On Doing Things Intentionally

Pierre Jacob, Cova Florian, Dupoux Emmanuel

On Doing Things Intentionally

FLORIAN COVA, EMMANUEL DUPOUX AND PIERRE JACOB

Abstract: Recent empirical and conceptual research has shown that moral considerations have an influence on the way we use the adverb 'intentionally'. Here we propose our own account of these phenomena, according to which they arise from the fact that the adverb 'intentionally' has three different meanings that are differently selected by contextual factors, including normative expectations. We argue that our hypotheses can account for most available data and present some new results that support this. We end by discussing the implications of our account for folk psychology.

Introduction

What makes an action intentional and which aspects of an agent's action can be achieved intentionally? Philosophical accounts have traditionally emphasized three factors: *foreknowledge, choice* and *control*. For an outcome to be caused intentionally, the agent must first *believe* that this outcome has a chance to occur as a result of his action. The agent must also *choose* to bring about this outcome. Finally, the agent must *control* its occurrence. On this raw canvas, different accounts of intentional actions can be and have been built. For example, some spell out the choice condition by requiring the agent to have a *desire* about the outcome, while others claim instead that the agent must have the *intention* to bring about this outcome. Malle and Knobe (1997) have empirically investigated which factors lay people consider relevant for intentional action and their results closely match these models, as people insisted on the five following components: awareness and belief, desire and intention, and skill. Furthermore, several distinct parameters of an agent's action can be achieved intentionally. Of course, if the agent's action is successful, then the agent's goal (to change some existing state of affairs into a new one) will be achieved intentionally. But as it turns out, the agent's means for achieving his goal too can be chosen intentionally or not. Even a side-effect (or a by-product) of the agent's action can be achieved intentionally or not.

Recently, some new empirical findings have suggested that considerations other than the agent's awareness, belief, desire, intention and skill could play a role in our ascriptions of intentionality to an agent's action: namely moral considerations

This research was supported in part by a Grant from the French Agence nationale de la recherche (ANR) (ANR Blanche: SoCoDev). In addition to the comments by two anonymous referees for this journal, we are grateful to Elisabeth Pacherie for her comments.

Address for correspondence: Florian Cova, Institut Jean Nicod, Ecole Normale Supérieure, 29 rue d'Ulm, 75005, Paris, France.
Email: florian.cova@gmail.com

Mind & Language, Vol. 27, No. 4 September 2012, pp. 378–409.

(Knobe, 2003a, 2003b). Some have claimed that these experiments show that the moral valence of an outcome (i.e. of the agent's bringing about that outcome) is taken into account when we determine whether it is intentional. Others have resisted this claim and proposed alternative accounts of these data.

In what follows, we defend a novel account of these empirical findings, inspired by, and adapted from, Nichols and Ulatowski's 'interpretive diversity' hypothesis (2007). This account relies on two assumptions: (i) the adverb 'intentionally' has three different, though related, meanings, and (ii) the context (partially) determines which of the three meanings will be used. Along the way, we argue that our own view can account for all available data, including data that other accounts so far have not been able to accommodate, and that it makes new predictions that were confirmed by experiments that we have run.

In §1, we describe the original experiments that uncovered the phenomenon. In §2, we survey a category of accounts of this phenomenon according to which 'intentionally' can have different meanings and argue that none of them is fully satisfying. In §3, we present our own account and argue that it succeeds where others fail. In §§4, 5, 6 and 7, we review the available experimental evidence and argue that our account can accommodate all of it. Also, we draw new empirical predictions from our account and put them to test. In §8, we end by discussing the implications of our theory for folk psychology.

1. Two Puzzles for Intentional Action: The Knobe Effect and the Skill Effect

Evidence for the putative influence of moral considerations on ascriptions of intentionality arises from the study of two phenomena, both of which were discovered by the philosopher Joshua Knobe: the 'Knobe Effect' and the 'Skill Effect'. In what follows, we describe Knobe's original experiments that lead to the discovery of these two puzzles.

1.1 The Knobe Effect

The Knobe Effect can be described as the observation that whether a side-effect is considered intentional depends on the moral valence of this side-effect. Consider the following scenario (Knobe, 2003a):

> *Harm Case*: 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*.

In his original study, Knobe found that 82% of the people surveyed answered 'yes' to the question, 'did the chairman of the board intentionally harm the

environment?' When given the same vignette, but this time with the word '*harm*' changed into '*help*' (the *Help Case*), only 23% responded positively when asked if the chairman of the board intentionally helped the environment. This striking asymmetry has since been replicated in other languages and cultures (Knobe and Burra, 2006), in young children (Leslie *et al*., 2006; Pellizzoni *et al*., 2009), in people suffering from Asperger Syndrome (Zalla and Machery, ms) and in patients with cerebral lesions to the prefrontal cortex (Young *et al*., 2006). More recently, it has been shown that the means whereby an agent achieves her goal exhibit asymmetries similar to the Knobe Effect (Cova and Naar, forthcoming a).

1.2 The Skill Effect

The other phenomenon, the Skill Effect, can be described as the fact that moral considerations modulate the impact of the 'skill' factor on ascriptions of intentionality. Consider the following scenario:

> *Bull's-eye (Skill)*: Jake desperately wants to win the rifle contest. He knows that he will only win the contest if he hits the bull's-eye. He raises the rifle, gets the bull's-eye in the sights, and presses the trigger.
>
> Jake is an expert marksman. His hands are steady. The gun is aimed perfectly . . . The bullet lands directly on the bull's-eye. Jake wins the contest.

In this case, 79% of participants answered that Jake intentionally hit the bull's-eye. Now, consider the *Bull's-eye (No-Skill)* case in which the second paragraph is modified:

> But Jake isn't very good at using his rifle. His hand slips on the barrel of the gun, and the shot goes wild . . . Nonetheless, the bullet lands directly on the bull's-eye. Jake wins the contest.

In this case, only 28% of participants answered that Jake intentionally hit the bull's-eye. These results show that ascriptions of intentionality also depend on the degree of control the agent exerts on his action. But now consider the following pair of scenarios:

> *Aunt (Skill)*: Jake desperately wants to have more money. He knows that he will inherit a lot of money when his aunt dies. One day, he sees his aunt walking by the window. He raises his rifle, gets her in the sights, and presses the trigger.
>
> Jake is an expert marksman. His hands are steady. The gun is aimed perfectly . . . The bullet hits her directly in the heart. She dies instantly.

> *Aunt (No-Skill)*: [. . .] But Jake isn't very good at using his rifle. His hand slips on the barrel of the gun, and the shot goes wild . . . Nonetheless, the bullet hits her directly in the heart. She dies instantly.

In the *Skill* condition, 95% of participants answered that Jake intentionally killed his aunt. 76% did so in the *No-Skill* condition. Once again, the outcome is perceived as less intentional when the agent exerts less control. But this difference is much smaller in this case (19%) than in the *Bull's-eye* pair (51%). Moreover, in the *Bull's-eye (No-Skill)* case, most participants consider hitting the bull's eye as not intentional, whereas most participants consider killing the aunt intentional in the *Aunt (No-Skill)* case. These results suggest that the contribution of control to ascriptions of intentionality is greatly diminished when the outcome is morally bad. Another experiment by Knobe suggests that it is also diminished when the outcome is a morally good one (Knobe, 2003b).

These two sets of experiments suggest that moral considerations can (i) play a role in our ascriptions of intentionality and (ii) modulate the extent to which the *control* factor has an impact on these ascriptions. In the following section, we distinguish two main kinds of accounts of the Knobe Effect (which has been at the center of most theoretical attempts so far, at the expense of the Skill Effect) and motivate our preference for one particular kind of accounts.

