental state is critical to the Moral Grammar Hypothesis. Yet, mental state information is entirely absent from the kinds of stimuli that have been used to gather evidence in support of the theory, which are typically trolley-like cases of double effect. In fact, what

has gone largely unnoticed in research on double effect scenarios is that two possible intention structures are equally compatible with the causal structure of the case. To solve this poverty-of-the-stimulus problem, we propose that subjects deploy a *good intention prior*, namely, if the action of an agent has two effects – one good and one bad – the agent intends the good effects of her action and does not intend the bad effects.

We report two studies (one with adults, one with preschoolers aged three to five years old) that provide evidence for the use of this prior to disambiguate between intention structures in double-effect scenarios.

## Table of Contents

Title Page.....	i
Abstract.....	ii
Table of Contents.....	iv
List of Figures.....	v
Introduction.....	1
Experiment 1.....	5
Experiment 2.....	11
Discussion.....	15
Conclusion.....	16
References.....	17
Bibliography.....	19

## List of Figures

Figure 1: Possible transformations of causal structure into intention structures.....	3
Figure 2: Preschool subjects' deontic judgment of the action of the agent.....	7
Figure 3: Preschool subjects' Likert ratings of the action of the agent.....	8
Figure 4: Preschool subjects' judgments of the intention of the agent.....	10
Figure 5: Adult subjects' deontic judgments.....	13
Figure 6: Adult subjects' judgments of the intention of the agent.....	14

## Introduction

Philosophers and psychologists have long been interested in deontic judgments of cases of double effect -- morally charged scenarios in which one action has two effects and only one of those effects is intended (eg: Foot, 1967; Greene, Sommerville, Nystrom, Darley & Cohen, 2001). Cases of double effect have been critical test-cases for the study of moral judgment, giving us insight into how our moral faculty functions.

The Moral Grammar Hypothesis is one attempt to describe the cognitive mechanisms that underwrite our moral competence (Mikhail 2007, 2011). The Moral Grammar Hypothesis suggests that deontic judgments are the output of a modular system that runs a series of computations over highly structured, informationally rich mental representations. For this theory to be viable, a major question looms large: how do these structured, rich mental representations get formed from the impoverished stimuli available in the environment? Mikhail (2007, 2011) has suggested a series of computational conversion rules that translate the stimulus into a representation of the appropriate form – namely, one that can act as the input of the deontic computation.

In particular, the following series of conversion rules that operate on a stimulus (for example, a verbal or written narrative prompt) have been proposed: (1) Action descriptions (or act-tokens) are identified. (2) The temporal sequence of action descriptions is determined. (3) Causal dependencies of action descriptions are determined. (4) Valence of the action descriptions (moral goodness and badness) is determined. (5) Intentional structure of the actor is determined, allowing for denotation of ends, means, and side-effects. (6) Preliminary deontic assessment of morally salient acts is determined and these acts are situated at locations in the structured representation.

The output of steps (1)-(6) would be a representation that is appropriately structured and properly endowed with the information necessary to enter the computational process of deontic assessment.

