

Acting intentionally and the side-effect effect: 'Theory of mind' and moral judgment

Alan M. Leslie^{a, b}
Joshua Knobe^c
and
Adam Cohen^b

^a Department of Psychology,
Rutgers University

^b Center for Cognitive Science,
Rutgers University

^c Department of Philosophy,
Princeton University

Corresponding author: Alan M. Leslie, Center for Cognitive Science, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA. email: aleslie@ruccs.rutgers.edu.

Acknowledgments: This work was supported in part by grant BCS-0079917 from the National Science Foundation to AML.

ABSTRACT

The concept of acting intentionally is an important nexus where 'theory of mind' and moral judgment meet. Preschool children's judgments of intentional action show a valence-driven asymmetry. Children say that a foreseen but disavowed side-effect is brought about 'on purpose' when the side-effect itself is morally bad but not when it is morally good. This is the first demonstration in preschoolers that moral judgment influences judgments of 'on-purpose' (as opposed to purpose influencing moral judgment). Judgments of intentional action are usually assumed to be purely factual. That these judgments are sometimes partly normative — even in preschoolers — challenges current understanding. Young children's judgments regarding foreseen side-effects depend upon whether the children process the idea that the character does not care about the side-effect. As soon as preschoolers effectively process the 'theory of mind' concept, NOT CARE THAT P, children show the side-effect effect.idea

Introduction

The development of ‘theory of mind’ (for reviews, see Leslie, 2000; Leslie, Friedman, & German, 2004; Wellman, Cross & Watson, 2001) and moral judgment (for reviews, see Killen, Lee-Kim, McGlothlin & Stangor, 2002; Turiel, 1994) have been intensively and independently researched over the last twenty years. However, there has been relatively little work relating the two domains. One exception has been understanding intention in action. Actions often have consequences that draw forth moral judgment; and whether an action is judged intentional or not influences that moral judgment.

Piaget (1932) claimed that only children over seven years could conceive the role of intention in behavior. More recently this idea has been questioned. For example, there is increasing evidence that infants already regard action as goal-directed (Baldwin, 1993; Gergely, Bekkering & Király, 2002; Gergely, Nadasdy, Csibra & Biro, 1995; Leslie, 1994; Meltzoff, 1995; Woodward, 1998, 1999). Nuñez and Harris (1998) found that three-year-olds will assign more blame for intentional behaviors, and Siegal and Peterson (1998) found that 3-year-olds can make a three-way distinction for assigning blame, differentiating whether a falsehood is uttered intentionally, because of an innocent mistake, or because of negligence. The latter studies point to an early link between ‘theory of mind’ reasoning and moral judgment, in which intentional/unintentional judgments subsequently influence good/bad judgments. These previously studied judgments run *from* ‘theory of mind’ (purpose) *to* morality (good/bad).

A recent study with adults (Knobe, 2003a) has shown that the connection can also run in the opposite direction: from morality to theory of mind. Subjects were read vignettes about an agent who brings about a foreseen side effect—an effect that the agent foresees but does not care about. Subjects were then asked whether or not the agent brought about the side effect “intentionally.” Responses appeared to be sensitive to the moral valence of the effect itself, with morally bad effects being considered intentional and morally good effects being considered unintentional. These judgments emerged with help versus harm variants of the following scenario (Knobe 2003a, 2003b, 2004).

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, and it will also *help/harm* the environment.’

The chairman of the board answered, ‘I don’t care at all about *helping/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 *helped/harmed*.

Subjects were then asked, “Did the chairman help/harm the environment *intentionally*?” Results show that adult judgments of whether or not the chairman brought about the environmental side-effect intentionally depend crucially upon whether or not the side-effect was helpful or harmful. If it was helpful, then the chairman did *not* do so intentionally. If the side-effect was harmful, then the chairman *did* do so intentionally. The asymmetry in response appears to be part of a more general pattern. Across a wide range of stimuli and methodologies, subjects tend to regard morally bad side-effects as intentional and morally good side-effects as unintentional (Knobe

2003a; 2004; Knobe & Burra, in press; Knobe & Mendlow, in press; Malle, in press; McCann 2004; Nadelhoffer, in press). We refer to this tendency as the ‘side-effect effect.’