2. Pluralist Accounts of the Knobe Effect: A Critical Survey

Many accounts have been proposed of these phenomena (particularly of the Knobe Effect). These accounts can be roughly grouped into two categories: monist and pluralist accounts. *Monist* accounts consider that there is *one* folk concept of intentional action and that the effects we described must be accounted for either as a consequence of this concept (e.g. Knobe, 2006; Sripada, 2010; Sripada and Konrath, 2011) or as a consequence of biases, rules or heuristics that interfere with the application of this concept (e.g. Adams and Steadman, 2004a, 2004b, 2007; Nadelhoffer, 2004a, 2004b, 2006; Wright and Bengson, 2009). *Pluralist* accounts claim that the adverb 'intentionally' can have *more than one* meaning or expresses more than one concept, and that asymmetries must be understood as a consequence of this semantic plurality. In this section, we describe the existing pluralist accounts and explain why, although they are not satisfactory, we think they are nonetheless on the right tracks.

2.1 Nichols and Ulatowski's 'Interpretive Diversity' Hypothesis

Nichols and Ulatowksi were the first to propose that 'intentionally' could have different meanings. According to Nichols and Ulatowski's 'interpretive diversity' hypothesis, people actually ascribe two different meanings to the noun phrase 'intentional action': (i) 'having a motive' and (ii) 'having foreknowledge'. Furthermore, one and the same person can adopt one or the other according to the context. In the *Harm Case*, when they use 'intentionally', most people mean 'done with foreknowledge' and they judge the chairman as having harmed the environment 'intentionally' (because he knew his action would harm the environment). But,

in the *Help Case*, when they use 'intentionally', most people mean 'done with a motive' and they consider the chairman as not having helped the environment 'intentionally' (because he lacked a motive for helping the environment).

The first problem for Nichols and Ulatowski's account is that they do not propose hypotheses about the factors that select one of the two meanings rather than the other. It is not sufficient to be told that a word has two meanings, because if true, this does not tell us which of the two meanings a speaker has in mind in a particular use of the word. Furthemore, as an account of the Knobe effect, their approach is, at best, incomplete since it doesn't allow precise predictions for other cases. Nevertheless, one could argue that this doesn't prove that their approach is unable to explain the Knobe effect, just that it needs further precision. But we think that there are cases that it can't accommodate, whatever these specifications are.

Let's begin with the 'foreknowledge' interpretation: does it mean that the agent must be certain that a given outcome will occur? Or does it only mean that he must be conscious that this outcome has a chance (even a slight one) to occur as a result of his action? We think that Nichols and Ulatowski have to adopt the second solution in order to accommodate the following data: we gave 20 subjects a version of the *Harm Case* in which the chairman is told that the program has only a 10% chance of harming the environment. Against the odds, the program ends up harming the environment. In this case, most subjects judged that the chairman intentionally harmed the environment. So, consciousness of a small chance for an event to occur should be enough for the 'foreknowledge' meaning. But, if this is the case, then Nichols and Ulatowski's account cannot explain why people judge that the agent did not intentionally hit the bull's-eye in the *Bull's-eye (No-Skill)* case. According to the 'motive' interpretation, his action should be judged intentional, since he had the desire to hit the bull's-eye. But it should also be judged intentional according to the 'foreknowledge' interpretation, since the agent is conscious that there is a chance (even small) that he hits the bull's-eye. So, Nichols and Ulatowski cannot explain why people judge that hitting the bull's-eye is not intentional in this case.

It could be answered that Nichols and Ulatowski's account was only aimed at explaining subjects' answers in the case of side-effects, and that it is unfair to dismiss it on the basis of results drawn from researches on the Skill Effect. Fair enough! But Nichols and Ulatowski also have troubles with judgments of intentionality in the case of side-effects. Let's consider the following case (inspired by Nanay, 2010):

> *Apple Tree*: A company has decided to expand its building. The vice-president of the company has been assigned the task to prepare the building's new plans. Once the plan is finished, the vice-president goes to submit them to the chairman of the board. On his way, he thinks, 'The chairman will surely be happy. Expanding our building will increase our profits. Moreover, to start the expansion, it will be necessary to cut down the apple tree in front of the chairman's window. The chairman has always hated this tree that has annoyed him ever since he moved into this office. That's what he told me many times.'

> The vice-president submits the plans to the chairman, 'Expanding the building will help us increase profits. Moreover, to expand our building, we will need to cut down the apple tree that is in front of your office.' The chairman answers, 'So what? Although that apple tree has annoyed me ever since I moved into this office, I don't care at all about its being cut down. All I care about is making profits. Start the expansion.'
>
> They started the expansion and the apple tree was removed.

We gave this case to 34 subjects and asked on seven-point scales (i) whether the chairman intentionally had the apple tree cut down (on a 7-point scale ranging from −3 = 'NO' to 3 = 'YES') and (ii) how much the chairman wanted to have the apple tree cut down (on a 7-point scale ranging from −3 = 'he didn't want to' to 3 = 'he really wanted to', 0 being 'he didn't care'). The results showed that participants tended to consider that the chairman did not intentionally have the apple tree cut down (M=−0.62, with only 29% of subjects giving a answer superior to 0) but tended to consider that he wanted to have the apple tree cut down (M=0.90, with 57% of subjects giving a answer superior to 0). So, we have a case in which most subjects judge that (i) the agent desired the outcome, (ii) the agent had foreknowledge of the outcome (since this is explicitly told in the scenario) and yet (iii) the agent did not intentionally bring about this outcome: these results suggest that the two meanings described by Nichols and Ulatowski do not exhaust our use of the word 'intentionally'.

2.2 Cushman and Mele

Cushman and Mele (2008) have defended a somewhat similar (but more precise) hypothesis, according to which, when ascribing intentionality, (i) all people consider 'having a desire' as a sufficient condition, (ii) only about 20% consider 'foreknowledge' as a sufficient condition in general, and (iii) most people consider that, nevertheless, 'foreknowledge' can be a sufficient condition in the case of bad actions. Nevertheless, this hypothesis cannot account for certain cases,[1] such as the following (drawn from Mele and Cushman, 2007):

[1] Take for example the *Apple Tree* case: in this case, the agent knows that the apple tree will be cut down (there is 'foreknowledge') and people tend to say that the agent desires the tree to be cut down (there is 'desire'). So, we should expect people to tend to say that the chairman intentionally had the apple tree cut down—which is not the case. One could reply that the agent's desire in this case is not strong enough. But, even so, there is still a problem: there is a second version of the *Apple Tree* case, in which the agent is portrayed as having a long-time relationship with the tree. In this case, people tend to say he didn't desire to have the tree cut down. Still, they tend to say he intentionally had the tree cut down (Cova and Naar, forthcoming b). So, we have here a case in which we have an intentional action without desire, and without the side effect being a bad one (since, if having the tree cut down was a bad side effect, 'foreknowledge' would be a sufficient condition, and Cushman and Mele should expect high intentionality ratings in both *Apple Tree* cases).

> *Pond*: Al said to Ann: 'You know, if you fill in that pond in the empty lot next to your house, you're going to make the kids who look for frogs there sad.' Ann replied: 'I know that I'll make those kids sad. I like those kids, and I'll definitely regret making them sad. But the pond is a breeding ground for mosquitoes; and because I own the lot, I am responsible for it. It must be filled in.' Ann filled in the pond, and, sure enough, the kids were sad. Did Ann intentionally make the kids sad?

Participants had to answer the question using a scale from 1 ('no') to 7 ('yes'). The mean intentionality rating was 3.19 (with only 28% saying 'yes'). So, in this case, most people judged unintentional a side-effect that was (i) bad and (ii) foreseen[2], in contradiction with Mele and Cushman's account that would predict that, in this case (involving a bad action, 'making the kids sad'), most people would consider 'foreknowledge' as a sufficient action, and would judge that Ann intentionally made the kids sad.