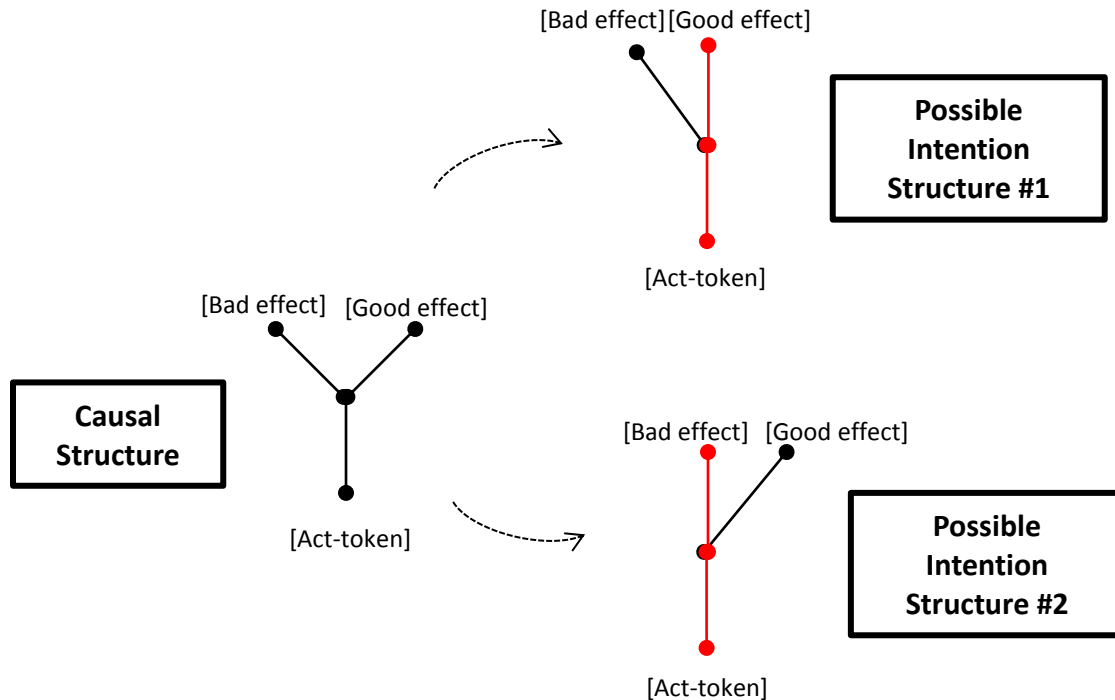
Steps (1)-(6) describe how a stimulus may be converted into a highly structured mental representation. Some of those computational steps may be outsourced by the moral module to other cognitive modules. For example step (3), determining causal dependencies, could be partially computed by the combined outputs of a module specified for mechanical object processing (for example, the ToBy system suggested in Leslie 1994) and a module specified for actional processing (for example, ToMM system 1 suggested in Leslie 1994). In step (6), a separate cognitive system may be tuned to detecting the elements of harmful battery and then then constructing a battery representation and placing it within the larger act representation (Mikhail, forthcoming).

Our focus in this paper is on step (5) – determining intentional structure of the actor. Determining the intentional structure of the moral agent's mental state is critical to the Moral Grammar Hypothesis. Yet, mental state information is (typically) entirely absent from the kinds of stimuli that have been used to gather evidence in support of the theory, which are typically Trolley-like cases of double effect (eg: Mikhail 2002; Cushman, Young & Hauser 2006; Greene, Cushman, Stewart, Lowenberg, Nystrom & Cohen 2009; Schwitzgebel & Cushman 2012). Put another way, the stimulus subjects receive is entirely impoverished on the level of intentional structure, creating a puzzle for how step (5) in the computation is carried out at all.

Here is another way of describing the problem posed by step (5) of the conversion rule sequence. Steps (1)-(4) together generate tree structures that arrange action descriptions by their causal dependencies. If the stimulus describes a case of double effect, then it can be described by one causal tree which is compatible with two different intention trees. (See figure 1 below.)



Despite this ambiguity, a broad range of subjects consistently choose one intentional structure and reject the other in trolley-like cases (Mikhail 2002; O'Neill & Petrinovich 1998; Hauser, Cushman, Young, Jin, & Mikhail, 2007). How do subjects do this?



**Figure 1: Possible transformations of causal structure into intention structures.** When one action has two effects (for instance, one good effect and one bad effect), a single causal structure is compatible with two different intentional structures. In this figure, the red line indicates the agent’s action plan – her sequences of actions she intends to bring about her goal. As shown here, it is possible that the agent intends the good effect and that the bad effect is a foreseen but unintended side-effect of the basic act-token (possible intention structure #1). The reverse possibility is also a viable transformation of the causal structure (possible intention structure #2).

To answer the puzzle of how subjects successfully compute step (5), we posit a “presumption of innocence” – that is, a Good Intention Prior (Mikhail 2007, 2011). The Good Intention Prior

(which is deployed when no other intention information is available) contains the following information: *if the action of an agent has two effects – one good and one bad – the agent intends the good effects of her action and does not intend the bad effects*. This prior provides information that allows for the selection of one of the two possible intention trees that are compatible with the causal tree computed in step (4). We present the results of two experiments – one with preschoolers and one with adults – which provide support for the use of the Good Intention Prior to disambiguate intention structures in double-effect scenarios.

## Experiment 1

We suggest that in typical double effect cases (in which no intention information is presented in the stimulus) subjects assume that the agent intended the good effects and not the bad effects of her action. However, if intention information is supplied, we suggest that the prior is not deployed and, instead, the intention information provided by the stimulus is used to generate a structural description of the action. If this is true, then consider a case in which intention information is given explicitly in the stimulus that is more-or-less equivalent with the information that would have been supplied by the prior had no intention information been present. In such a case, subjects' judgments (on measures of intention and deontic status) should be the same as they would have been had no intention information been supplied at all. In contrast, for a case that supplies intention information that is at odds with the information supplied by the prior, then we would expect subject judgments (on measure of intention and deontic status) to change to reflect the new intention information.