There are asymmetries in moral judgment which appear to invoke the principle that we are morally obliged to avoid doing harm while under no or lesser obligation to do good. Studies of children eight to nine years old suggest that their moral judgment conforms to this principle (Grueneich, 1982b; Sedlak, 1979). The principle is reflected in Anglo-American jurisprudence: in general, the law requires us to avoid doing harm but does not oblige us to do good. However, this previously recognized asymmetry is entirely within the moral domain; the side-effect effect crosses domains, and in a unusual direction: from morality to purpose.

When does the side-effect effect develop? One prerequisite is that children appreciate that the actor knows the side-effect will occur. Given a large body of findings, this prerequisite should be easily met by three-year-olds (e.g., Roth & Leslie, 1998; Wimmer, Hogrefe & Perner, 1988). Three-year-olds sometimes have difficulty appreciating that an actor does *not* know something, defaulting to an assumption of shared knowledge. This default is all our task will require. A second prerequisite is that children appreciate that the actor *does not care* about the side-effect. Nothing is currently known about when children first understand *caring/not caring*. Therefore, we tested this understanding among preschoolers in Experiment 1, as a prelude to testing the side-effect effect among preschoolers in Experiment 2.

Experiment 1

Three-year-olds are able to predict the affective reaction of a person whose desire is fulfilled (happy) or unfulfilled (sad; Wellman & Woolley, 1990). But do they understand the feelings of a person who does not care about an outcome? By this age, children already appreciate that people have different likes and dislikes; indeed, even infants appreciate this (Repacholi & Gopnik, 1997). But can they understand when someone neither likes nor dislikes something—when the person simply does not care? We tested preschoolers on the tripartite distinction among liking, disliking, and not caring about food items. Characters either obtained or did not obtain an item, and the children were asked to predict how the characters would feel. Because our next experiment would involve outcomes affecting third parties, in Experiment 1 we also tested children on scenarios in which characters did and did not care about another person.

Method

Design and Procedure

Subjects were tested individually in quiet locations in local nursery schools. Toy props were used in presenting the stories. Sessions were videotaped and scored later. Subjects were tested on two story types. In the first type, a character was described as loving, hating, or not caring about a food item. In each of these conditions, the character either did or did not receive the item. The children were asked to predict how that character would feel (happy, sad, or "just okay"). In the second story type, a character was described as either caring or not caring about another character. The second character then became happy or sad, and the children were asked to predict how the first character would feel (happy, sad, or "just okay").

The children were introduced to the response scale and trained on its use. The scale showed a smiling face at one end of a black line, and a frowning face at the other extreme. At the midpoint was a neutral face, and at points in between the midpoint and extremes were a slightly smiling face and a slightly frowning face, respectively. The children were asked if they liked ice cream, told they could show how much they liked or hated it by pointing to a place on the scale, and encouraged to do so. This procedure was repeated for broccoli and for an item that they indicated was "just okay," typically water. Only children who understood and readily used the scale were tested further.

Following this training, the children were given the test stories in fixed order. Responses were scored as -2 for the frowning face (very unhappy), -1 for the slightly frowning face (a little unhappy), 0 for the midpoint (okay), +1 for the slightly smiling face (a little happy), and +2 for the smiling face (very happy). Three age groups (3-, 4-, and 5-year-olds) were tested. For the food stories, three caring conditions (love, hate, do not care) were crossed with two outcome conditions (get, do not get). For the people stories, two caring conditions (care, do not care) were crossed with two affect conditions (second character becomes happy, second character becomes sad).

Subjects

Fifty-four children were seen, but 13 were eliminated from the study, 5 for failing to cooperate, 2 for not having English as their first language, and 6 for failing scale training. Of the remaining 41 subjects, 11 (6 girls) were 3-year-olds between the ages of 41 and 48 months ($M = 45.1$, $SD = 1.9$), 13 (7 girls) were 4-year-olds between the ages of 49 and 60 months ($M = 54.1$, $SD = 3.4$), and 17 (7 girls) were 5-year-olds between the ages of 61 and 79 months ($M = 68.7$, $SD = 5.2$).