2.3 Sousa and Holbrook

In a more recent study, Sousa and Holbrook (2010) have proposed an account of the Skill Effect that relies on a distinction between two folk concepts of intentional action: according to the 'simple concept' of intentional action, it is sufficient to do something with the intention to do it intentionally, while, according to the 'composite concept' of intentional action, intention is a necessary but insufficient condition. It must be supplemented with other conditions such as the fact that the outcome was brought about following the agent's plan, and on the basis of a reliable capacity. Sousa and Holbrook's hypothesis is that blameworthy actions (such as murder) make the 'simple concept' more salient. When the agent tries to hit a bull's-eye, both concepts are equally salient, and a certain number of subjects take into account whether the agent succeeded by sheer luck and on the basis of a reliable capacity, leading to a difference between the *Bull's-eye (Skill)* and *Bull's-eye (No-Skill)* cases. But, when the agent tries to murder his aunt, the simple concept is the most salient, leading most subjects to take into account only the agent's intention, and reducing to almost nothing the difference between the *Aunt (Skill)* and the *Aunt (No-Skill)* cases.

---

[2] One might say that, in the *Pond* case, the side effect is not really 'bad'. First, notice that Cushman and Mele (2007) explicitly designed this case as involving a 'bad side-effect'. Second, even if we focus here on the *Pond* case, it is not unique. Phelan and Sarkissian (2008) designed a similar puzzling scenario (*City Planner*) in which a city planner starts a plan that will clean up the toxic waste polluting a former industrial area but will also increase the level of joblessness. Only 29% of participants considered that the city planner intentionally increased the level of joblessness. Lanteri (2009) used another scenario (*Lever*) in which a trolley is diverted in order to save five people, causing the death of a sixth person as a side effect. In this case, only 29% of participants answered that killing the person was intentional. These cases ('increasing joblessness' and 'killing a person') are clear cases of morally bad actions.

Though Sousa and Holbrook succeed in explaining the Skill Effect, they acknowledge that their account can't be applied to the Knobe Effect: the chairman's harming the environment would be unintentional according to both the simple and the composite concepts. As one can see, none of these hypotheses can at the same time explain the Knobe Effect and the Skill Effect. So, there's no existing pluralist account that can claim to be a complete theory of our use of 'intentionally' and our ascriptions of intentionality. Nevertheless, we think that these theories are on the right track when they claim that 'intentionally' is a polysemous word and that this polysemy is the source of the Knobe and the Skill Effects. The first reason is that informal observations and reports show that, when tested, people often ask what we, experimentalists, mean by 'intentionally' and start to argue with one another about the meaning of this word once the experiment is over.[3] This strongly suggests that 'intentionally' can be understood in different ways. Nichols and Ulatowski's and Sousa and Holbrook's studies have confirmed these informal observations through controlled experiments by showing that the kind of justifications people give to their intentionality judgments can greatly differ according to the case they have been considering. This fluctuation in the criteria people use (or claim to use) to attribute intentionality probably reflects changes in the concept of intentional action that is used.

A second reason is that, in accordance with people's changing justifications, criteria that matter for intentional action in one experiment can fail to have any impact on ascriptions of intentionality in another set up. The Skill effect is a good example (with the 'control' condition being a condition for intentional action in one case but not in another), but it is not the only one. Let's consider for example the subjective probability of success—that is the probability assigned by the agent to his successfully achieving a certain outcome. In some cases, this probability seems to matter. Consider for example Nadelhoffer's **C1**, **C2**, **C3** and **C4** cases (2006c). In these cases, a hunter kills a deer to win a hunting competition but, as a side-effect, he also hits and kills a bird-watcher. Among the cases, Nadelhoffer varied the probability that the hunter ascribes to his killing the bird-watcher. The smaller the probability, the less intentionality participants ascribed to the hunter's killing of the bird-watcher. So, we would be inclined to think that subjective probabilities matter for intentional action. But let's now consider the following case (drawn from Mele and Cushman, 2007):

> *Bowl*: Earl is an excellent and powerful bowler. His friends tell him that the bowling pins on lane 12 are special 200-pound metal pins disguised to look like

[3] In pilot studies for experiments reported in this article, we tested groups of friends in universities. Each one of them answered separately, but, once the survey was filled, they would often ask each other what their answer was. More than once, this started a dispute about what 'intentionally' really means: some would argue that 'intention' was necessary for something to be done intentionally, while others would claim that it wasn't.

> normal pins for the purposes of a certain practical joke. They also tell him that it is very unlikely that a bowled ball can knock over such pins. Apparently as an afterthought, they challenge Earl to knock over the pins on lane 12 with a bowled ball and offer him ten dollars for doing so. Earl believes that his chance of knocking over the pins on lane 12 is very slim, but he wants to knock them down very much. He rolls an old bowling ball as hard as he can at the pins, hoping that he will knock down at least one. To his great surprise, he knocks them all down! The joke, it turns out, was on Earl: the pins on lane 12 were normal wooden ones. Did Earl intentionally knock down the pins?

On a scale from 1 (not intentional) to 7 (intentional), participants gave a mean answer of 6.36 (that is, very intentional). But, in this case, Earl believes that it is very unlikely that he will succeed in knocking down the pins (he ascribes very low subjective probabilities). Still, most participants rate his action as intentional.

So, from one case to another, not only do people explicitly give different criteria for intentional action, but also they are sensitive to different factors. We take this as a strong motivation for a pluralist account: the fact that ascriptions of intentionality depend on so many factors is evidence that 'intentionally' has several meanings. All we now need is a theory that can (i) specify the different meanings of 'intentionally', (ii) tell how different contexts preferentially elicit each of these meanings and (iii) accommodate all the existing data. In the following section, we advance a tentative hypothesis that could fulfill all three requirements.

3. Revisiting the 'Interpretive Diversity' hypothesis

In this section, we introduce our account of the Knobe and Skill Effects and argue that it accommodates all the cases we have encountered so far. Our new account can be considered as belonging to the class of pluralist accounts, since it claims that 'intentionally' has various meanings that are differently elicited depending on which features of the case are emphasized.

3.1 General Presentation

If 'intentionally' has various meanings, which are they? One meaning seems straightforward: to do *x* intentionally, you must 'aim at *x*' or 'intend to do *x*', that is, do *x* because you *want x* to occur. The relevant notion of intention here is arguably what philosophers have called *prior* intention (as opposed to intention *in action*).[4] An agent's prior intention is closely linked to her desire. You must be motivated to engage in an action whose outcome will be x. For example, in order to understand an agent's action in a given context, one must identify the agent's prior intention: the very same hand movement can be recruited by an agent whose

[4] Cf. Searle, 1983; Bratman, 1987; Pacherie, 2000.

prior intention (or desire) is to frighten a bird, to greet an invitee or to wave good-bye to a departing host guest. This is roughly the definition given by many dictionaries.[5] Let's call this first meaning the *conative* meaning of 'intentionally'. This *conative* meaning is close to the 'desire' meaning hypothesized by Nichols and Ulatowski's and Sousa and Holbrook's 'simple concept' of intentional action.

Now, it is necessary that an agent's act arises from some desire or other, but it is not sufficient for the agent's action to count as intentional. Consider a puppet on a stage, whose intentionality entirely derives from the intentionality of another agent and whose movements are fully guided by another's intentions and desires. Clearly this puppet cannot be said to be acting intentionally because it lacks the relevant intentions and desires. Now consider an agent with her own non-derived intentions and desires, but who is acting under the coercion of another agent whose desires and intentions are being manipulatively induced by an evil neuroscientist, or whose life is being threatened, or who is being blackmailed. In this case, the agent is acting out of her own intentions and desires: she does not have some weird desire implanted in her brain by an evil neuroscientist; or she genuinely wants to save her life or the life of her beloved. But she is not really acting intentionally because the intentions and desires on the basis of which she acts were not formed *autonomously*: they were forced onto her, i.e. they were formed under some constraint or coercion. If so, then acting intentionally requires the agent's action to flow from intentions and desires that she agent autonomously (or freely) formed. Let's call this the *autonomy* meaning of 'intentionally'.

