Moral norms inform mental state ascriptions

Kevin Uttich (uttich@berkeley.edu)
Tania Lombrozo (lombrozo@berkeley.edu)
Department of Psychology, 3210 Tolman Hall
Berkeley, CA 94720 USA

Abstract

Theory of mind, the capacity to understand and ascribe mental states, has traditionally been conceptualized as analogous to a scientific theory. However, recent work in philosophy and psychology has documented a "side-effect effect" suggesting that moral evaluations influence mental state ascriptions, and in particular whether a behavior is described as having been performed 'intentionally.' This evidence challenges the idea that theory of mind is analogous to scientific psychology in serving the function of predicting and explaining, rather than evaluating, behavior. In three experiments, we demonstrate that moral evaluations do inform ascriptions of intentional action, but that this relationship arises because behavior that conforms to norms (moral or otherwise) is less informative about underlying mental states than is behavior that violates norms. This analysis preserves the traditional understanding of theory of mind as a tool for predicting and explaining behavior, but also suggests the importance of normative considerations in social cognition.

Keywords: Social Cognition; Morality; Theory of mind.

Introduction

Consider sitting at a commencement address and thinking, "that speaker must love to wear billowy black gowns." This attribution is odd, because we know that academic norms dictate commencement attire. But upon viewing someone dressed in full regalia at a café, it might be appropriate to infer an underlying mental state, such as a false belief that it's commencement or a desire to look scholarly, because the norm does not apply. These examples illustrate that norms inform mental state ascriptions. More precisely, prescriptive norms provide "desire-independent reasons" for action (Searle, 2001), with the consequence that norm-conforming behavior offers a weak basis for inferring underlying mental states, while norm-violating behavior offers a strong basis for inferring underlying mental states, as the reason for action must be sufficiently strong to outweigh the desire-independent reason to observe the norm.

The capacity to understand and attribute mental states is often characterized as a theory of mind (e.g. Gopnik, 1999). Like a scientific theory, Theory of Mind (ToM) posits unobserved entities (internal states) to support explanation and prediction. Knowing that a man in a café desires to appear scholarly, for example, can explain eccentric attire, and supports predictions about whether he is more likely to smoke a pipe or a cigar. But for the commencement speaker, eccentric attire is better explained by appeal to a conventional norm, and smoking habits are better predicted on the basis of statistical norms. These observations suggest that norms *should* inform mental state ascriptions if ToM is to accomplish the function of predicting and explaining behavior.

This paper explores the relationship between norms and mental state ascriptions by considering the relationship between moral norms and ascriptions of intentional action. Previous work suggests that ascriptions of intention have an impact on moral evaluations (Malle & Nelson, 2003). For example, an intentional killing is typically judged a murder, while an unintentional killing is considered manslaughter (e.g. California Penal Code). But recent findings suggest that the reverse may likewise hold – that moral evaluations can influence ascriptions of intentional action (Knobe, 2006). Specifically, Joshua Knobe has uncovered an intriguing asymmetry in judgments concerning intentional action for morally good versus bad side effects, a phenomenon known as the side-effect effect. Consider the following vignette, which Knobe presented to participants in his initial studies:

> The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.'
> The chairman of the board answered, 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program.'
> They started the new program. Sure enough, the environment was harmed.

When participants were asked if the chairman intentionally harmed the environment, 82% said yes. However, when the new program's side effect was to *help* the environment, only 23% of subjects said the chairman intentionally helped the environment (Knobe, 2003a). Because the harm and help vignettes seem to differ only in the moral valence of the side effect, the results suggest that *moral* considerations somehow influence ToM judgments.

Broadly speaking, responses to the side-effect effect have fallen into two distinct camps (see Nado, 2008, for a similar taxonomy). 'Competence' models take the effect as evidence that ToM competencies are shaped by the role ToM judgments play in assigning praise and blame: Knobe writes, "…moral considerations are actually playing a role in the fundamental competencies underlying our use of the concept of intentional action" (Knobe, 2006). This interpretation not only challenges the idea that the influence of ToM judgments on moral judgments is one-way (see Fig.