### *Methods*

Subjects were tested individually in quiet locations in their preschools or in the lab. Subjects were first trained on a Likert scale, with X's on one end and stars at the other. Children were taught that the ends of the scale could be used to talk about things that were "really bad" and "really good" and that the intermediate points were for things that were "a little bad" and "a little good", with the point in the middle being for things that were "just OK." Children were asked to practice using the scale. Then, children were then told stories in which a simple morally good or bad action took place. Children were asked to issue a straight-forward deontic judgment of the action ("Should he/she have done that?") and were asked to rate the action on the Likert Scale. Only children who expressed competence making simple deontic assessments and using the scale to describe moral behavior were tested further.

Children were then told a story in which a girl prevented a squirrel from eating five children's cookies (by putting up a gate) and with the same action caused the squirrel to eat one child's cookie. The causal structure of the girl's action was unambiguous. However, two intentional structures are compatible with the causal structure: it is possible that the girl's intention was to save the five children's cookies (and that the harm to the one child was a foreseen but unintended side-effect) or that the girl intended to cause the squirrel to eat the one child's cookie (and that saving the five was a foreseen but unintended side-effect).

Subjects received the story in one of three conditions. Subjects in the Neutral Condition received no explicit information about the agent's intention. Subjects in the Good Condition received information that the agent intended the good effects of her action. Subjects in the Bad Condition received information that the agent intended the bad effects of her action. A series of control questions were asked to ensure subject memory and comprehension of the story. Children were then asked three test questions. First, they were asked to judge the agent's intention: "Did [actor's name] make this one kid sad on purpose?" Second, they were asked to issue a deontic judgment of the agent's action: "In this story [actor's name] used her gate. Should she have done that?" Finally, subjects were asked to rate the actor's action on the Likert Scale: "Was what [actor's name] did good, bad, or just OK?"

### *Subjects*

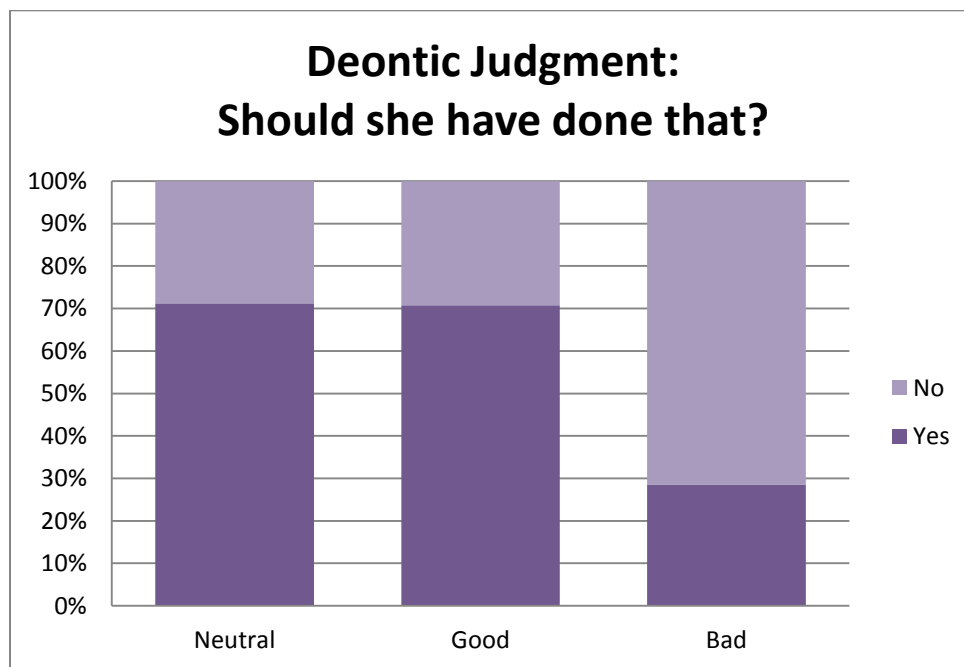
Thirty-five children between the ages of 36 months and 72 months received the Neutral Condition ( $M = 52.8$  ;  $SD = 8.4$ ), 20 of which were girls. Forty-one children between the ages of 41 months and 72 months received the Good Condition ( $M = 53.6$ ;  $SD = 8.4$ ), 20 of which were girls. Twenty-one children between the ages of 42 months and 68 months received the Bad Condition ( $M = 51.9$ ;  $SD = 7.5$ ), 10 of which were girls.

