Actor-Observer Differences in Intentional Action Intuitions

Adam Feltz (ADFeltz@schreiner.edu)
Philosophy and Interdisciplinary Studies, 2100 Memorial Blvd.
Kerrville, TX 78028, USA

Maegan A. Harris (MAHarris@schreiner.edu)
Schreiner University, 2100 Memorial Blvd.
Kerrville, TX 78028, USA

Ashley L. Perez (ALPerez@schreiner.edu)
Schreiner University, 2100 Memorial Blvd.
Kerrville, TX 78028, USA

Abstract

Empirically minded researchers have begun exploring the "folk" notion of intentional action, often with surprising results. In this paper, we extend these lines of research and present new evidence from a radically new paradigm in exploration of folk intuitions about intentional action. Our results suggest that in some circumstances people make strikingly different judgments about intentions and intentionality as a function of whether the person brings about or observes an event. Implications for action theory and the experimental study of folk intuitions are discussed.

Keywords: Experimental Philosophy; Intentional Action; Actor-Observer Differences; Side-Effect Effect

Determining whether a person's behavior was intended or intentional is crucial for a host of important judgments such as assigning blame and praise. This part of human experience has been of central concern for philosophers of action (Mele, 1992). Many of these philosophers take themselves to be exploring the everyday or "folk" concept of intentional action (Adams, 1986; McCann, 1986, 2005; Mele, 1992). Some philosophers even write that "a philosophical analysis of intentional action that is wholly unconstrained by that [folk] concept runs the risk of having nothing more than a philosophical fiction as its subject matter" (Mele, 2001, 27). Empirically minded researchers (e.g., experimental philosophers) have helped shed light on this folk notion of intentional action, often with surprising results. In this paper, we extend these lines of research and present evidence using a new paradigm to study folk intuitions about intentional action. Our results suggest that in some circumstances people make strikingly different judgments about intentions and intentionality partially as a function of whether a person brings about or observes an event. Implications for traditional action theory and the study of philosophically relevant folk intuitions are discussed.

Experimental Philosophy and Action Theory

Arguably the best known studies in the experimental investigation of intentional action intuitions are Knobe's (2003a) harmful (underlined) and helpful (bracketed) chairman cases:

Harm/Help: The vice-president of the 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 [and] it will also harm [help] the environment." The chairman of the board answered, "I don't care at all about harming [helping] the environment. I just want to make as much profit as I can. Let's start the new program. They started the program. Sure enough, the environment was harmed [helped]. (191)

The only difference between the two cases is the moral valence of the consequence of the chairman's decision. Remarkably, this shift in the moral valence of the consequence drastically changed people's intentionality judgments about the consequence: 82% of participants judged that the chairman brought about the harm to the environment intentionally whereas only 23% judged the chairman brought about the help to the environment intentionally. This general effect (the side effect-effect or the Knobe effect) has been replicated using similar scenarios (Cushman and Mele, 2008; Knobe, 2003a, 2003b, 2004a, 2004b) across cultures (Knobe and Burra, 2006), as well as across ages (Leslie, Knobe, and Cohen, 2006).[1]

Knobe-style cases feature *side effects*. If a consequence of an intended action is foreseen but not intended, then that consequence is a side effect of the intended action.[2] Side effects have been considered important test cases of some theories of intentional action. Just to take one example, Knobe-style cases have been argued to challenge a prominent view in intentional action—the Simple View (SV). According to the SV, if an agent intentionally performs an action *A* then the agent intends to *A*. Some philosophers have argued that the SV is supported by folk intuitions (Adams, 1986; McCann 1986, 2005). However, as judgments in Harm suggest, sometimes the folk make judgments that are contrary to the SV. If the harm to the environment is a side effect, then it is

[1]See Feltz (2007b) for a more detailed overview.

[2]See Cushman & Mele (2008) for a detailed definition of a side effect.

not intended. But, most people think that the harm is brought about intentionally. Hence, in some circumstances, many have the intuition that one can harm the environment intentionally without intending to do so. This pattern of intuitions seemingly falsifies that the SV is supported by folk intuitions (Nadelhoffer, 2006).