1a & 1b), but also the idea that the function of ToM is to predict and explain behavior – instead, ToM may be a multipurpose tool partially shaped by its role in moral evaluation.

'Performance' models instead suggest that the effect results from a bias in ToM judgments. On this view, moral evaluations are not *conceptually* related to ToM judgments, but do exert an extraneous influence. For example, conversational pragmatics (Adams and Steadman, 2004a) or the desire to blame an agent for a negative outcome may lead participants to (mistakenly) describe the side-effect as intentional (Malle & Nelson 2003; Mele, 2001; Nadelhoffer, 2004b). This view preserves the traditional function of theory of mind, adding the claim that moral evaluations can have a biasing effect (see Fig. 1c).

We propose a third way of explaining the side-effect effect and understanding the relationship between ToM and moral judgment. Perhaps moral judgments inform ToM, but not because moral considerations partially constitute or bias ToM concepts. Rather, as suggested in the introduction, actions that violate norms (e.g. harming the environment) may provide more information about an agent's mental states than do actions that conform to norms. We call this the 'inferential' model to emphasize that actions that differ in their relationship to norms (moral or otherwise) may support different inferences about underlying mental states (see Fig. 1d).
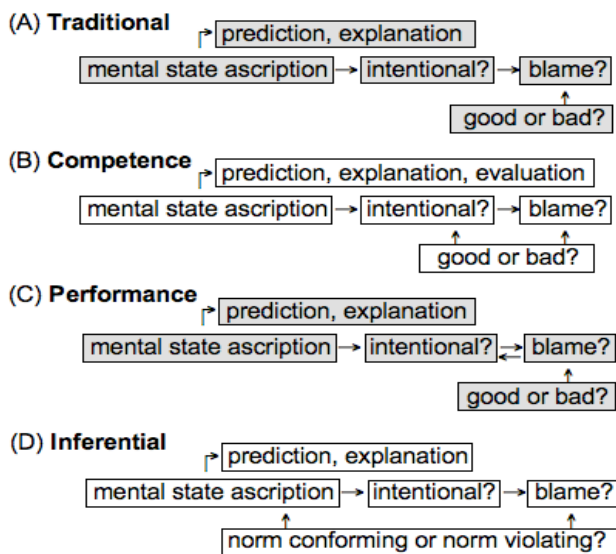

Figure 1: Schematic representations of proposed relationships between ToM and Moral Evaluations (modified and expanded from Knobe and Mendlow, 2004).

The inferential model differs from competence models in preserving the traditional function of theory of mind: prediction and explanation. The inferential model concedes that moral judgments influence ToM, but this influence is seen as *evidential*, not *constitutive*. The inferential model also differs from performance models in regarding the influence of moral judgment on ToM as a rational strategy for achieving the function of ToM, and not as a bias or extraneous pressure.

In this paper we test the inferential model as a hypothesis about the relationship between moral evaluation and theory of mind. Specifically, we examine the prediction that norm-violating behavior provides stronger evidence than does norm-conforming behavior concerning an agent's mental states, and this in turn informs predictions and impacts judgments of intention (see Fig 1d). Experiment 1 considers whether behavior with bad side effects provides more predicatively useful information than behavior with good side effects. Experiment 2 considers whether differences between good and bad side effects stem from the moral valence of the side effects themselves, or from the evidence that moral valence provides about the nature of an agent's behavior. Finally, in Experiment 3 we examine whether differences between good and bad side effects are the result of goodness and badness per se, or of the relationship between behavior and norms.

Experiment 1

Experiment 1 examines whether *norm-violating* (morally bad) behavior has a greater impact on predictions than does *norm-conforming* (morally good) behavior. Only the inferential model (Fig 1) incorporates a causal path from moral valence to future predictions. We test this causal path by presenting participants with vignettes involving agents who bring about good or bad side effects, but instead of having participants judge whether the side effect is intentional, they make two predictions about future behavior. The *specific* prediction considers whether the agent is more likely to repeat the same norm-conforming or norm-violating behavior in the future. The *general* prediction concerns broader adherence to norms, and thus examines whether the inferred properties of the agent are restricted to the specific outcome in the vignette (e.g. harming the environment) or generalize more broadly (e.g. harming in general). In addition to the vignette involving a CEO introduced above, a new vignette with a doctor was added in case negative preconceptions about CEOs play a role in the side-effect effect.