### *Results*

First we will consider subjects' deontic judgments of the case. Subjects in the Neutral Condition judged the case permissible – that is, they responded “yes” to the question “Should she have done that?” (binomial test,  $p < 0.001$ , one-tailed; Bayes factor 69:1 against the null).

Subjects in the Good Condition judged the case permissible (binomial test,  $p < 0.01$ , one-tailed; Bayes factor 13:1 against the null). Subjects in the Bad Condition judged the case impermissible (binomial test,  $p < 0.05$ , one-tailed; Bayes factor 3.4:1 against the null).

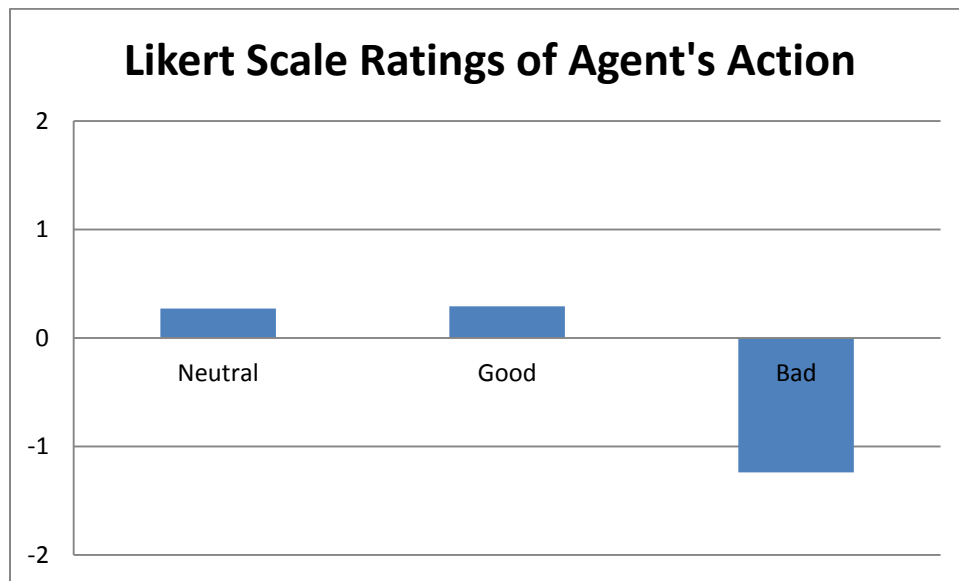
Importantly, there was no significant difference between the Neutral and Good Conditions ( $\chi^2 = 0.002$ ,  $p = 0.96$ , two-tailed). In fact, Bayesian analysis favors the null hypothesis that subject's deontic judgments in Conditions 1 and 2 are drawn from the same distribution (Bayes factor 4.4:1 for the null). There was a significant difference between the Neutral and Bad Conditions ( $\chi^2 = 10$ ,  $p < 0.005$ , one-tailed; Bayes factor 89:1 against the null) as well as a significant difference between the Good and Bad Conditions ( $\chi^2 = 8.4$ ,  $p < 0.005$ , one-tailed; Bayes factor 45:1 against the null). See Figure 2.



**Figure 2:** Preschool subjects' deontic judgment of the action of the agent.

Next, we will consider subjects' Likert ratings of the action of the agent. Likert scale ratings were scored as follows: really bad = -2; a little bad = -1; just OK = 0; a little good = 1; really good = 2. Subjects in the Neutral Condition and the Good Condition both rated the agent's action as slightly above the midpoint of the scale (Neutral:  $M = 0.27$ ,  $SD = 1.4$ ; Good:  $M = 0.29$ ,  $SD = 1.3$ ). Subjects in the Bad Condition rated the agent's action as bad ( $M = -1.2$ ;  $SD = 0.88$ ).

Importantly, there was no significant difference between the Neutral and Good Conditions (independent-sample t-test,  $t(74) = 0.077$ ,  $p = 0.94$ ). Furthermore, Bayesian analysis shows that there is evidence in favor of the null hypothesis that Likert Ratings in these conditions were drawn from the same distribution (Bayes factor 5.5:1 for the null). There was a significant difference between the Neutral Condition and the Bad Condition (independent-sample t-test,  $t(54) = 4.8$ ,  $p < 0.0001$ ) as well as between the Good Condition and the Bad Condition (independent-sample t-test,  $t(60) = 5.1$ ,  $p < 0.0001$ ). See Figure 3.