Results

For the food stories, children in the three age groups responded similarly and appropriately according to caring status and outcome (see Fig. 1). They judged that a character who loved something and got it would be happy and that a character who loved something and did not get it would be sad; they judged that a character who hated something and got it would be sad and that a character who hated something and did not get it would feel neutral; most important, they judged that a character who did not care about something would feel neutral whether or not the character got it.

Scores for the food stories were entered into a 2 (outcome) \times 3 (caring) \times 3 (age) repeated measures analysis of variance (ANOVA), with caring condition nested under outcome condition. There were main effects of outcome, $F(1, 38) = 18.6$, $p < .001$, $\eta^2 = .33$, and of caring, $F(2, 76) = 5.1$, $p = .008$, $\eta^2 = .12$, and a large Outcome \times Caring interaction, $F(2, 76) = 60.3$, $p < .001$, $\eta^2 = .61$. The main effect of age was not significant ($F < 1$), but there was an Outcome \times Caring \times Age interaction, $F(4, 76) = 3.15$, $p = .02$, $\eta^2 = .14$. The main effect of outcome reflected an overall tendency to judge that characters who got the food would be happy and those who did not would be sad, regardless of caring status. The main effect of caring reflected an overall tendency to judge that characters who hated the food would be sad, regardless of outcome. The large interaction of outcome with caring reflected affect judgments appropriate to whether the character loved, hated, or did not care about the food and whether the character received the item or not.

Inspection of the data suggested that the three-way interaction with age was produced by the judgments of the character who hated the item. We checked this by calculating the difference between sad/happy judgments in the get and not-get conditions for each subject in each of the caring conditions. These difference scores were then entered into three separate one-way (age) ANOVAs, one for each level of the caring factor. A significant effect was found only for judgments of the character who hated the item, $F(2, 38) = 3.99$, $p = .027$, $\eta^2 = .17$, all other comparisons, $F < 1$. Post hoc Bonferroni correction showed a significant difference only between the 3- and the 5-year-old groups ($p = .048$). Inspection of the mean judgment scores suggested that in the case of the character who hated and got an item, there was a marked age-related trend from predicting a little sadness (youngest group) to predicting a great deal of sadness (oldest group), and that there was a slight tendency for the oldest children to predict that the character who hated the item and did not get it would feel some happiness, whereas the other age groups thought this character would have neutral feelings. No age-related differences were found in the judgments of characters who loved or did not care about the item.

For the stories with a second character who was made happy or sad, the children in the three age groups responded similarly and appropriately according to caring status and outcome (see Fig. 2). Data were analyzed by a 2 (caring: cares, does not care) \times 2 (affect: sad, happy) \times 3 (age: 3-, 4-, 5-year-olds) repeated measures ANOVA. There was no main effect of age ($F < 1$), and age was not involved in any significant interactions. Caring also did not have a significant main effect ($F < 1$). There was a significant main effect of affect, $F(1, 38) = 11.3$, $p = .002$, $\eta^2 = .23$, reflecting a tendency to judge that overall the first character would have a greater degree of