Finally, in addition to having autonomous desires and prior intentions, an agent who engages in a particular action must select a particular motor sequence (or behavior) that she thinks is her best means of fulfilling her desire or prior intention, given the environmental constraints that she faces. This is roughly what philosophers call an *intention in action* or a *motor intention*. For example, if an agent's prior intention (or desire) is to turn on the light, then she must either form the motor intention to press a switch or else form the communicative intention to ask another to turn on the light. Furthermore, the environmental constraints faced by the agent may change in the course of the agent's action. If the environmental constraints vary, the agent of an intentional action must be able to perceive the changes and to adapt her behavior to the changes in order to fulfill her prior intention or desire to achieve a particular outcome. What matters to the intentionality of the agent's action is the ability of the agent to vary its behavior in accordance with the changes of the environmental constraints.[6] Unless the agent exhibits this behavioral variability in response to environmental changes, her behavior will be judged purely accidental,

[5] For example, the *Cambridge Dictionary Online* defines an 'intentional action' as an action that was 'intended or planned'.

[6] Developmental psychologists have emphasized the importance of what Biro *et al.* (2007), Biro and Leslie (2007) and Csibra (2008) call *equifinal variations of behavior* as a cue that shapes the interpretation by preverbal human infants of an agent's movements as goal-directed.

i.e. she will have achieved a given outcome (if she does) by sheer luck. Let us call this sense of 'intentionally' the *control* sense.[7]

In accordance with these semantic and pragmatic observations, we hypothesize that 'intentionally' can have three different meanings, which can be selectively foregrounded by the context:

i. First (*conative*) meaning: asking whether *A* did *x* intentionally amounts to asking whether *A* had the desire to bring *x* about;
ii. Second (*autonomy*) meaning: 'intentionally' is to be understood by contrast with 'unwillingly', so that asking whether *A* did *x* intentionally amounts to wondering whether *A* was forced or coerced into doing *x*;
iii. Third (*control*) meaning: 'intentionally' is to be understood by contrast with 'accidentally' or 'by sheer luck', so that asking whether *A* did *x* intentionally amounts to wondering whether *A* had control over the occurrence of *x* or whether *A* succeeded by sheer luck.

On the basis of a given scenario, subjects form expectations about an agent's psychological attitudes and character. Thus, when they interpret the intentionality-question that they must answer, subjects pick out one of these three meanings on the basis of their expectations about the agent's psychology. In the next sections, we give a more accurate description of this hypothesis and apply it to the different cases we surveyed.

3.2 Descriptions of the Three Meanings

All three meanings share some requirements that an event must fulfill if it is to be considered intentional. These requirements are the following:

i. The event must originate in an action that was directed by an intention in action (that is, a goal-directed action, not just a reflex);[8]

[7] Each of the three meanings of 'intentionally' has specific antonyms. First, if one causes an effect *carelessly* or *recklessly*, one causes it in violation of the conative sense of 'intentionally' (because this effect was not what the agent wanted to achieve). Secondly, if one causes an effect *unwillingly* (i.e. against one's own will), one causes it in violation of the autonomy sense of 'intentionally'. (Arguably, to cause an effect unwillingly is not just to cause it without wanting to, i.e. carelessly or recklessly, but also to cause it while wanting *not* to cause it.) Finally, if one causes an effect *accidentally* (or by luck), one causes it in violation of the control sense of 'intentionally'. In accordance with the *conative* meaning, various Google searches returned 37,900 results for 'intentionally or carelessly' and 1,150,000 results for 'intentionally or recklessly'. In accordance with the *autonomy* meaning, we found 2,300 results for 'intentionally or by necessity', 3,340 for 'intentionally or reluctantly', 12,000 results for 'intentionally or unwillingly', 69,800 for 'intentionally of forced' and 213,000 for 'intentionally of by force'. Finally, in accordance with the *control* meaning, a Google search of 'intentionally or accidentally' returned 908,000 results (plus 146,000 results for 'intentionally or by accident').

[8] In Mele and Cushman's *Weed* case (2007), a person named Jen drives to the hardware store to buy weed spray and get rid of weeds when she loses control of her car and ends up crushing the weeds. In this case, most people considered her eliminating the weeds as not intentional.

ii. The event must originate in an action the agent consciously planned (not an action performed out of pure habit);[9]
iii. The agent *either* (i) specifically tried to bring about the event *or* (ii) knew that the event had a non-zero chance to happen.[10]

In light of these necessary conditions, we propose the three following different meanings for 'intentionally':

i. *Meaning 1*: an event is intentional if the agent's *desire* for the occurrence of the event turns out to be at least as strong as (or stronger than) we expected it to be.[11]
ii. *Meaning 2*: an event is intentional if the agent is *less reluctant to bring about* the event that we would expect him to be.
iii. *Meaning 3*: an event is intentional if the agent brought about this event, not by sheer luck, but by exerting *control* upon its occurrence.

Notice that our hypotheses are in agreement with traditional accounts of intentional action: whereas *foreknowledge* is a necessary condition for all three meanings, *choice* (under the form of 'desiring to' and 'not being reluctant to') is at the core or Meanings 1 and 2, and *control* is at the heart of Meaning 3. So, our hypotheses highlight the very same parameters as traditional accounts.

3.3 Eliciting Features

Now, we must specify which features of the scenario will lead people to favor one meaning over the others. We think there are two main eliciting factors: (i) which desire the agent can be *expected* to have and (ii) the possibility of *failure* (of the agent's action).

Let's start with expectations. There are two different kinds of expectations: *normative* and *descriptive*. When we normatively expect *A* to *x*, we mean that *A should morally x*.[12] When we *descriptively* expect A to *x*, we attribute a *high probability* to A's *x*-ing. In the following, when we use the word 'expectation', we mean both types of expectations.

---

[9] In Mele and Cushman's *Drive* case (2007), a man drives home by habit, though he had originally planned to go to his favorite store. In this case, only 28% of participants answered that he intentionally drove home.

[10] The agent's foreknowledge can be very elusive. For experiments on this precise topic, see Pellizzoni *et al*., 2010.

[11] Here and in the rest of the article, by 'desire' we mean a very broad and general kind of *pro-attitude*. Note that we don't propose a 'pure desire' account of intentional action: we don't claim that 'desire' is a sufficient condition for intentional action, since it must be accompanied by foreknowledge and furthermore the outcome must be the result of an intention in action. Also note that the relevant desire is the one that actually explains the agent's behavior.

[12] Normative expectations can also be described as what we believe a morally reasonable agent would desire in a given context.

The second relevant feature is the possibility of failure. Some actions appear to us as more likely to fail than others: for example, ringing the doorbell typically seems to us less likely to fail than hitting a distant target with an arrow. There is less room for failure in the first than in the second case and we consider that, in the second case, success depends on a particular skill.