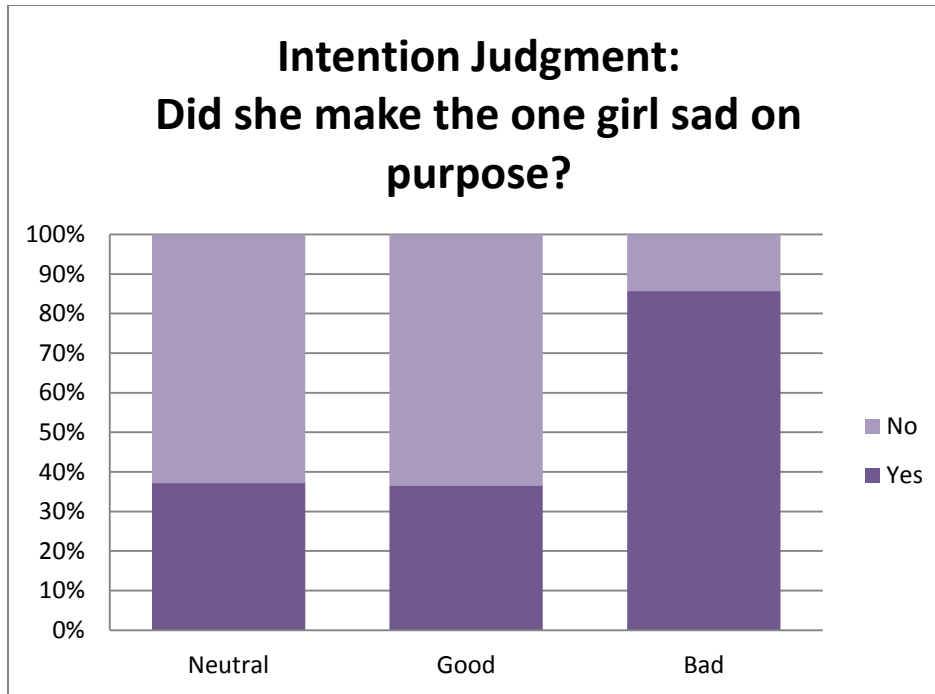


**Figure 3:** Preschool subjects' Likert ratings of the action of the agent.

Finally, we will consider subject's intentionality judgments. Subjects in the Neutral Condition judged that the agent did not intend the bad effect of her action (though this judgment

was only marginally significant: binomial test,  $p < 0.1$ , one-tailed). Subjects in the Good Condition also judged that the agent did not intend the bad effect of her action (also marginally significant: binomial test,  $p < 0.1$ , one-tailed). Subjects in the Bad Condition judged that the agent did intend the bad effect of her action (binomial test,  $p < 0.001$ , one-tailed).

Importantly, there was no significant difference between the Neutral and Good Conditions ( $\chi^2 = 0.002$ ,  $p = 0.96$ , one-tailed). In fact, Bayesian analysis provides evidence in support of the null hypothesis (Bayes factor 3.7:1 for the null). Given that Bayesian analysis suggests that subjects' judgments about the intention of the agent in the Neutral and Good Conditions are drawn from the same distribution, when we combine those response sets we find a significant pattern of subjects judging that the agent did not intend the bad effect of her action (binomial test,  $p < 0.01$ , one-tailed; Bayes factor 3.9:1 against the null). There was a significant difference between the Neutral and Bad Conditions (Fischer's Exact,  $p < 0.001$ ; Bayes factor 204:1 against the null) as well as between the Good and Bad Conditions (Fischer's Exact,  $p < 0.001$ ; Bayes factor 331:1 against the null). See Figure 4 below.



**Figure 4:** Preschool subjects' judgments of the intention of the agent.



## Experiment 2

Experiment 2 was designed to show that adults respond to double effect scenarios in a similar way to the preschoolers in Experiment 1. We used the classic “Trolley” paradigm as our stimulus.

### *Methods*

Subjects read a story in which an agent prevented a train from killing five people who were standing in its path (by throwing a switch) and with the same action caused the train to turn down a side-track and kill one person standing there. Just like in the preschool version of the story, the causal structure of the agent’s action was unambiguous. However, two intentional structures are compatible with the causal structure: it is possible that the agent’s intention was to save the five people (and that harm to the one person was a foreseen but unintended side-effect) or that the agent intended to harm the one person (and that saving the five was a foreseen but unintended side-effect).