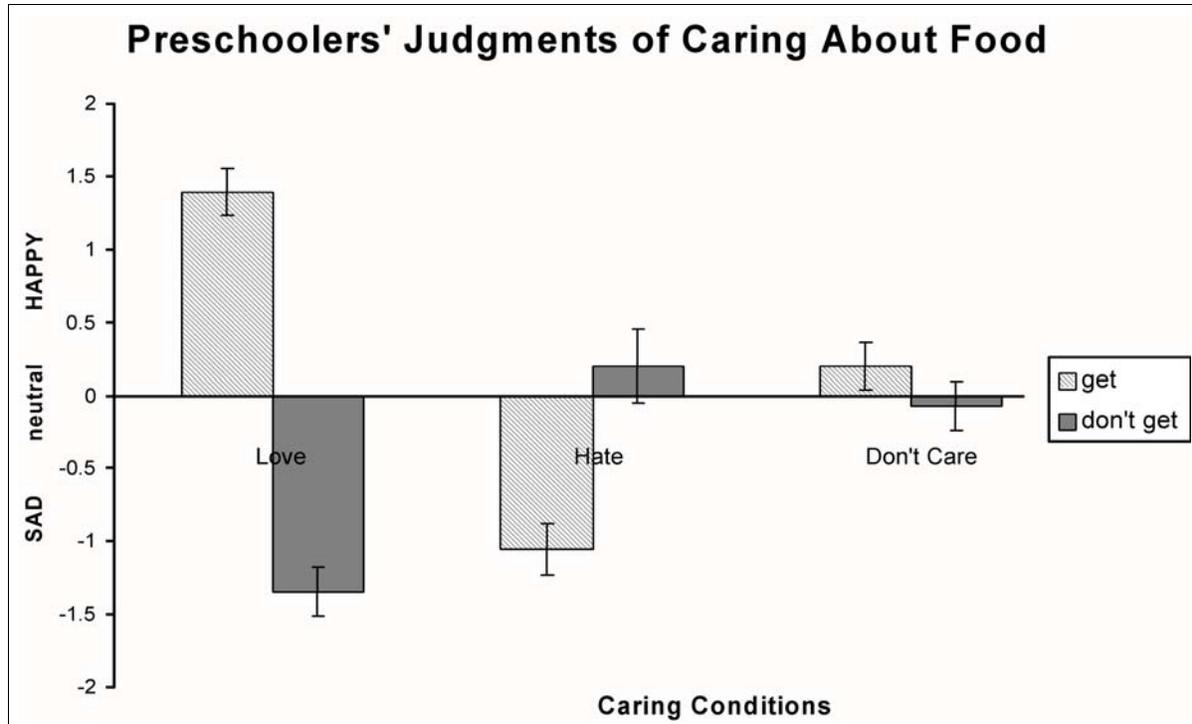


Figure 1. Preschoolers judge characters' affect appropriately depending upon both whether a character loves, hates or doesn't care about a food item and whether or not the item was obtained by the character. The only difference between 3- and 5-year-olds was in the case of *hates* and at all ages children predicted neutral affect for *don't care*.

sadness than of happiness. As expected there was a significant interaction of caring and affect, $F(1, 38) = 7.93$, $p = .008$, $\eta^2 = .17$. When the first character cared about the second character, the children predicted that the first character's affect would mirror that of the second. When the first character did not care about the second, however, the children judged that the first character would have neutral affect regardless of the affect of the second character. This pattern was observed in all age groups.

Discussion

The children appeared to judge that two characters would tend to share sadness more than happiness. They judged that one person is more likely to feel sad regarding another person's misfortune than to feel happy for another's good fortune, even when that person cares about the other. Most important for present purposes, preschool children appear to have a basic grasp of the concepts of caring and not caring. In Experiment 1, they appeared to grasp that not caring implied a more neutral affect with regard to all outcomes tested, whether they involved getting or not getting a food item or the prospect of someone being happy or sad. Even 3-year-olds appear to satisfy an important prerequisite for understanding side-effect scenarios. In our second experiment, we tested whether they would exhibit the side-effect effect.

Experiment 2

Method

Design and Procedure

Preschool children were told stories in which an actor carried out an action to bring about one effect while knowing that this action would also have a certain side effect (see the appendix). Each child heard one story. In the good condition, the side effect was to make another person happy. In the bad condition, the side effect was to make another person upset. In either case, the actor was described as not caring that the other person would be happy or upset. Toy props were used in telling the stories, and the children were asked control questions to monitor their attention to and understanding of the basic plot. If a child did not answer a control question correctly, he or she was corrected by the experimenter, and the relevant portion of the story was retold; the child's comprehension was then checked again. No child failed a control question the second time.

The children were then asked two test questions. First, they were asked the caring question: "Does [actor's name] care that [other character's name] will be/get happy/upset?" If a child failed to answer this question correctly, then the story was retold up to that point and the question was asked again. If the child's second answer was still incorrect, the child was scored as failing the caring question; otherwise, he or she was scored as passing. The second test question was the intentionality question: "Does [actor] make [other character] happy/upset on purpose?" Each child's first answer to this question was recorded as either "yes" or "no."