In the next two sections, we suggest that folk intuitions surrounding intentional action may be much more complicated than originally thought and may be influenced by a variety of factors including one's perspective.

Actor-observer differences

Actor-observer differences refer to a common effect where people who engage in behaviors (actors) see things differently than those who watch behaviors (observers). The traditional conception of the actor-observer asymmetry posits that the "actor's view of his behavior emphasizes the role of environmental conditions at the moment of action. The observer's view emphasizes the causal role of stable dispositional properties of the actor" (Jones & Nisbett, 1972, 80). While it is debatable whether this traditional conception is completely accurate (Malle, Knobe, & Nelson, 2007), some actor-observer asymmetries have been revealed in decisions made in risky environments (Fernandez-Duque & Wifall, 2007), moral judgments (Nadelhoffer & Feltz, 2008), and action explanations (Malle & Knobe, 1997; Malle, Knobe, & Nelson, 2007). To illustrate, consider one case from Nadelhoffer and Feltz (2008) where an actor-observer asymmetry was found:

Trolley: A trolley is hurtling down the tracks. There are five workers on the track ahead of the trolley, and they will definitely be killed if the trolley continues going straight ahead since they won't have enough time to get out of harm's way. There is a spur of track leading off to the side where another person is working. The brakes of the trolley have failed and there is a switch which can be thrown to cause the trolley to go to the side track. Imagine that you are an innocent bystander who happens to be standing next to the switch. You realize that if you do nothing, five people will definitely die. On the other hand, you realize that if you throw the switch, you will definitely save the five workers. However, you are also aware that in doing so the worker on the side track will definitely be killed as the result of your actions.[3]

Observers received the same scenario except 'you' was replaced with 'John' (along with appropriate verb conjugations). Participants were asked if flipping the switch is morally permissible and rated how much control over the situation one has. People given the 'John' version were more likely than those given the 'you' version to judge (a) that flipping the switch was "morally permissible" and (b) that John had control over the events.[4]

But why do actors and observers sometimes display this asymmetry? According to Malle, Knobe, & Nelson, one reason is that "we can expect that actors normally have better access to their own reasons than observers do and that they are normally more motivated to portray themselves as active, conscious, and rational agents" (2007, 508). Hence, because actors may be motivated to (a) portray themselves in a positive light and (b) have special access to their own reasons, they are prone to judge their own behaviors differently from others' behaviors. This explanation could account for the asymmetry in the Trolley example. Because actors are motivated to portray themselves in a positive light and flipping the switch results in the awful killing of a person, they are less likely to judge it permissible for them to flip the switch. However, because they are relatively less interested in portraying others in a positive light, they judge that it is permissible for others to flip the switch. But actors who realize that flipping the switch is the optimal decision even if it kills a person may excuse themselves by judging they had no control over the situation.

Given that there are actor-observer differences in a wide variety of contexts, we thought that similar actor-observer differences would be found in judgments about intentions and intentionality. In our first experiment, we used a new method in the study of folk intuitions about intentional action borrowed from experimental economics. We had participants engage in a real decision making process with real rewards and penalties. Because participants *actually* became actors, we hypothesized this methodology would have the greatest chance of revealing actor-observer differences in intentional action intuitions.

Experiment 1

We constructed a decision making environment where participants could (a) engage in helpful and harmful behaviors and (b) observe others' helpful and harmful behaviors. We call *Actors* those who generate a behavior. We call *Observers* those who watch a behavior. In the *Harm* condition, an actor generates a harm to one other person. In the *Help* condition, an actor generates a benefit to one other person. We hypothesized that actors would judge behaviors as (a) less intended and (b) less intentional than when they judge behaviors as observers.