These two kinds of features guide the choice of the relevant meaning of 'intentionally' when subjects are faced with vignettes (the context) and asked whether the agent acted intentionally. Why? Remember that subjects have (or think they have) to determine what the experimenter wants to know (and thus means by 'intentionally'). Thus, it is likely that their 'deduction' of the appropriate meaning of 'intentionally' will be oriented by the content of the vignette and by what they think the more interesting question is in the current context. Thus we make the following predictions:

- If we are led by the context to expect agent *A* to have the desire to bring about event *x*, then Meaning 1 will be preferentially elicited. Indeed, if we normatively expect *A* to have the desire that *x*, then the more interesting question will be whether he fulfilled this expectation. And if *A* is descriptively expected to desire *x*, then we will wonder whether our expectation was correct.
- If we are led to expect agent *A* to be reluctant to bring about event *x*, then Meaning 2 will be preferentially elicited. Indeed, if we normatively expect *A* to be reluctant to *x*, then the more interesting question will be whether he fulfilled this expectation. And if *A* is descriptively expected to be reluctant to *x*, then we will wonder whether our expectation was correct.
- If there was a serious possibility that *A* failed to cause *x*, then Meaning 3 will be elicited, because we will wonder whether *A*'s success was due to sheer luck or to *A*'s control over his action.

Of course, this is far too simple: there are cases in which these rules lead to the prediction that all three meanings will be elicited. Imagine the case of a psychopath with sniper skills who intentionally shoots an innocent person in the head at a great distance. Suppose we ask: 'did the psychopath intentionally kill the person?' In this case, all three meanings will be elicited because (i) we descriptively expect our psychopath to have the desire to kill a person, (ii) we normatively expect him to be reluctant to kill a person and (iii) hitting a target at a great distance requires either skill or luck and there is a serious possibility of failure.[13]

Arguably, there is no algorithm that would allow us to predict for each case which meaning will be the most salient. Saliency is a contextual matter: which meaning will be chosen depends on which feature is made salient by the context.

[13] In this particular case, our account predicts that our psychopath killed the person intentionally according to the three meanings, and thus that people would overwhelmingly answer 'yes' to the question: 'did the psychopath intentionally killed the person?'

Furthermore, our expectations about, e.g., the strength of an agent's desire are shaped by of all the relevant contextual information about the agent's psychology. For example, imagine a scenario in which an agent, who has been described as ambivalent or even indifferent about money but who nonetheless decides to pick up a five Euros bill which he has just seen on the ground. We are likely to judge that he intentionally picked up the money. This might seem like a problem for our account of meaning 1. But it is not. Our judgment is the result of our ability to make interpretive adaptations in response to incoming information. After being told that the agent is indifferent about money, we expected him to care less about money than *most people*. Now, we learn that he picked up a five Euros bill on the ground. To make sense of the agent's novel action, in the light of our background expectations about him, we ascribe to him a desire for this bill, on this occasion, which is *stronger* than the desire we would have ascribed to him on the basis of the background description of his overall attitude towards money alone. This is in accordance with our account of Meaning 1.[14]

However, there are some general principles of interpretation:

- *Standard pragmatic assumption*: The more a scenario highlights a feature, the more this feature is salient.
- *Priority of normative expectations: ceteris paribus*, normative expectations are more salient than other features.

The first principle is quite evident: the more you draw the subject's attention to a particular feature, the more this feature becomes salient. The second principle is less evident but can be justified in different ways. A priori, it is likely that moral features are salient to moralizing creatures like us: we are fast and skilled in detecting violations of normative expectations. A posteriori, the Skill effect shows that, when moral expectations are present, other features (like control) tend to be overlooked. Furthermore, the second principle can also be further illustrated by a case that might seem to threaten our account of meaning 2 of 'intentionally'. Suppose that we learn (from a historian) that Stalin actually had regrets about sending so many people to the Gulag. Surely, we would be surprised: we wouldn't have expected Stalin to feel regret. Thus, it would show that Stalin was more reluctant to send people to the Gulag than we expected him to be on the basis of what we knew about him. However, we wouldn't say that he didn't do it intentionally. It may seem as if, in light of Stalin's regret, our judgment of intentionality is in violation of our account of meaning 2. But it is not. We are surprised by Stalin's regret in light of our descriptive expectations about Stalin's psychology. But if we compare Stalin's attitudes to what we would *normatively* expect from other people, we would conclude that his reluctance to engage in this action was not strong enough to prevent him from performing it, since he actually did it. So, from a *normative* point

[14] We are grateful to an anonymous referee for this journal for raising such a case.

of view, Stalin was *less reluctant* than we expected him to be—and this warrants our ascription of intentionality. In accordance with the second principle, this example also highlights the fact that, in cases of morally bad actions, our judgments are shaped by *normative*, rather than descriptive, expectations.[15]

Of course, these meanings and principles are only tentative hypotheses that cannot be fully justified independently from the data. Nevertheless, we will show that they are successful in accounting for the data that have been collected so far, and this success is the main reason why we should accept them.

3.4 Accounting for the Data

So, can our hypotheses account for all the different cases described so far? Let us begin with the original Knobe Effect. In the *Harm Case*, the chairman is (normatively) expected to be reluctant to harm the environment: thus, Meaning 2 is preferentially elicited. But, because the chairman does not care about the environment, he is much less reluctant to harm the environment that we would normatively expect him to be. So, harming the environment is intentional. In the *Help Case*, the chairman is (normatively) expected to have the desire to help the environment: Meaning 1 is elicited. But, the chairman does not have the desire to help the environment as much as we would expect him to. So, helping the environment is not intentional.

In the *Apple Tree* case, we are told that the apple tree annoys the chairman, so we (descriptively) expect him to have the desire to get rid of the tree (which elicits Meaning 1). But he claims not to care and so does not seem to really have the desire to get rid of the tree: thus, getting rid of the tree is not intentional, because the chairman's desire to get rid of the tree is not as great as we expected.

In the *Pond* case, making the kids sad is something we (normatively) expect Ann to be reluctant to do. So, Meaning 2 is elicited. But the case portrays Ann as reluctant to make the kids sad: she does so only because she is forced to. Because she has a good reason to act, most people judge that, although she finally decides to act, she does so reluctantly enough. So, making the kids sad is not intentional.

Finally, our hypotheses can account for the Skill Effect. In the two *Bull's-eye* cases, there is a descriptive expectation (Jack wants to win the contest) and a serious possibility of failure (hitting the bull's-eye requires skill). But much of the scenario focuses on Jack's aiming at the bull's-eye and trying to hit it, while Jack's desire to win the contest is quickly mentioned in the first sentence of the scenario. So, Meaning 1 is not really salient while Meaning 3 is strongly elicited. This is why the main factor that drives attributions of intentionality is the amount of control Jake exerts on his action. In the *Skill* condition, Jack exerts much control on his action:

[15] We are grateful to an anonymous referee for helping us to better link our account of Meaning 2 and our second general tentative principle of interpretation.

thus, it is intentional. In the *No-Skill* condition, on the contrary, Jake succeeds only by sheer luck: thus, it is not intentional.

In the two *Aunt* cases, the outcome (killing his aunt) raises very strong normative expectations (we normatively expect Jake to be reluctant to kill his aunt) and it is mentioned in several places (it is the object of Jack's desire, it constitutes the end of the scenario and it is part of the content of the question). So, Meaning 2 is preferentially elicited. What is most important is not the amount of control exerted by Jack; it is Jack's desires. As in both cases Jack is not reluctant to kill his aunt, killing his aunt is intentional in both cases. Nevertheless, a small difference remains between the two cases because a small number of subjects are using Meaning 3 and are sensitive to differences in control.

So, our hypotheses can account for the data just surveyed. In the following sections, we argue that the explanatory power of our hypotheses is strong enough to account for most of the experimental literature on the use of 'intentionally'. For each of our three meanings, we will describe the empirical predictions that can be derived and search for empirical data that confirm or disconfirm these hypotheses. Where no data were available, we ran our own experiments.

4. Empirical Predictions Related to Meaning 1

In this section, we describe the empirical predictions that can be derived from the hypothesis that people use 'intentionally' with the first meaning ('having the desire to') in cases in which the intentionality question bears on an outcome about which we would expect the agent to have a desire.