Subjects were tested individually in quiet locations in local nursery schools. Sessions were videotaped and scored later.

Subjects

One hundred thirty children were seen, but no data were collected from 8 children because of inattentiveness, excessive shyness, or experimenter error. Of the remaining 122 subjects, 26 (19 girls) were 3-year-olds between the ages of 39 and 48 months ($M = 43.1$, $SD = 3.1$), 49 (29 girls) were 4-year-olds between the ages of 48 and 59 months ($M = 53.6$, $SD = 3.7$), and 47 (22 girls) were 5-year-olds between the ages of 60 and 80 months ($M = 66.7$, $SD = 4.9$). The children were assigned randomly to one of the two conditions (bad, good), with approximately equal numbers in each.

Results

Figure 3 shows the percentages of children in each age group and condition who answered "yes" and "no" to the intentionality question. Three-year-olds showed a simple bias to respond "yes"; 81% said that the side effect was brought about on purpose (binomial test, $N = 26$, $x = 5$, $p = .002$, two-tailed). Four- and five-year-olds, by contrast, showed the predicted side-effect effect, answering "yes" if the side effect was harmful and "no" if the side effect was beneficial: 4-year-olds—Upton's $\chi^2(1, N = 49) = 4.61$, $p = .016$, $\phi^2 = .096$; 5-year-olds—Upton's $\chi^2(1, N = 47) = 6.2$, $p = .006$, $\phi^2 = .135$ (p s one-tailed).

Somewhat surprisingly in view of Experiment 1, a substantial proportion of subjects failed the caring question by declaring (twice) that the actor did care that the other character was happy or upset. Children who failed this question (mean age = 51.0 months, $SD = 8.8$ months) were significantly younger than those who passed (mean age = 59.7 months, $SD = 9.1$), $t(120) = 5.1$, $p < .001$, two-tailed, $d = 0.96$. We therefore reanalyzed the results separating children who failed this question from those who passed. In the youngest group, only 27% of the children