Participants

Participants (N = 40) were recruited via email at a small southern university.[5] Participants were tested in 6 groups consisting of no more than 12 participants and no fewer than 4. Participants received $10 for attending. They also had the opportunity to earn an additional $10 depending on their performance in the experiment (Range = $16-$20). Participants were told that they would be paid as a function how many Experimental Currency Units (ECUs) they earned in the experiment. The payoff function was not disclosed.

[3]These scenarios modified cases used by Petrinovich and O'Neil (1996), but Trolley Problem cases are well known in the literature.

[4] Ninety percent in the 'John' version thought it was permissible versus 65% in the 'you' version. Also, the mean control rating in the 'John' version was 4.28 and 5.12 for the 'you' version (on a 7 point ascending scale).

[5]The expense of the experiment necessitated a small sample size.

Each participant was an actor and an observer (counterbalanced for order). However, each participant was in only one of the Help or Harm conditions. Because we were interested in intuitions about actions, all participants who did not perform the desired action (contributing to Account A, see below) were excluded. Five participants were thereby excluded in Harm. For the purposes of analyzes, there were 20 participants in Help and 15 in Harm.

Methods and Materials

Participants completed the experiment on a computer programed using Z-Tree software (Fischbacher, 2007). Actors in the Harm condition were instructed to indicate how many of their 10 "tokens" they wished to invest in 'Account A'. They were told that for every token they invested in Account A, they would earn 12 ECUs. For every token they did not invest in Account A, they would earn 10 ECUs. However, for every token invested in Account A, they would generate a 3 ECU penalty to one other person in the experiment. Actors in Help were given the same instructions as Actors in Harm but instead of generating a 3 ECU penalty, the actor generated a 3 ECU bonus by contributing to Account A. Observers in Harm read a display indicating that somebody else had contributed 10 tokens to Account A generating a 30 ECU penalty to them. Observers in Help read a display stating that another participant contributed 10 tokens to Account A generating a 30 ECU bonus for them. There was one unpaid practice round followed by one paying round in each condition.

After each instance of acting or observing, participants were asked to rate on a 7 point scale (1=disagree, 7=agree) their level of agreement with the appropriate version of each of the following sentences: 1. You/the other participant intended to generate the penalty/bonus; 2. You/the other participant intentionally generated the penalty/bonus; 3. You/the other participant are/is blameworthy/praiseworthy for generating the penalty/bonus. Participants were also given the opportunity to explain their answers in a few sentences. So, each participant answered 3 actor questions and 3 observer questions in only one of Harm or Help conditions and had the opportunity to explain their answers in each condition.

Results and Discussion

To test our hypothesis, a mixed-model Analysis of Variance (ANOVA) was preformed with Harm/Help and observer order as between participants variables and answers to the Actor Intended and Observer Intended prompts as within participants variables.

Table 1: Actor/Observer

Actor Intended	$M = 3.3$, $SD = 2.08$
Actor Intentionally	$M = 3.5$, $SD = 2.17$
Observer Intended	$M = 4.0$, $SD = 2.8$
Observer Intentionally	$M = 4.11$, $SD = 2.15$

The predicted difference in actor/observer judgments was found for intention judgments, F (1, 31) = 4.51, p =. 04, η_p^2 = .13. Neither order F (1, 31) = 1.12, p =. 29, η_p^2 = .04 nor condition interacted with judgments $F < 1$. A similar mixed-model ANOVA found the predicted differences in intentionality judgments, F (1, 31) = 4.14, p =. 05, η_p^2 = .12. Order did not interact with judgments, $F < 1$.

Theoretically, there should be differences in people's Harm and Help judgments (Knobe, 2003a) and a moderately sized non-significant trend toward an interaction for Harm/Help was observed, F (1, 31) = 2.61, p = .12, η_p^2 =.08. To help illuminate these possible differences, each condition (Harm or Help) was selected and four mixed-model ANOVAs were conducted with order as between participants factors and judgments about (1) Actor Intention/Observer Intention and (2) Actor Intentional/Observer Intentional as within participants factors.