4.1 Immunity to Changes in Objective and Subjective Control

Let us first call 'objective control' the control that the agent exerts on his action and 'subjective control' the control that the agent *believes* himself to be able to exert on his action. If the agent brings about outcome x with low objective control and/or low subjective control, does this make this outcome less intentional on Meaning 1? Clearly, it should not: having (or believing to have) low control upon his action doesn't reduce the agent's desire to x. So, we should expect decreases in objective and subjective control not to have any effect on people's use of 'intentionally' when they use Meaning 1. Here is a scenario drawn from the literature (Nadelhoffer, 2005):

> *Nuclear (Good)*: A nuclear reactor is in danger of exploding. Fred knows that its exploding can only be prevented by shutting it down, and that it can be shut down only by punching a certain ten-digit code into a certain computer. Fred is alone in the control room. Although he knows which computer to use, he has no idea what the code is. Fred needs to think fast. He decides that it would be better to type in ten digits than to do nothing. Vividly

> aware that the odds against typing in the correct code are astronomical, Fred decides to give it a try. He punches in the first ten digits that come into his head, in that order, believing of his doing so that it 'might thereby' shut down the reactor and prevent the explosion. Amazingly, he punches in the correct code, thereby preventing a nuclear explosion and saving thousands of people.

In this case, 73% of participants said that Fred intentionally prevented the explosion (a morally good event that we would expect the agent to have a strong desire to cause). Our hypothesis can account for this result if we suppose that most (but obviously not all) participants have adopted Meaning 1.

4.2 Making Morally Good Side-effects Intentional

Our hypothesis also makes the prediction that a morally good side-effect can be intentional if the agent genuinely has the desire that it occurs. Wible (2009) used a case similar to the *Help Case*, but in which the chairman answered: 'Great! I care about helping the environment. I am happy that we can help the environment and make a profit at the same time. Let's start the new program.' In this case, 55% of participants thought the chairman intentionally helped the environment, which is higher than in the original *Help Case*. Nevertheless, 55% is not that high. One possibility is that, in this case, most participants do not really believe that the chairman genuinely cares about the environment but that he thinks that helping the environment will improve his reputation. (For similar results, see Guglielmo and Malle, 2010.)

To address this concern, we designed the following scenario on the model of the *Help Case*:

> *Very Nice Chairman*: The vice-president of a company goes to the chairman of the board and says: 'You asked us to imagine new programs that would enable us to make more money. We can propose two programs. Program A will enable us to make a lot of money for a very small cost. Program B will generate as much money as Program A for the same cost, but will have the supplementary effect of helping the environment. Nevertheless, it will be impossible to prove that it is our action that helped the environment, and that won't help our reputation. What program do you want us to start?' The chairman of the board answers: 'Let's start Program B.' Program B is started and the environment is helped.

Participants were asked (i) if the chairman intentionally helped the environment and (ii) whether helping the environment was the chairman's goal, a means to achieve his goal or a side-effect of his action. Among the 25 participants, 20 (80%) answered that helping the environment was a side-effect. Among the 20 participants who answered that helping the environment was a side-effect, 16 (80%) answered

that the chairman intentionally helped the environment.[16] These results confirm our prediction: a morally good side-effect becomes intentional when the agent has a genuine desire that it occurs.

5. Empirical Predictions Related to Meaning 2

In this section, we present empirical predictions related to the use of Meaning 2 when ascriptions of intentionality bear on an outcome that we would expect the agent to be reluctant to bring about.

5.1 Immunity to Changes in Objective Control

As Meaning 2 bears only upon the agent's mental states, changes in objective probabilities won't have an effect on ascriptions of intentionality. Here is an example drawn from the literature (Nadelhoffer, 2005) that parallels the examples given for Meaning 1:

> *Nuclear (Bad)*: Fred has just been fired from the nuclear power plant. In a desperate fit of anger, he decides to cause the reactor to meltdown. Fred knows that the only way the reactor can be forced to melt down is by punching a certain ten-digit code into a certain computer. Fred is alone in the control room. Although he knows which computer to use, he has no idea what the code is. Fred needs to think fast before the other employees return. Vividly aware that the odds against typing in the correct code are astronomical, Fred decides to give it a try. He punches in the first ten digits that come into his head, in that order, believing of his doing so that it 'might thereby' cause the reactor to meltdown. Amazingly, he punches in the correct code, thereby causing a serious nuclear meltdown and killing thousands of people.

In this case, 83% of the participant said that Fred intentionally caused the explosion. So, it seems that objective control really doesn't matter if the outcome is one that we would expect an agent to be reluctant to cause.

5.2 Sensitivity to Changes in Subjective Control under Particular Conditions

The *Nuclear (Bad)* case could lead us to think that changes in subjective control do not affect intentionality ratings for an outcome we would expect the agent to

[16] It is very likely that most of these participants considered that the chairman deserved praise for having helped the environment. So, these results seem to contradict Nadelhoffer's claim that if an agent is praiseworthy for having caused x, then x cannot be a side effect. For a discussion of this problem, see Feltz, 2007 and Nadelhoffer, 2007.

be reluctant to cause. But, in this particular case, the outcome is the agent's goal, so it is clear that the agent is not reluctant to bring it about. In fact, our account predicts that decreases in subjective control will cause decreases in ascriptions of intentionality when they cause people to consider the agent to be more reluctant to cause the outcome. It is easier to believe someone did not want something bad to happen when this person thought he had little chance to bring this outcome than when he was sure it would happen.[17]

5.3 Sensitivity to Regret

The more the agent will express regret for having brought about the outcome, the less the outcome will be judged intentional. This phenomenon has indeed been observed by Sverdlik (2004), Sripada (2010) and Guglielmo and Malle (2010).

5.4 Sensitivity to the Value of the Goal: A Review of the Literature

Imagine a pair of cases in which the agent causes the same means or side-effect he claims to be reluctant to cause. But, in the first case, he acts for the sake of a relatively futile goal while, in the second case, he has a good reason for his action. We predict that the intentionality rating will be higher in the first scenario, because, as the agent acts for some unimportant reason, it shows that he is not really reluctant to bring about the means or side-effect—at least, not as much as we would have expected him to be. So, the intentionality of 'bad' outcomes can be reduced if the agent acts on the basis of sufficiently 'good' reasons.

The fact that the better the goal, the less intentional is a bad means or a bad side-effect can be found in multiple cases provided by the literature on the Knobe Effect:

5.4.1 Phelan and Sarkissian, 2009. Drawing on Knobe's original *Thompson Hill* scenario (Knobe, 2003a), Phelan and Sarkissian designed the following pair of cases:

> *Lieutenant (Important Goal)*: A lieutenant was talking with a sergeant. The lieutenant gave the order: 'Send your squad to the top of Thompson Hill'. The sergeant said: 'But if I send my squad to the top of Thompson Hill, we'll be moving the men directly into the enemy's line of fire. Some of them will surely be killed!'
>
> The lieutenant answered: 'Look, I know that they'll be in the line of fire, and I know that some of them will be killed. But I don't care at all about what

[17] That might be what is going on in Nadelhoffer's **C1**, **C2**, **C3** and **C4** cases (described earlier). A possibility is that people ascribe more reluctance to a hunter who thinks there's only a small chance of hitting the bird-watcher. A less interesting possibility is that people use the hunter's estimations as a clue to infer what are the real probabilities.

happens to our soldiers. *It's imperative to the success of this campaign that we take Thompson Hill.*'

The squad was sent to the top of Thompson Hill. As expected, the soldiers were moved into the enemy's line of fire, and some of them were killed.