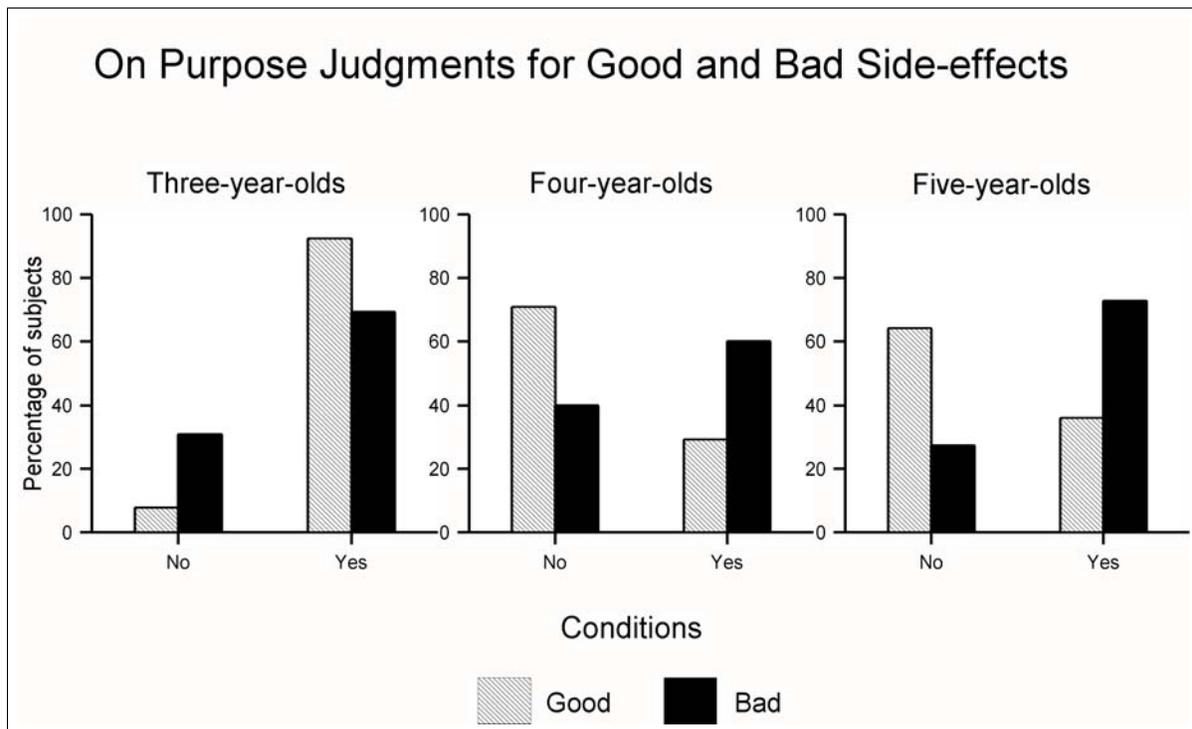


Figure 3. Proportion of subjects answering 'yes' or 'no' to the question whether a story character had caused another character to be happy (good) or upset (bad) on purpose. The youngest children showed a simple 'yes' bias. The other two groups judged 'on purpose' if the side-effect was bad and 'not on purpose' if the side-effect was good.

passed this question, and neither those who passed nor those who failed showed signs of the side-effect effect. Among 4-year-olds, 65% passed the caring question (binomial test, $N = 49$, $x = 17$, $p = .022$, one-tailed). The 4-year-olds who failed showed neither the side-effect effect (Fisher's exact test, $p > .2$, one-tailed) nor a bias to respond "yes" (binomial test, $N = 17$, $x = 7$, $p = .32$, one-tailed). Those who passed, however, showed the expected side-effect effect, Upton's $\chi^2(1, N = 32) = 2.93$, $p = .044$, one-tailed, $\phi^2 = .094$. In the oldest age group, 79% of the children passed the caring question. Those who failed this question showed no pattern in their answers to the intentionality question, whereas those who passed showed the side-effect effect, Upton's $\chi^2(1, N = 37) = 7.73$, $p = .003$, one-tailed, $\phi^2 = .215$.

Discussion

The principal finding of Experiment 2 is that 4- and 5-year-olds already exhibit the side-effect effect. In these age groups, the effect was evident only in those children (the majority) who passed the caring question. Figure 3 might suggest that only judgments concerning beneficial outcomes change with age, conceivably because children learn that "on purpose" applies only to actions with harmful outcomes and therefore start to answer "no" in the case of beneficial outcomes.¹ However, Feinfield, Lee, Flavell, Green, and Flavell (1999) found that 4-year-olds do attribute purpose to nonharmful actions. Therefore, it is unlikely that our results reflect the emergence of a general restriction of the phrase "on purpose" to bad outcomes by 4- and 5-year-olds. Instead, it is more likely that the entire side-effect effect emerges at 4.

A subsidiary finding was that most 3-year-olds failed the caring question and did not show the side-effect effect. The youngest subjects apparently found the stories too complex and resorted to a "yes" strategy in answering. The developmental shift seems to be from a general failure to process the scenarios at age 3 to the emergence of the side-effect effect at age 4, when the children became able to process that the character did not care that the side effect would occur.

General Discussion

Our principal finding is that preschool children already base attributions of intentional action on the moral valence of the relevant consequences. If the foreseen side consequences of an action are harmful, preschoolers judge that the actor has brought those consequences about on purpose, despite the fact that the actor has renounced that intention and indicated he or she does not care about those consequences. By contrast, if the foreseen side consequences are beneficial, preschoolers judge that the actor has not brought about those consequences on purpose. This adultlike pattern of judgment (Knobe, 2003a, 2003b) is already evident by 4 years of age. Before discussing this main finding further, we turn to our subsidiary findings.