In Harm, predicted differences were found for Intention judgments, F (1, 13) = 5.63, p= .03, η_p^2 = .3. Order did not interact with judgments F (1, 13) = 1.05, p = .33, η_p^2 = .07. Predicted differences were also found for Intentional judgments F (1, 13) = 9.15, p = .01, η_p^2 = .41. Order did not interact with judgments ($F < 1$). In Help, no actor-observer differences were detected (all F's < 1).

Table 2: Harm Actor/Observer

Actor Intended	$M = 2.13$, $SD = 1.13$
Actor Intentionally	$M = 2.5$, $SD = 1.96$
Observer Intended	$M = 3.2$, $SD = 2.01$
Observer Intentionally	$M = 3.73$, $SD = 2.25$

This experiment also allowed us to explore some other possibly interesting actor-observer differences. We thought that actors would display a reversed side effect-effect while observers would display the traditional side effect-effect. As side effects can occur when a behavior is judged intentional but not intended, we selected only those participants who did not judge the behavior in the relevant condition to be intended. After excluding those who did not intend the behavior (responding 4 or less), 14 participants remained in Harm and 10 remained in Help. A univariate ANOVA indicated the predicted shift in judgments in Harm that trended toward significance: Harm $M = 2.35$, $SD = 1.2$, Help M = 3.1, SD = 1.97, F (1, 23) = 2.46, p =. 13, η_p^2 = .11. However, order appeared to interact with judgments, F (1, 23) = 2.46, p = .13, η_p^2 = .11. To eliminate any possible order effect, only first responses were analyzed.[6] After eliminating those who were in the actor condition second, did not contribute to Account A, and responded that they intended the bonus or penalty, a very large marginally significant dif-

[6] Participants could not go back to the previous condition after they had entered their answers. Once participants gave their actor judgments, they could not go back and change them after they entered the observer condition.

ference was observed: Harm (N = 7, M = 2.43, SD = 1.9), Help (N = 3, M = 5.0, SD = 1.0), F (1, 8) = 4.68, p = .06, η_p^2 = .37.[7] However, we did not find the predicted side effect-effect for observers (all $F's$ < 1).

Finally, previous research indicates that some intentional action intuitions are predictable by the global personality trait extraversion (Cokely & Feltz, 2009a). Extraversion is a member of the Big Five personality model and is represented in almost all modern personality models (John, 1999). The current experiment allowed us to test for possible actor-observer differences in relation to extraversion. To this effect, participants also completed the Brief Big Five Inventory at the end of the experiment (Gosling, Rentfrow, & Swan, 2003). Extraversion was negatively correlated with judgments in Harm for Actor Intention, r (15) = -.64, p = .01 and Actor Intentionally, r (15) = -.55, p = .03 but was not correlated with Observer Judgments p < .05. To illustrate the difference, a rough median split of extraverts was created. Those who were relatively more introverted (scoring 9 or less) were more likely than extraverts (scoring higher than 9) to respond that they intended (Introverted M = 2.63, SD = 1.3, Extraverted M = 1.57, SD = .53, F (1, 14) = 3.92, p = .07, η_p^2 = .26) or intentionally (Introverted M = 3.5, SD = 2.2, Extraverted M = 1.29, SD = .49, F (1, 14) = 5.67, p = .04, η_p^2 = .34) brought about the harmful behavior. Order did not interact with judgments ($F's$ < 1). Of note, there was a strong overall correlation of intention and intentionality judgments: Other Intention/Other Intentional r (35) = .83, p < .001, Self Intention/Self Intentional, r (35) = .77, p < .001.

Experiment 2

Experiment 1 suggested that providing the right environment could engender an Actor-Observer difference in judgments about intentions and intentionality. However, a question remains whether Actor-Observer differences can occur in traditional pencil-and-paper surveys where participants are asked to imagine themselves in the role of the chairman. To address this possible worry that the effect found in Experiment 1 is not the result of the testing environment but rather is a more general phenomenon, Experiment 2 was designed to suggest that Actor-Observer differences are not likely to be found when participants are merely asked to imagine that they are the chairman.