In the *Unimportant Goal* version, the case was modified so that the lieutenant has a less good reason to send the soldiers to Thompson Hill (he just wants to take it). In the *Important* version, 50% of participants answered that the lieutenant intentionally caused the soldiers' deaths, against 76% in the *Unimportant* version. We can explain these results: since the lieutenant has a better reason to send his soldiers in the *Important* case than in the *Unimportant*, it is reasonable to think that the lieutenant cares more about his soldiers in the *Important* case and is thus more reluctant to cause their deaths.

5.4.2 Nadelhoffer, 2006. In a 2006 study, Nadelhoffer used the following pair of cases:

> *Officer*: Imagine that a thief is driving a car full of recently stolen goods. While he is waiting at a red light, a police officer comes up to the window of the car while brandishing a gun. When he sees the officer, the thief speeds off through the intersection. Amazingly, the officer manages to hold on to the side of the car as it speeds off. The thief swerves in a zigzag fashion in the hope of escaping—knowing full well that doing so places the officer in grave danger. But the thief doesn't care; he just wants to get away. Unfortunately for the officer, the thief's attempt to shake him off is successful. As a result, the officer rolls into oncoming traffic and sustains fatal injuries. He dies minutes later.

Faced with this scenario, 37% of the participants said the thief intentionally brought about the officer's death. Another group of participants had to read the following scenario:

> *Thief*: Imagine that a man is waiting in his car at a red light. Suddenly, a car thief approaches his window while brandishing a gun. When he sees the thief, the driver panics and speeds off through the intersection. Amazingly, the thief manages to hold on to the side of the car as it speeds off. The driver swerves in a zigzag fashion in the hope of escaping—knowing full well that doing so places the thief in grave danger. But the driver doesn't care; he just wants to get away. Unfortunately for the thief, the driver's attempt to shake him off is successful. As a result, the thief rolls into oncoming traffic and sustains fatal injuries. He dies minutes later.

In this case, only 10% of the participants answered that the driver intentionally brought about the car thief's death. This difference is statistically significant. This result can be explained by the mere fact that protecting one's own good is a better goal than stealing the property of others.

5.4.3 Tannenbaum, Ditto and Pizarro. Tannenbaum, Ditto and Pizarro (ms; see also Sargent *et al*., ms) used the two following scenarios:

> *Civilian casualties (American forces)*: Recently, an attack on Iraqi insurgence leaders was conducted by American forces. The attack was strategically directed at a few key rebel leaders that have been responsible for a number of recent attacks on American forces, and it was strongly believed that eliminating these key leaders would cause a significant reduction in the casualties of both American military forces and American civilians working in Iraq. It was known that in carrying out this attack there was a chance of Iraqi civilian casualties, although these results were not intended and American forces sought to minimize the death of civilians. The attack was successful—it eliminated all of the targets and is certain to ensure the safety of American soldiers and civilians. Unfortunately, a number of Iraqi civilians were killed and injured in the attack. American representatives say that while this was not intended, it was necessary to take the risk of civilian deaths in order to ensure the future safety of a larger number of individuals. They also stated that sometimes it is necessary to allow the death of innocent people in order to promote greater good.

> *Civilian casualties (Iraqi forces)*: Recently, an attack on American forces was conducted by Iraqi insurgence leaders. The attack was strategically directed at a few key military outfits that have been responsible for a number of recent attacks against the Iraqi rebels, and it was strongly believed that eliminating these key outfits would cause a significant reduction in the casualties of both Iraqi military forces and Iraqi civilians. It was known that in carrying out this attack there was a chance of American civilian casualties, although these results were not intended and American forces sought to minimize the death of civilians. The attack was successful—it eliminated all of the targets and is certain to ensure the safety of Iraqi soldiers and civilians. Unfortunately, a number of American civilians were killed and injured in the attack. Iraqi rebels say that while this was not intended, it was necessary to take the risk of civilian deaths in order to ensure the future safety of a larger number of individuals. They also stated that sometimes it is necessary to allow the death of innocent people in order to promote greater good.

Both scenarios were given to participants classified according to their political orientation. The results show that the more conservative the participants were, the more they tended to consider the death of American civilians (second scenario) as more intentional than the death of Iraqi civilians (first scenario). On the contrary, this difference tended to disappear when the participants were liberal. From Knobe's or Machery's point of view, this can only be explained by claiming that conservatives give less value to the life of an Iraqi civilian than to the life of an American civilian, while liberals give the same value to both. But this is not a very charitable interpretation of the data. According to us, the key change

between these two scenarios is the goal that is aimed at by the agents. According to conservatives, the goal of American forces is good, while the goal of Iraqi rebels is not. Civilian casualties are thus considered as a sufficient reason for Iraqi rebels to refrain from acting, but not for American forces. This is why killing civilians is judged more intentional in the second scenario. From a liberal point of view, the goal of American forces in Iraq is not so highly valued. So killing innocents is a sufficient reason for both American Forces and Iraqi rebels to refrain from acting. Thus, we can explain the asymmetry between conservatives and liberals in these cases without postulating that conservatives give less worth to the lives of non-American people. Moreover, our hypothesis fits the results obtained by Tannenbaum, Ditto and Pizarro who found that the difference between liberals and conservatives is mediated by the intention participants ascribed to the agents: conservatives were far more likely than liberals to ascribe to Iraqi insurgence leaders (than to American forces) a greater desire and intention to harm civilians.

5.5 Sensitivity to the Goodness of the Goal: Experiment 1

5.5.1 Material and Hypotheses. We designed four scenarios. The first two scenarios (*Greedy Chairman + Bad Side-effect* and *Greedy Chairman + Good Side-effect*) are similar to the original *Harm* and *Help* cases. The other two scenarios (*Caring Chairman + Bad Side-effect* and *Caring Chairman + Good Side-effect*) form a similar pair, except that the chairman's goal is not to make money but to save his employee's jobs. Here is an example:

> *Caring Chairman + Bad Side-effect*: The vice-president of a company went to the chairman of the board and said: 'We have designed a new program. It will help us save our employees' jobs and will harm the environment.' The chairman of the board answered: 'I don't care at all about harming the environment. All I care about is saving my employees' jobs. Let's start the new program.' They started the new program and the environment was harmed.

These four cases allow us to systematically vary two factors: (i) the moral valence of the SIDE-EFFECT (HARM, i.e. harming the environment, or HELP, i.e. helping the environment) and (ii) the value of the chairman's GOAL (GREEDY, i.e. making more money, or CARING, i.e. saving his employees' jobs). Our hypothesis makes the two following predictions: (i) intentionality ratings will be lower when the chairman's goal is a 'good reason' and (ii) this effect will be limited to cases in which the side-effect is 'bad' because Meaning 2 is used only in these cases.

5.5.2 Participants and Ratings. 119 students participated in this experiment. Each participant received only one case. Each participant had to answer the question 'Did the chairman intentionally harm / help the environment?' using a scale ranging from −5 to 5 with −5 = 'NO' and 5 = 'YES'.

	Helping the environment	*Harming the environment*
Greedy Chairman	−3.03 (0.61)	3.87 (0.38)
Chairman	−3.36 (0.48)	2 (0.56)

Table 1 *Means (and Standard Errors) of the participants' answers for Experiment 1.*

5.5.3 Results. Results are summarized in Table 1. A two-factor ANOVA revealed a significant effect of SIDE-EFFECT, $F(1,115)=143.2$, $p<.001$, showing that we were successful in replicating the Knobe effect for our two pairs of scenarios, and a significant effect of GOAL, $F(1,115)=4.9$, $p<.05$. No significant interaction effect was found, $F(1,115)=2.2$, $p=.14$. These results are coherent with our first hypothesis.