Paralleling the preschool version, subjects received the story in one of three conditions. Subjects in Condition 1 received no explicit information about the agent’s intention. Subjects in Condition 2 received information that the agent intended the good effect of his action. Subjects in Condition 3 received information that the agent intended the bad effect of his action. Subjects were then asked two test questions. First, they were asked to issue a deontic judgment of the agent’s action: “Is it morally permissible for [actor’s name] to throw the switch?” Second, they were asked to judge the agent’s intention by responding “true” or “false” to the following statement: “Hank threw the switch in order to kill the man.” (Subjects also responded to a series of additional true/false questions about the story, which are not relevant to the present analysis.)

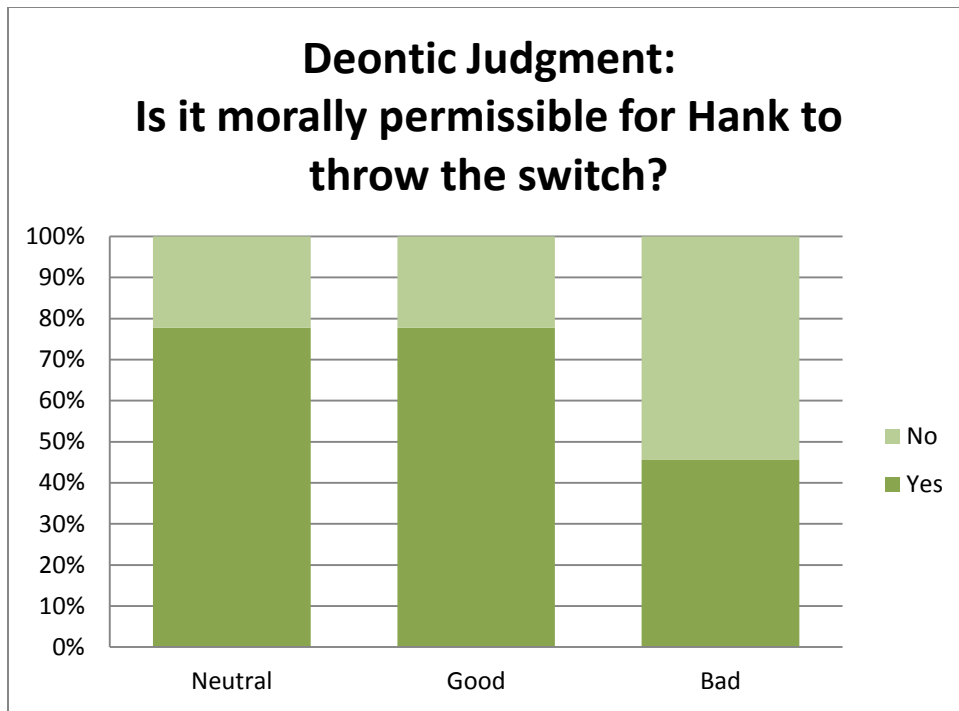
### *Subjects*

One hundred five subjects were recruited from the Mechanical Turk website (mturk.com). Subjects were excluded from analysis who failed to follow the instructions. All subjects were paid for participating.

### *Results*

First we will consider subjects' deontic judgments of the case. Subjects in the Neutral and Good Conditions judged the case permissible (binomial test,  $p < 0.001$ , one-tailed; Bayes factor for both conditions 480:1 against the null). Subjects in the Bad Condition were split between judging the case permissible and judging the case impermissible (binomial test,  $p = 0.33$ , one-tailed). Bayesian analysis shows that there is some evidence in favor the null hypothesis (that subjects are choosing between deontic judgments with equal probability; Bayes factor 8.3:1 for the null).