The inferential view predicts that because norm-violating behavior provides more information about underlying mental states than does norm-conforming behavior, participants who learn about the CEO or doctor who generates a bad side-effect will make predictions about future behavior that differ more from baseline predictions than will participants who learn about the CEO or doctor who generates a good side-effect. It's important to consider the *magnitude* of deviations from baseline predictions and not absolute predictions, because all models might predict that norm-violators are more likely to violate norms and norm-conformers to conform to norms; the issue is whether these behaviors provide an equivalent amount of information concerning underlying mental states.

Participants

Participants were 156 University of California-Berkeley undergraduates (70% women; mean age = 20, s.d. = 3) who participated for course credit.

Materials and Procedure

Participants were randomly assigned to one of three conditions: *baseline*, *norm-conforming*, or *norm-violating*. Participants in the *norm-conforming* and *norm-violating* conditions were presented with two short vignettes, the CEO vignette (Knobe, 2003a) from the introduction as well as the analogous DR vignette:

> **DR Vignette:** A team of doctors is treating a patient. One doctor on the team came to the senior doctor and said, "We are thinking of starting a treatment. It will lower the patient's blood pressure but it will also help [hurt] the patient's stomach problems."
>
> The senior doctor answered, "Stomach problems are not our concern. I just want to lower the patient's blood pressure as much as I can. Let's start the treatment."
>
> They started the treatment. Sure enough the patient's stomach problems were helped [hurt].

After each vignette participants were asked to make two ratings about the future actions of the agent in the story, a *specific* prediction and a *general* prediction. These questions are below, with the text for the CEO vignette in brackets:

> **Specific prediction:** In the following month the doctor [chairman] will make another decision that results in either:
> **A.** An action that has a positive consequence beyond what the doctor is treating. [that helps the environment]
> Or **B.** An action that has a negative consequence beyond what the doctor is treating. [that harms the environment]
> Which decision do you think the doctor [chairman] will make?
> **General prediction:** The next month the doctor [chairman] will make another decision that results in either:
> **A.** Exceeding ethical standards.
> Or **B.** Violating ethical standards.
> Which decision do you think the doctor [chairman] will make?

Subjects rated the likelihood of each event on a scale from 1 to 7, where 1 indicated "very likely to choose A," 4 "equally likely to choose A or B," and 7 "very likely to choose B."

Participants in the *baseline* condition were introduced to the agents (e.g. "There is a chairman of the board who makes the final decisions for his company") and made all four prediction judgments, but were given no information about the agents' past behavior.

The order of story presentation (CEO first or DR first) as well as the direction of the 7-point scale (from conforming to violating or vice versa) was counterbalanced across participants.

Results and Discussion

To examine whether participants' prediction ratings varied across conditions, the data were analyzed in an ANOVA with condition as a between-subjects variable (*baseline*, *norm-conforming*, *norm-violating*), vignette as a within-subjects variable (CEO, DR), and prediction question as a within-subjects variable (specific, general). This revealed a main effect of condition ($F_{(2,153)}=14.36$, $p<.001$), as well as a main effect of vignette ($F_{(1,153)}=83.43$, $p<.001$). Overall, participants rated negative actions more probable in the *norm-conforming* condition than in the *baseline* condition, and in the *norm-violating* condition than in the *norm-conforming* condition (see Fig. 2). Ratings in the *norm-conforming* condition may have been more negative than in the *baseline* condition because failing to endorse a fortuitous side effect (e.g. helping the environment) is itself a norm violation (see Mele & Cushman, 2007). The main effect of vignette resulted from the fact that predictions concerning the CEO were generally more negative than those concerning the doctor.