Why were the results of Experiments 1 and 2 apparently discrepant with respect to whether 3-year-olds understand the concept of not caring? The stories in Experiment 2 were more complex than those in Experiment 1, and perhaps with simpler stories, 3-year-olds would have succeeded in Experiment 2 as well. Another difference is that the stories in Experiment 1 concerned whether a character cared about a food item or person, whereas the stories in Experiment 2 concerned whether a character cared that a state of affairs would occur ("care about X" vs. "care that P"). In the theory-of-mind literature, it has been claimed that concepts like

"care about X" can be grasped with a nonrepresentational understanding of mind, whereas grasping concepts like "care that P" requires a representational understanding of mind. Although previous research has not explored understanding of the concept "not caring," this argument concerning the importance of whether the object of a mental state is an entity or a proposition has been made with regard to the concept of desire (Wellman, 1990). However, as a number of authors have pointed out, to understand "he wants a sandwich" requires understanding the implicit propositional object in "he wants that he eat a sandwich" (e.g., Leslie, 1994; Nichols & Stich, 2003). The present results add a further twist to this issue, but do not resolve it.

Previous developmental research has shown various asymmetries within the domain of moral judgment (Grueneich, 1982; Sedlak, 1979), but these asymmetries cannot by themselves explain the side-effect effect described here. Moral asymmetry says nothing about how or why moral judgments should determine judgments of purpose. Law courts commonly assume that judgments of purpose are purely factual judgments to be decided by juries. The same assumption is made in the literature on the child's theory of mind. The side-effect effect suggests, however, that such judgments may sometimes be partly factual and partly moral. To the extent that such judgments are moral, theory of mind is unlike a scientific theory, and its development is not reducible to discovering matters of fact.

One key task now is to determine how the mechanisms underlying the side-effect effect develop. Given that the effect is partly moral, it cannot be the product of a domain-general process of scientific-theory construction. But the present results suggest also that it is not explained by domain-general processes of cultural transmission. First, the effect was found in subjects who were only 4 years old—an age at which children probably have not heard anyone say that a good or bad side effect was not brought about on purpose or was brought about on purpose, respectively. Second, the effect showed a surprising pattern of development. The youngest children had difficulty processing that an agent may not care about a particular side effect. Then, as soon as the children were old enough to correctly attribute not caring, they showed the pattern characteristic of the adult side-effect effect. Perhaps this pattern is not acquired gradually, but rather emerges immediately following its prerequisite.

At this point, we see two promising domain-specific hypotheses regarding underlying mechanisms. The first is that the theory-of-mind mechanism generating attributions of intentional action may have a parameter for the moral valence of outcomes (and perhaps for other kinds of valence). The value of this parameter would influence judgments of purpose, but be obtained from processes external to theory of mind, such as moral judgment. The second hypothesis is that the side-effect effect is the product of an innate "moral faculty" (Dwyer, 1999; Harman, 1999; Hauser, in press). Such a faculty could take in information about the situation and the agent's mental states. Then it could use this information to determine whether or not the behavior was morally bad and, on that basis, produce as output a judgment of whether or not it was performed intentionally. It will be interesting to see whether future research will show that the available data can be explained using only domain-general learning mechanisms, and if not, whether it will be able to distinguish between our two domain-specific hypotheses.

Appendix.

Protocol for experiment 2

Upset Condition

Here is a boy named Andy, and he's over at his house [put him on one side of the table]. And here is a girl named Janine, and she's over at her house [put her on the other side of the table]. And look what Andy has with him, he has a [frog - let child answer]. Now Andy loves frogs, but Janine hates frogs.

Now can you remember, does Andy love frogs?

Does Janine love frogs?

Andy wants to bring the frog over to Janine's house, but she will get upset.

Why will she get upset?

Now listen very carefully. Andy does *not* care that Janine will get upset. He is going to bring the frog over anyway.

Caring Question: Does Andy care that Janine will get upset?

So Andy brings the frog over to Janine's house and she gets upset.

Now I have a question for you.

On Purpose Question: Does Andy make Janine upset on purpose?

Happy Condition

Here is a boy named Andy, and he's over at his house [put him on one side of the table]. And here is a girl named Janine, and she's over at her house [put her on the other side of the table]. And look what Andy has with him, he has a [frog - let child answer]. Now Andy loves frogs, and Janine loves frogs.

Now can you remember, does Andy love frogs?

Does Janine love frogs?

Andy wants to bring the frog over to Janine's house. If he brings the frog over, she will be happy.