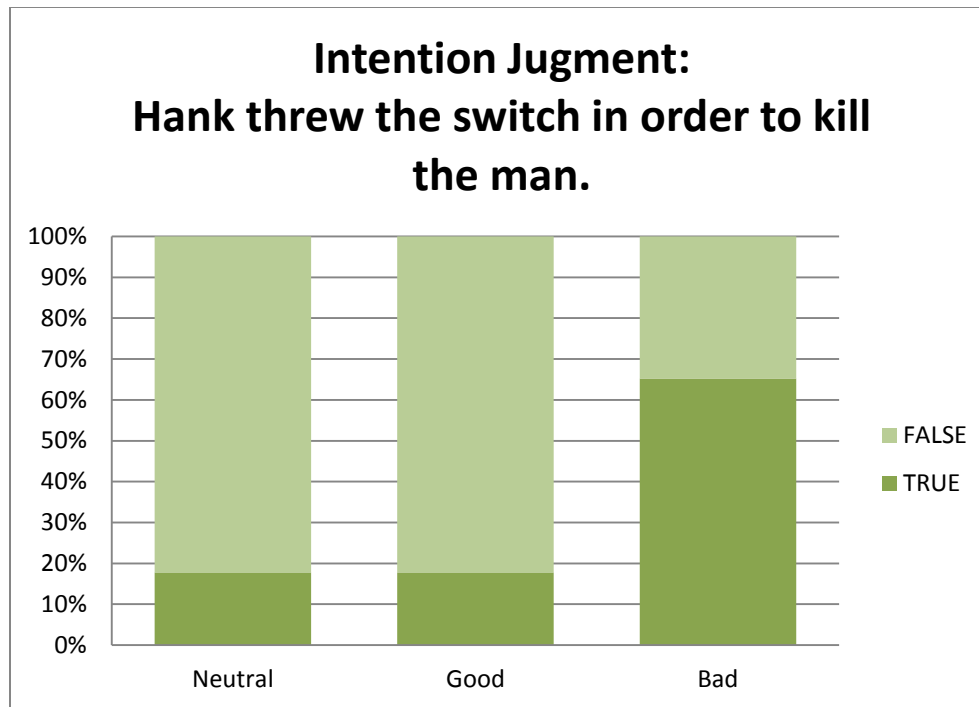
Importantly, there was no significant difference between the Neutral and Good Conditions. Bayesian analysis provides evidence in favor of the null hypothesis (Bayes factor 4.6:1 for the null). There was a significant difference between the Neutral Condition and the Bad Condition ( $\chi^2 = 8.61$ ,  $p < 0.005$ , one-tailed; Bayes factor 35.4:1 against the null) as well as a significant difference between the Good and Bad Conditions ( $\chi^2 = 8.61$ ,  $p < 0.005$ , one-tailed; Bayes factor 35.4:1 against the null). See Figure 5 below.



**Figure 5:** Adult subjects' deontic judgments.

Next, we will consider subjects' intentionality judgments of the case. Subjects in the Neutral and Good Conditions judged that the agent did not intend the bad effect of his action (binomial test,  $p < 0.0001$ , one-tailed; Bayes factor 7100:1 against the null in both conditions). Subjects in the Bad Condition judged that the agent did intend the bad effect of his action (binomial test,  $p < 0.05$ , one-tailed; Bayes factor 3.0:1 against the null).

Importantly, there was no significant difference between the Neutral and Good Conditions. In addition, Bayesian analysis provides evidence in favor of the null hypothesis (Bayes factor 5.03:1 for the null). There was a significant difference between the Neutral and Bad Conditions ( $\chi^2 = 19.1$ ,  $p < 0.0001$ , one-tailed; Bayes factor  $1.2 \times 10^4$ :1 against the null) as well as a significant difference between the Good and Bad Conditions ( $\chi^2 = 19.1$ ,  $p < 0.0001$ , one-tailed; Bayes factor  $1.2 \times 10^4$ :1 against the null). See Figure 6 below.



**Figure 6:** Adult subjects' judgments of the intention of the agent.

These findings suggest that subjects in the neutral condition, in the absence of receiving countervailing information about the agent's intention, presumed that the agent intended the good effects of her action and not the bad effects.

## Discussion

The principle finding is that -- in both the preschool experiment and the adult experiment, which used different prompts with similar structures -- there was no difference in subject judgments on measures of intention and deontic status for a case where no intention information is provided and a case where the agent's good intention is made explicit. In contrast, when the prompt contained information that the agent intended the bad effect, there were significant differences on measures of intention and deontic status when that case was compared with each of the other two cases. This suggests that when no intention information is present, subjects are deploying a Good Intention Prior, which allows them to assume that the agent intends the good effects and not the bad effects of her action. This intention information is integrated into the agent's structural description of the act, preparing the representation for the next step of the computation in which moral grammatical rules are applied and a deontic judgment is generated.

A surprising difference between Experiment 1 and Experiment 2 was in subjects' judgments of the deontic status of the condition in which a bad intention was made explicit. Preschoolers robustly judged this case as impermissible while adults were split between permissible and impermissible. It remains an open question whether this is a developmental change or a figment of the difference between the adult prompts and the preschool prompts. Future research will investigate each of these possibilities.