The key hypothesis that predictions in the *norm-violating* condition should differ more from baseline than do those in the *norm-conforming* condition can be examined by looking for significant differences across these conditions, as both yielded more negative ratings than *baseline*. An ANOVA like that above but restricted to the *norm-violating* and *norm-conforming* conditions reproduced the main effect of vignette ($F_{(1,102)}=50.86$, $p<.001$) and revealed a main effect of condition ($F_{(1,102)}=8.75$, $p<.01$) as well as a 3-way interaction between vignette, prediction, and condition ($F_{(1,102)}=4.80$, $p<.05$). With post-hoc t-tests, the *norm-conforming* and *norm-violating* conditions differed significantly on both CEO predictions (specific: $t(102)=3.43$, $p<.001$; general: $t(102)=2.18$, $p<.05$), and were suggestive for the DR predictions (specific: $t(102)=1.11$, $p=.271$; general: $t(102)=1.91$, $p=.059$).[1] These findings confirm the prediction that norm-violating behavior provides more information about an agent's future behavior than norm-conforming behavior.

According to the inferential view, participants use behavior, including its relationship to norms, as a source of evidence concerning the agent's mental states. Mental state

---

[1] To verify that the DR vignette generates a side-effect effect, a different group of 72 participants was randomly assigned to either the CEO or the DR vignette in a condition involving either a helpful or a harmful side-effect. On a 7-point scale, participants judged whether it was appropriate to say that the agent *intentionally* brought about the side-effect. This experiment revealed a main effect of condition ($F_{(1, 68)} = 121.5$, $p < .001$) as well as an interaction between condition and vignette ($F_{(1, 68)} = 9.82$, $p = .003$). The help/harm asymmetry was smaller for the DR (2.3 for help versus 4.5 for harm) than for the CEO (1.4 for help versus 5.3 for harm), but even the DR vignette involved a significant effect of condition ($t(34) = 5.13$, $p < .001$).

ascriptions (e.g. "dislikes the environment") in turn generate different predictions. If this is the case, then participants in Knobe's original demonstration of the side-effect effect (2003a) and in subsequent studies may not have generated asymmetrical judgments about intentional action as a result of the moral valence of the side-effect per se, but rather because the moral valence of the side effect provided evidence about the nature of the *behavior* leading to the side effect, which in turn supports mental state ascriptions. This is examined in Experiment 2.
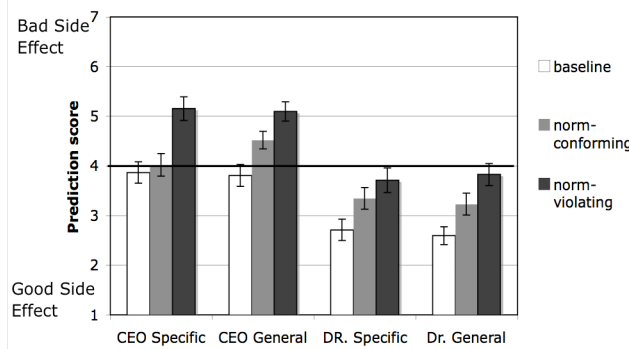

Figure 2: Prediction scores from Experiment 1 on a scale from 1 (good side effect likely in future) to 7 (bad side effect likely in future).

Experiment 2

According to the inferential view, the moral valence of an outcome (good or bad) can impact ascriptions of intentional action because outcomes that differ in moral valence provide different evidence concerning underlying mental states (see Fig. 1d). Critically, the link between outcomes and mental states must be mediated by inferences about the agent's *behavior*. Experiment 2 examines the role of an agent's behavior in generating the side-effect effect by varying the nature of the agent's behavior – in this case, a decision to pursue a given plan – while holding constant the outcome of that behavior – in this case, environmental help or harm. This was achieved by adding probability information to the CEO vignettes. A decision to pursue a plan with a 95% chance of environmental harm is a greater norm violation than a decision to pursue the same plan with a 5% chance of harm, and should be correspondingly less informative about the agent's mental states, even if both cases generate harm (see Nadelhoffer, 2006 for a related manipulation).

Participants

There were 104 participants (58% women; mean age = 22, s.d. = 4), 74 University of California-Berkeley undergraduates participating for course credit and 30 individuals who participated online.