Participants

One hundred and one participants were recruited from Amazon's Mechanical Turk to complete the survey for a small reward ($0.15). Participants were excluded if they reported that their first language was not English or if they failed the comprehension question. After excluding these participants, 95 remained.

[7] The small sample size and unequal cells are problematic. The small sample size in Help was anticipated because it is unlikely that good behaviors would be judged unintended by actors. See Feltz (2007a) and Nadelhoffer (2007) for a discussion.

Methods and Materials

Participants were redirected from Amazon's Mechanical Turk to complete the surveys at SurveryMonkey.com. There were four different scenarios: 1. Harm Observer, 2. Harm Actor, 3. Help Observer, and 4. Help Actor. The following were the Help and Harm cases in the Actor condition:

Actor Harm/Help: Imagine that you are the chairman of the board. The vice-president of a company comes to you and says, "We are thinking of starting a new program. It will help us increase profits for this year's balance sheet, but in ten years it will start to [harm/help] the environment." Imagine that you answered, "I don't care at all about [harming/helping] the environment. I just want to make as much profit for this year's balance sheet as I can. Let's start the new program." The program was started. Sure enough, ten years later, the environment started to be [harmed/helped].

Immediately following the scenario, participants were asked to rate their level of agreement with the following sentences (1 = strongly disagree, 4 = neutral, 7 = strongly agree):

1. You intended to [harm/help] the environment;
2. You intentionally [harmed/helped] the environment;
3. You are [blameworthy/praiseworthy] for harming/helping] the environment.

Participants were also asked the following comprehension question:

4. How long did it take before the [harm/help] began?

The following were Help and Harm in the Observer condition:

Observer Harm/Help: 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 for this year's balance sheet, but in ten years it will start to [harm\help] the environment." The chairman answered, "I don't care at all about [harming/helping] the environment. I just want to make as much profit for this year's balance sheet as I can. Let's start the new program." They started the new program. Sure enough, ten years later, the environment started to be [harmed/helped].

Immediately following the scenario, participants were asked to rate their level of agreement with the following sentences (1 = strongly disagree, 4 = neutral, 7 = strongly agree): 1. The chairman intended to [harm/help] the environment; 2. The chairman intentionally [harmed/helped] the environment; and 3. The chairman is [blameworthy/praiseworthy] for [harming/helping] the environment. Participants were also asked the following comprehension question: 4. How long did it take before the [harm/help] began?

Each participant was an Actor and an Observer in only one of the Harm (N = 46) or Help (N = 49) conditions. The order of presentation was counterbalanced. Once participants completed their responses to one condition, they could not go back and change their answers.

Results and Discussion

Univariate ANOVAs found a large Knobe-like effect for Actor judgments about intentions, F (1, 93) = 231.92, p < .

001, η_p^2 = .71. Order did not interact with judgments p > .22. A similar effect was found for Observer judgments about intentions, F (1, 93) = 64.54, p < .001, η_p^2 = .41. Order did not interact with judgments F < 1. A Knobe-like differences was also found for intentionality judgments for Actors, F (1, 93) = 369.8, p < .001, η_p^2 = .80 and Observers, F (1, 93) = 122.29, p < .001, η_p^2 = 57. Order did not reliably interact with judgments for Actors (F < 1) or Observers (p > .09).

A mixed-model repeated measures ANOVA with Actor-Observer judgments as within-participants variables and order as between participants factor did not reveal a reliable Actor-Observer in Harm, all *F's* < 1. However, a significant difference was found for judgments in Help Intention F (1, 47) = 9.77, p = .003, η_p^2 = .17 and Help Intentionally F (1, 47) = 8.46, p = .006, η_p^2 = .15.