Why will she be happy?

Now listen very carefully. Andy does not care that Janine will be happy. He is going to bring the frog over just for himself.

Caring Question: Does Andy care that Janine will be happy?

So Andy brings the frog over to Janine's house and she is happy.

Now I have a question for you.

On Purpose Question: Does Andy make Janine happy on purpose?

References

- Baldwin, D.A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, **20**, 395–418.
- Dwyer, S. (1999). Moral competence. In K. Murasugi and R. Stainton (Eds.), *Philosophy and Linguistics*. Westview Press.
- Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, **415**, 755.
- Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, **56**, 165–193.
- Harman, G. (1999). Moral philosophy and linguistics. In K. Brinkmann (Ed.), *Proceedings of the 20th World Congress of Philosophy: Volume 1: Ethics*. Philosophy Documentation Center, 107-115.
- Hauser, M. (in press). *Moral minds: The unconscious voice of right and wrong*. New York, NY: Harper Collins.
- Killen, M., Lee-Kim, J., McGlothlin, H., & Stangor, C. (2002). How children and adolescents evaluate gender and racial exclusion. *Monographs of the Society for Research in Child Development*, **67**(4).
- Knobe, J. (2003a). Intentional action and side effects in ordinary language. *Analysis*, **63**, 190–193.
- Knobe, J. (2003b). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, **16**, 309–324.
- Knobe, J. (2004). Intention, intentional action and moral considerations. *Analysis*, **64**, 181–187.
- Knobe, J. (forthcoming). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies*.
- Knobe, J. & Burra, A. (in press). What is the relation between intention and intentional action? *Journal of Culture and Cognition*.
- Knobe, J. & Mendlow, G. (in press). The good, the bad, and the blameworthy: Understanding the role of evaluative considerations in folk psychology. *Journal of Theoretical and Philosophical Psychology*.
- 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: Cambridge University Press.
- Leslie, A.M. (2000). ‘Theory of mind’ as a mechanism of selective attention. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences*, 2nd Edition, (pp. 1235–1247). Cambridge, MA: MIT Press.
- Leslie, A.M., Friedman, O., & German, T.P. (2004). Core mechanisms in ‘theory of mind’. *Trends in Cognitive Sciences*, **8**, 528–533.
- Malle, B.F. (in press). The moral dimension of people's intentionality judgments. *Journal of Culture and Cognition*.
- McCann, H. (in press). Intentional action and intending: Recent empirical studies. *Philosophical Psychology*.
- Meltzoff, A.N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, **31**, 838–850.

- Nadelhoffer, T. (in press). On praise, side effects, and folk ascriptions of intentionality. *The Journal of Theoretical and Philosophical Psychology*.
- Nichols, S. & Stich, S. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford: Clarendon Press.
- Nuñez, M., & Harris, P.L. (1998). Psychological and deontic concepts: Separate domains or intimate connection? *Mind & Language*, **13**, 153–170.
- Piaget, J. (1932). *The moral judgment of the child*. London: Routledge & Kegan Paul.
- Repacholi, B.M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, **33**, 12–21.
- Roth, D., & Leslie, A.M. (1998). Solving belief problems: Toward a task analysis. *Cognition*, **66**, 1–31.
- Siegal, M., & Peterson, C.C. (1998). Preschoolers' understanding of lies and innocent and negligent mistakes. *Developmental Psychology*, **34**, 332–341.
- Turiel, E. (1994). The development of social-conventional and moral concepts. In B. Puka, (Ed). *Fundamental research in moral development. Moral development: A compendium, Vol. 2*. (pp. 255–292).
- Wellman, H.M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.
- Wellman, H.M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, **30**, 239–277.
- Wellman, H.M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false-belief. *Child Development*, **72**, 655–684.
- Wellman, H.M., & Woolley, J.D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, **35**, 245–275.
- Wimmer, H., Hogrefe, G.J., & Perner, J. (1988). Children's understanding of informational access as source of knowledge, *Child Development*, **59**, 386–396.
- Woodward, A.L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, **69**, 1–34.
- Woodward, A.L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior & Development*, **22**, 145–60.