Materials and Procedure

Participants were randomly assigned to one of four versions of the CEO vignette: *norm-conforming* or *norm-violating*

crossed with 5% versus 95%. Unlike Experiment 1, the chairman was told the probability of the environmental side effect occurring before making a decision. This was indicated in the first paragraph:

> The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. It will help us increase profits, but there is a 5% [95%] chance that it will also help [harm] the environment."

In all conditions the side effect, environmental help or harm, occurred. Participants were then asked, "How appropriate is it to say the CEO intentionally helped [harmed] the environment?", and provided a rating on a scale from 1 to 7, where 1 indicates "not at all appropriate," 4 "neither appropriate or inappropriate," and 7 "Very appropriate."

Results and Discussion

A 2 (norm-conforming, norm-violating) X 2 (5%, 95%) ANOVA revealed a significant effect of norm status ($F(1,100)=63.64$, $p<.001$), a marginal effect of probability ($F(1,100)=2.78$, $p=.099$), and a significant interaction ($F(1,100)=4.10$, $p<.05$; see Fig. 3). Post-hoc tests comparing the 5% and 95% groups revealed significant differences in the *norm-violating* condition (5%-4.31, sd=1.9; 95%-5.5, sd=1.42; $t(50)=-2.53$, $p<.01$) but not in the *norm-conforming* condition (5%-2.38, sd=1.49; 95%-2.27, sd=1.68; $t(50)=.26$, $p=.80$).

These results suggest that participants' ascriptions of intentional action are not simply a function of the *outcome* of an action, but rather of the *behavior* that generated that outcome, which itself reflects mental states.
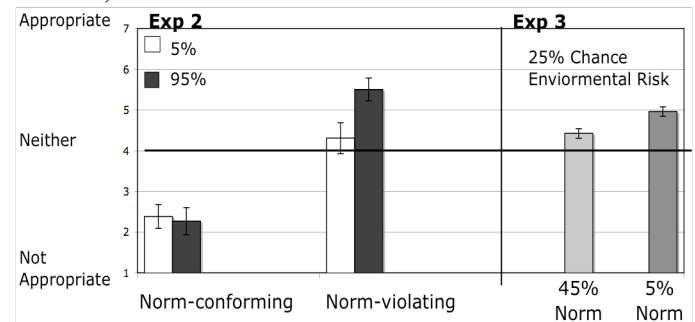

Figure 3: Experiment 2 and 3 ratings of how appropriate it is to call an action intentional as a function of probability and norm status. Ratings were made on a scale from 1 (not appropriate to say outcome brought about intentionally) to 7 (appropriate to say outcome brought about intentionally) with 4 (neither appropriate nor inappropriate) as a midpoint.

Experiment 3

Experiments 1 and 2 suggest that norm-violating behavior is more informative than norm-conforming behavior, and that this asymmetry derives from the moral valence of an agent's behavior, not only its outcome. However, these experiments

do not provide direct evidence for the role of *norms* as opposed to a specific asymmetry between moral goodness and badness, and hence do not provide support for this aspect of the inferential model (see Fig. 1). Experiment 3 investigates whether the asymmetry in the side-effect effect indeed reflects the relationship between behavior and norms. To accomplish this, the CEO vignette involving harm was modified to incorporate a probability, as in Experiment 2, and information about an industry norm for acceptable risk. Specifically, participants read about a CEO who pursued an action with a 25% chance of causing environmental harm, but where the industry standard for pursuing a plan with environmental risk was either 45% or less (making the behavior *norm-conforming*) or 5% or less (making the behavior *norm-violating*). As in Experiment 2, the negative side effect always occurred, but in addition, the probability of harm was constant across conditions. The only difference across conditions was the status of the chairman's decision with respect to the norm. If the inferential model is correct, participants should judge the action more intentional in the *norm-violating* case than in the *norm-conforming* case, no matter that both cases involve the same morally bad side-effect.

Participants

Participants were 431 University of California-Berkeley undergraduates (65% women; mean age = 20, s.d. = 3) who received the questionnaire as part of a larger online packet completed for course credit. There were 218 participants in the *norm-conforming* condition and 213 participants in the *norm-violating* condition.