Table 3: Means for Paper Survey

	Harm	Help
Actor Intend	*M* = 4.94 *SD* = 1.65	*M* = 1.14 *SD* = 0.54
Actor Intentionally	*M* = 5.83 *SD* = 1.36	*M* = 1.35 *SD* = 0.88
Observer Intend	*M* = 4.94 *SD* = 1.65	*M* = 1.96 *SD* = 1.94
Observer Intentionally	*M* = 5.83 *SD* = 1.24	*M* = 2.04 *SD* = 1.99

The results of Experiment 2 suggest that the Actor-Observer asymmetry produced in Experiment 1 is not likely to exist when participants are only encouraged to imagine they are the chairman.

General Discussion

Consistent with and extending previous research, our results suggest that in some circumstances people tended to judge their own behaviors differently than they judge the identical behavior of others. In addition, our evidence suggests that a well-known result in experimental philosophy—the traditional side effect-effect—can be reversed. Finally, replicating previous work (Cokely & Feltz, 2009a), extraversion was systemically and predictably related to some intention and intentionality judgments.

These results provide further evidence that impression management can play a key role in people's intention and intentionality judgments. An important clue for this interpretation comes from the results of the Harm case. Participants were much less likely to judge that they intended the Harm or intentionally brought it about compared to their judgments as observers. Presumably, participants did not want to be a "bad guy" by bringing about the bad side effect whereas they were relatively less interested in managing their impression of others. Hence, they were more motivated to respond that they did not intend or intentionally bring about the Harm. In addition, extraverts were much more likely to respond this way in Harm. Because extraverts are socially minded individuals, they would be relatively more concerned with possible social aspects of their behavior. However, because the behavior in Help is beneficial, there is less motivation to mitigate possibly negative implications of that behavior. So in Help, the responses between actors and observers would be similar.

These data also provide some important insights into the side effect-effect. We found strong correlations between people's intention and intentionality judgments. Those who favor the SV may take these as supporting data. However, defenders of the SV should be cautious for two reasons. First, correlation indicates that there is *some* relation between intention and intentionality judgments. These correlations *do not* necessarily indicate that an intention to *A* is a *necessary* condition for *A-ing* intentionally. These results are equally consistent with intending to *A* is a *sufficient* condition for *A-ing* intentionally when one *A's*—a condition that most theories of intentional action would endorse under normal conditions (e.g., no causal deviance). Second, we have some evidence that a new but equally problematic side effect-effect exists. For actors who did not think they intended to bring about the penalty or bonus, the moral valence of the consequence influenced their intentionality judgments. Specifically, participants were more likely to judge they brought about the beneficial consequence intentionally than the harmful consequence. These results suggest that at least some folk do not treat an intention to *A* as necessary for *A-ing* intentionally, contrary to the SV.

Third, our results reinforce the importance of individual differences in judgments about intentions and intentionality and provide more evidence that philosophically relevant intuitions are systematically fragmented (Feltz & Cokely, 2009; Cokely & Feltz, 2009b). Those who were extraverted were less likely to judge that they intended or intentionally brought about the penalty. Importantly, we were able to predict a priori who were likely to make those judgments. If there are predictable and systematic differences in intuitions regarding intentions and intentionality, then perhaps there is not a single folk concept of intentional action, but several (Cushman & Mele, 2008).

Finally, we would like to note one limitation of previous work in experimental philosophy that has relied on "pencil and paper" surveys. Rather than simply asking participants to respond to a scenario they read, we asked participants to perform an action and observe an action. We find that participants are less likely to think that a harm they actually bring about is intentional compared to a harm somebody else brings about to them. Hence, using this alternative method uncovered actor-observer differences in intuitions about intentions and intentionality, found an intriguing possible reversal of the side effect-effect, and provided additional evidence that folk intuitions about intentional action are predictably fragmented. We hope that the present experiments open up new methodological avenues for the experi-

mental investigation of folk intuitions about intentional action.

Acknowledgments

We would like to thank Al Mele, Joshua Knobe, Shaun Nichols, Mark Isaac, and participants at the NEH Summer Institute for Experimental Philosophy for very helpful comments.

References

Adams, F. (1986). Intention and intentional action: The Simple View. *Mind and Language, 1,* 281-301.