## Conclusion

The final question that remains to be addressed is whether the Good Intention Prior – the presumption of innocence that we apply to agents when no intention information is available – is an original sin. We have provided some evidence in favor of this presumption being “original” to cognition; even children as young as 3-years of age were shown to apply the prior. Of course, it is possible that even children this young may have learned this information from the data available in their environment. To differentiate between these hypotheses, future studies should examine whether children who are reared in less positive environments have a similar presumption.

Finally, is the presumption of innocence a sin or a virtue? Rand, Greene & Nowak (2012) reported a series of studies in which participants in economic game setting who were prompted to respond quickly and intuitively were more generous than subjects who were prompted to respond slowly or deliberatively. The authors suggest that this strategy is used by those who take themselves to be operating in a cooperative environment. Delton, Krasnow, Cosmides & Tooby (2011) note that the long-standing finding of subject generosity in one-shot economic games continues to be difficult to explain. The authors suggest an evolutionary explanation that explains generosity in these settings as a byproduct of the selected-for generosity in dyadic settings. Both groups of authors suggest that in an environment where cooperation is assumed (which also seems to be the default assumption of most participants), we seem to display immediate, automatic cooperative behavior. In a similar vein, we suggest that presuming the innocence of the agents around us is another virtue that is both derived from and allows for group coherence.

## References

- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuitions in moral judgment: Testing three principles of harm. *Psychological Science*, 17(12), 1082-1089.
- Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences*, 108(32), 13335-13340.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5-15.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105-2108.
- Greene, J.D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, 111(3), 364-371.
- Hauser, M., Cushman, F., Young, L., Jin, R.K., & Mikhail, J. (2007). A dissociation between moral judgment and justification. *Mind & Language*, 22(1), 1-21.
- Leslie, A.M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In L. Hirschfeld and S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119-148). New York, NY: Cambridge University Press.
- Mikhail, J. (forthcoming). Any Animal Whatever? Harmful Battery and its Elements as Building Blocks of Moral Cognition, 124 *Ethics*.
- Mikhail, J. (2002). Aspects of the theory of moral cognition: Investigating intuitive knowledge of the prohibition of intentional battery and the principle of double effect. *Georgetown University Law Center Public Law & Legal Theory Working Paper No. 762385*. Available at: <http://ssrn.com/abstract=762385>.
- Mikhail, J. (2007). Universal moral grammar: Theory, evidence, and the future. *Trends in Cognitive Science*, 11(4), 143-152.
- Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. New York, NY: Cambridge University Press.
- O'Neill, P., & Petrinovich L. (1998). A preliminary cross cultural study of moral intuitions. *Evolution and Human Behavior*, 19(6), 349-367.
- Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, 489(7416), 427-430.

Schwitzgebel, E., & Cushman, F. (2012). Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. *Mind & Language*, 27(2), 135-153.



## Bibliography

- Chemla, E., Egré, P., & Schlenker, P. (2013). Predicting moral judgments from causal judgments. *Philosophical Psychology*, 1-28.
- Feinfield, K. A., Lee, P. P., Flavell, E. R., Green, F. L., & Flavell, J. H. (1999). Young children's understanding of intention. *Cognitive Development*, 14(3), 463-486.
- Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33(2), 101-121.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*, Henry Holt and Co. Inc., New York, NY, 2.
- Moore, A., Clark, B. and Kane, J. "Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment." *Psychological science* 19.6 (2008): 549-557.
- Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, 100(3), 530-542.
- Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, 17(3), 145-171.
- Pellizzoni, S., Siegal, M., & Surian, L. (2010). The contact principle and utilitarian moral judgments in young children. *Developmental Science*, 13(2), 265-270.
- Sinnott-Armstrong, W., Mallon, R., McCoy, T., & Hull, J. G. (2008). Intention, temporal order, and moral judgments. *Mind & Language*, 23(1), 90-106.
- Spelke, E. S., Bernier, E. P., & Skerry, A. E. (2013). Core Social Cognition. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us* (pp. 11-16). Oxford University Press.
- Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb intervention myopia in moral intuitions. *Psychological science*, 18(3), 247-253.
- Wiegmann, A., & Waldmann, M. R. (2014). Transfer effects between moral dilemmas: A causal model theory. *Cognition*, 131(1), 28-43.
- Zimmerman, A. (2013). Mikhail's Naturalized Moral Rationalism. *Jerusalem Review of Legal Studies*, jls016.