Materials and Procedure

Participants were randomly assigned to either a *norm-conforming* (45%) or a *norm-violating* condition (5%) involving a modified CEO vignette. The vignettes always involved a bad side effect with a 25% probability of occurring, but involved different industry standards:

A regulatory agency for the Gizmo industry exists in order to provide environmental standards even though it does not have the authority to ensure compliance with these standards. This regulatory agency has established an environmental standard, which states that a company may only start new programs if the chance of environmental harm due to the program is under 5% [45%].

The vice-president of a company in the Gizmo industry went to the chairman of the board and said, "We are thinking of starting a new program. It will help us increase profits, but there is a 25% chance that it will also harm the environment. The industry standard is to only start programs of this type when the chance for harm is under 5% [45%]."

The environmental harm always occurred. Participants were then asked, "How appropriate is it to say the CEO intentionally harmed the environment?" and provided ratings as in Experiment 2.

Results and Discussion

Subjects in the *norm-violating* (5%) condition generated significantly higher ratings of intentional action (4.96, sd=1.66) than did subjects in the *norm-conforming* (45%) condition (4.42, sd=1.73; $t(429)=-3.28$, $p<.01$; see Fig. 3). This suggests that in evaluating whether an action is intentional, participants consider the relationship between behavior and norms, and not merely the behavior. These results are predicted by the inferential model, and make a strong case for the claim that moral considerations determine whether behavior is norm-conforming or norm-violating, which in turn informs mental state ascriptions.

General Discussion

These studies demonstrate the predictive value of norm-violating behavior (Experiment 1), and highlight the role of an agent's behavior (Experiment 2) and its relationship to norms (Experiment 3) in mental state ascriptions. Participants use norms as a point of reference to determine what outcomes and actions reveal about an agent's mental states, which in turn inform ascriptions of intentional action. So while there does seem to be an influence of moral evaluation in ToM, the relationship may be best described as *evidential*.

These studies also lend support to the inferential model (see Fig. 1). While other models could be modified to accommodate these findings, the inferential model has the advantage of specifically predicting these results. Moreover, the inferential model can accommodate several cases in the literature that have proved difficult for other accounts of the side-effect effect. Because the model allows for multiple sources of predictive information and emphasizes the relationship between an action and norms, it is equipped to deal with cases involving unusually motivated (Mallon, 2008) or conflicted (Phelan and Sarkissian, 2006) agents.

The inferential model preserves the traditional functions of ToM, though additional functions are certainly possible. Importantly, the inferential model also emphasizes a role for normative information in prediction and explanation. Developmental research has suggested that for children under the age of four, moral and conventional norms are a primary basis for explaining and predicting behavior (Kalish, 1998). For example, young children predict that an agent will conform to a norm, even if the norm is unknown to the agent or conflicts with the agent's desires (Kalish and Shiverick, 2004). Even in adults, not all ToM judgments are automatic (Apperly, 2006); it's possible that norms support many everyday predictions and explanations, with mental state inferences drawn only as needed.

Recognizing a role for norms in mental state ascriptions raises a number of important questions. For example, is the influence of norms on mental state ascriptions restricted to prescriptive norms, such as the conventional and moral norms considered here? We suspect a similar relationship

holds for statistical norms. A behavior that violates a statistical norm is not 'expected', and hence provides information about the agent's underlying mental states that differs from the default. If most people conform to a norm not to harm the environment, for example, observing someone avoid environmental harm is relatively uninformative: the behavior would have been predicted from the statistical norm. On the other hand, observing an agent violate this norm by harming the environment *is* informative: rather than ascribing default mental states, we ascribe an atypical (in this case negative) attitude towards the environment (see Lucas et. al., in press, for a similar argument with respect to preference attribution). As with prescriptive norms, this makes sense if the function of ToM is to track information that supports prediction and explanation.