Cushman, F., & Mele, A. (2008). Intentional action: Two-and-a-half folk concepts? In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy*. Oxford: Oxford University Press.

Cokely, E.T., & Feltz, A. (2009a). Individual differences, judgment biases, and Theory-of- Mind: Deconstructing the intentional action side effect asymmetry. *Journal of Research in Personality, 43,* 18-24.

Cokely, E. T., & Feltz, A. (2009b). Adaptive variation in folk judgment and philosophical intuition. *Consciousness and Cognition, 18*, 355-357.

Feltz, A. (2007a). Knowledge, moral praise, and moral side effects. *Journal of Theoretical and Philosophical Psychology, 27,* 123-126.

Feltz, A. (2007b). The Knobe effect: A brief overview. Journal of Mind and Behavior. *The Journal of Mind and Behavior, 28*, 265-277.

Feltz, A., & Cokely, E.T. (2008). The fragmented folk: More evidence of stable individual differences in moral judgments and folk intuitions. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Feltz, A., & Cokely, E.T. (2009). Do judgments about freedom and responsibility depend on who you are?: Personality differences intuitions about compatibilism and incompatibilism. *Consciousness and Cognition, 18,* 342-350.

Fernandez-Duque, D., & Wifall, T. (2007). Actor/observer asymmetry in risky decision making. *Judgment and Decision Making, 2,* 1-8.

Fischbacher, U. (2008). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics, 10,* 171-178.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*, 504-528.

John, O. (1999). The Big-five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. In L. Pervin & O. P. John (Eds.), *Handbook of Personality: Theory and Research*. New York: Guilford.

Jones, E., & Nisbett, R. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. Jones, D. Kanouse, H. Kelly, R. Nisbett, S. Vallins & B. Weiner (Eds.), *Attribution: Perceiving the Causes of Behavior*. Morristown, NJ: General Learning Press.

Knobe, J. (2003a). Intentional action and side-effects in ordinary language. *Analysis, 63*, 190-194.

Knobe, J. (2003b). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology, 16,* 309-323.

Knobe, J. (2004a). Folk psychology and folk morality: Response to critics. *Journal of Theoretical and Philosophical Psychology, 24,* 270-279.

Knobe, J. (2004b). Intention, intentional action and moral considerations. *Analysis, 64,* 181-187.

Knobe, J., & Burra, A. (2006). Experimental philosophy and folk concepts: Methodological considerations. *Journal of Cognition and Culture, 6,* 331-342.

Leslie, A., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect: Theory of mind and moral judgment. *Psychological Science, 17*, 421-427.

Malle, B., & Knobe, J. (1997). Which behaviors do people explain? A basic actor-observer asymmetry. *Journal of Personality and Social Psychology, 72,* 288-304.

Malle, B., Knobe, J., & Nelson, S. (2007). Actor-observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology, 93,* 491-514.

McCann, H. (1986). Rationality and the range of intention. *Midwest Studies in Philosophy, 10,* 191-211.

McCann, H. (2005). Intentional action and intending: Recent empirical studies. *Philosophical Psychology, 18,* 737-748.

Mele, A. (1992). Recent work on intentional action. *American Philosophical Quarterly, 29,* 199-217.

Mele, A. (2001). Acting intentionally: Probing folk intuitions. In B. F. Malle, L. J. Moses & D. A. Baldwin (Eds.), *Intentions and Intentionality*. Cambridge: MIT Press.

Nadelhoffer, T. (2007). Fringe benefits, side effects, and intentional actions: A reply to Feltz. *The Journal of Theoretical and Philosophical Psychology, 27,* 801-809.

Nadelhoffer, T. (2006). On trying to save the simple view. *Mind and Language, 21,* 565-586.

Nadelhoffer, T., & Feltz, A. (2008). The actor-observer bias and moral intuitions: Adding fuel to Sinnott-Armstrong's fire. *Neuroethics, 1,* 133-144.

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology, 17,* 145-171.