A related question concerns the interaction between multiple norms. While many moral norms are also statistical norms, there may be cases in which norm-conformance is rare, placing a moral norm in conflict with a statistical norm. How are mental state ascriptions made under such conditions? These cases may be uncommon because a moral norm would presumably be the statistical norm unless conformance had a cost. But as an illustrative example, consider the low-cost behavior of agreeing to donate one's organs in case of accidental death. Though it is generally believed that organ donation is morally good (*morally norm-conforming*), actual organ donor rates in the US are not very high (*statistically norm-violating*). In this case, it may be possible to see a reversal of the typical side-effect effect, where the morally good behavior (organ donation) is more informative and judged intentional.

While we've contested Knobe's (2003, 2006) interpretations of the side-effect effect as a challenge to the traditional function of theory of mind, our findings support the underlying claim that moral (and other) norms influence mental state ascriptions. The key lesson from our arguments and findings is that sensitivity to norms may improve rather than imperil our ability to predict and explain behavior, and accordingly deserves more attention in the study of social cognition (see also ; Kalish 2006; Wellman & Miller, 2006).

Acknowledgments

References

Adams, F. and Steadman, A. (2004a). Intentional action in ordinary language: Core concept or pragmatic understanding? *Analysis*, 64, 173-181.

Apperly, I.A., Riggs, K.J., Simpson, A., Chiavarino, C. and Samson, D. (2006). Is belief reasoning automatic? *Psychological Science,* 17, 841-844.

Cal. Penal Code § 187

Gopnik, A. (1999). Theory of mind. In R. Wilson & F. Keil (Eds*). The MIT Encyclopedia of the Cognitive Sciences* Cambridge, MA: M. I. T. Press.

Kalish, C.W. (1998). Reasons and causes: Children's understanding of conformity to social rules and physical laws. *Child Development*, 69, 706-720.

Kalish, C.W. (2006). Integrating normative and psychological knowledge: What should we be thinking about? *Journal of Cognition and Culture*, 6, 191-208.

Kalish, C.W. & Shiverick, S.M. (2004). Children's reasoning about norms and traits as motives for behavior. *Cognitive Development*, 19, 401-416.

Knobe, J. (2003a). Intentional action and side effects in ordinary language. *Analysis. 63,* 190-193.

Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies*, 130, 203-231.

Knobe, J. & Mendlow, G. (2004). The good, the bad, and the blameworthy: Understanding the role of evaluative considerations in folk psychology. *Journal of Theoretical and Philosophical Psychology 24*, 252-258.

Lucas, C., Griffiths, T. L., Xu, F., & Fawcett, C. (in press). A rational model of preference learning and choice prediction by children. *Advances in Neural Information Processing Systems*, 21.

Malle, B. F., & Nelson, S. E. (2003). Judging Mens Rea: The Tension Between Folk Concepts and Legal Concepts of Intentionality. *Behavioral Sciences and the Law, 21*, 563-580.

Mallon, R. (2008). Knobe vs. Machery: Testing the trade-off hypothesis. *Mind & Language*, 23, 247-255.

Mele, A. (2001). Acting intentionally: Probing folk notions. In B. F., Malle, L. J. Moses, & D. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition*. Cambridge, MA: M. I. T. Press.

Mele, A.R. & Cushman, F. (2007). "Intentional action, folk judgments, and stories: Sorting things out." *Midwest Studies in Philosophy*, 31, 184-201.

Nadelhoffer, T. (2004b). Praise, side effects, and intentional action. *The Journal of Theoretical and Philosophical Psychology*, *24*, 196-213.

Nadelhoffer, T. (2006). Bad acts, blameworthy agents, and intentional actions: Some problems for jury impartiality. *Philosophical Explorations*, 9, 203-220.

Nado, J. (2008). Effects of moral cognition on judgments of intentionality. *The British Journal for the Philosophy of Science*, 59, 709-731.

Phelan, M. & Sarkissian, H. (2006). The folk strike back; or, why you didn't do it intentionally, though it was bad and you knew it. *Philosophical Studies*, 138, 291-298.

Searle, J.R. (2001). Rationality in action. Cambridge, MA: M. I. T. Press.

Wellman, H.M. and Miller, J.G. (2006) Developing conceptions of responsive intentional agents. *Journal of Cognition and Culture*, 6, 27-55.