We also made the prediction that the difference in the ratings of intentionality between the Caring Chairman and the Greedy Chairman will be greater in the case of bad side-effects. We compared the response of the participants for the two scenarios involving a bad side-effect (*Caring Chairman + Bad Side-Effect* and *Greedy Chairman + Bad Side-Effect*) using a two-tailed t-test. We found that the difference was significant, $N=58$, $t=-2.8$, $df=47.6$, $p<.01$. No significant difference was found for the comparison between the two scenarios involving a good side-effect (*Caring Chairman + Good Side-Effect* and *Greedy Chairman + Good Side-Effect*).

5.6 Sensitivity to the Goodness of the Goal: Experiment 2

5.6.1 Material and Hypotheses. In this experiment, we tried to replicate the results of the previous experiment in a context that would elicit probabilistic, i.e. descriptive (or non-moral) expectations, rather than normative expectations (as in e.g. moral cases). We designed six scenarios that varied along the same two factors: the importance of the GOAL, and the goodness of the SIDE-EFFECT. In each scenario, the main character was a doctor. His goal could be *Important* (to cure a deadly disease) or *Unimportant* (to remove an ugly spot from a patient's face). In each scenario, the doctor proposed an effective treatment that had various side-effects. This side-effect could be *Good* (increasing the patient's memory), *Neutral* (imperceptibly modifying the patient's blood pressure) or *Bad* (decreasing the patient's visual acuity). In each of these side-effects, only the patient himself suffers (or benefits) from his decision. Here is an example:

> *Unimportant Goal + Bad Side-effect*: A patient with a skin disease goes to his doctor. The doctor proposes him a new treatment: 'As you already know, this spot will remain on your nose for the rest of your life. But, if I give you this new treatment, it is very likely that this spot will disappear and that your sight will be badly impaired.' The patient answers: 'I don't care about my sight. All I want is to get rid of this spot. Give me the treatment.' The treatment is given to the patient, whose sight is badly impaired.'

Our two hypotheses were the same as in the previous experiment.

5.6.2 Participants and Rating. 200 students participated in this experiment. Each participant received only one version of the scenario. Each participant had to answer the question 'Did the patient intentionally increase his memory / modify his blood pressure / decrease his sight?' using a scale ranging from −5 to 5 with −5 = NO and 5 = YES. Then came a second question in the form of a sentence to complete: 'In itself, decreasing his sight / modifying his blood pressure / increasing his memory seems to you to be something . . . ', followed by a scale ranging from −5 to 5 with the following indications: −5 = UNDESIRABLE, 0 = INDIFFERENT, 5 = DESIRABLE

5.6.3 Results. The second question (bearing on the desirability of the side-effect) was asked in order to verify whether our side-effects corresponded to our classification (Bad / Neutral / Good). Results are summarized in Table 2.

The order of the ratings was the one we expected (Bad < Neutral < Good). Nevertheless, the Neutral side-effect seemed to have been considered rather as a mildly bad side-effect. Table 3 sums up the results for the intentionality question.

A two-factor ANOVA revealed a significant effect of Side-Effect, $F(2,194)=18.5$, $p<.001$, and a significant effect of Goal, $F(1,194)=5.8$, $p<.05$. No significant interaction effect was found, $F(2,194)=0.2$, $p=.74$. Once again, as predicted by our hypothesis, the side-effect is considered to be more intentional when the goal is unimportant. Our hypothesis also predicted that this difference would be greater when the side-effect is really bad, given that a good side-effect or a neutral side-effect are never a good reason not to act, whatever the importance of the goal, but we failed to observe an interaction effect. Nevertheless, for the three SIDE-EFFECT conditions, we compared the responses in the *Important Goal* and the *Unimportant*

Bad	*Neutral*	*Good*
−3.85	−1.97	3.85

Table 2 *Means of the participants' answers to the desirability question in Experiment 2.*

	Good	*Neutral*	*Bad*
Unimportant Goal	−1.33 (0.69)	1.45 (0.49)	2.25 (0.44)
Important Goal	−1.93 (0.57)	0.20 (0.67)	0.95 (0.49)

Table 3 *Means (and Standard Errors) of the participants' answers to the intentionality question in Experiment 2.*

Goal conditions, using a one-tailed t-test, with the hypothesis that the mean will be greater in the *Unimportant Goal* condition. We obtained a significant difference in the *Bad Side-Effect* condition ($N=80$, $t=2$, $df=77$, $p<.05$), only a marginally significant difference in the *Neutral Side-Effect* condition ($N=60$, $t=1.5$, $df=52$, $p<.10$) and no significant difference in the *Good Side-Effect* condition. Though not enough to warrant our prediction, these results are encouraging for future empirical investigations of the interaction between the importance of goal and the valence of the side-effect.

6. Empirical Predictions Related to Meaning 3

In this section, we present empirical predictions related to the use of Meaning 3, which is elicited when ascriptions of intentionality bear on an outcome whose achievement requires either control or luck and when Meaning 1 or 2 is not already salient.

6.1 Immunity to Subjective Control

We assume that subjective probabilities won't have an effect on ascriptions of intentionality when Meaning 3 is used. This assumption is supported by the *Bowl* case, described earlier. In this case, Earl has a great objective control upon his action (he is a skillful bowler) but a very low subjective control (he believes that it is very unlikely that he will succeed in knocking down the pins). Still, most participants rate his action as intentional.

6.2 Sensitivity to Changes in Objective Control

When Meaning 3 is preferentially elicited, an outcome performed by an agent who exerts low objective control upon his action and succeeds by sheer luck should have very low intentionality ratings. This prediction is consistent with data we already mentioned, such as the *Bull's-eye (No-Skill)* cases.

6.3 Changing Answers by Making Some Elements More Salient

Our theory also makes the prediction that it is possible to change participants' answers to cases similar to the *Bull's-eye* and the *Aunt* cases by making some elements (and therefore a certain meaning) more salient. For example, in Knobe's *Bull's-eye (No-Skill)* case, there are two different elements: the fact that the agent wants to hit the bull's-eye (Meaning 1) and the fact that hitting the bull's-eye requires skill (Meaning 3). Knobe's results suggest that the second element is more salient than the first and that most participants use Meaning 3. But making the first element more salient could modify these results. Sripada (2010) gave participants a modified version of this case in which they were told that Jake wants to be a police

	Skill	*No-Skill*
'kill his aunt'	100%	100%
'shoot his aunt'	90%	84%
'hit his aunt's heart'	95%	49%

Table 4 *Results from Malle, 2006.*

officer more than anything else, that the rifle contest is part of the competition to enter the police academy and that Jack needs to hit the bull's-eye for in order to win the contest. In this version, 90% of participants answered that Jack intentionally hit the bull's-eye.

Another solution is to modify the description of the outcome in the questionnaire. We can imagine that one and the same outcome can be described in different terms, and that some will preferentially elicit a certain meaning. For example, Malle (2006) used variations of Knobe's *Aunt* cases in which only the question varied. Three questions were used: 'Did Jake intentionally kill his aunt?'; 'Did Jake intentionally shoot his aunt?'; and 'Did Jake intentionally hit his aunt's heart?' The results are presented in Table 4.

As we can see, the difference between the *Skill* and the *No-Skill* conditions is absent when the question bears on 'kill his aunt' but is present when it bears on 'hit his aunt's heart'. This can be explained by the fact that the description 'kill' stresses the immoral side of Jake's action (eliciting Meaning 2), while the description 'hit his aunt's heart' stresses the fact that Jake's action required skills (eliciting Meaning 3).

Here is another example drawn from Nadelhoffer, 2006c:

> *Dice (Roll)*: Brown is playing a simple game of dice. The game requires that Brown roll a six to win. So, hoping to get a six, Brown throws a die onto the table. Unluckily for the other players, the die lands six-up and Brown wins the game. Question: Did Brown intentionally *roll a six*?

> *Dice (Win)*: Brown is playing a simple game of dice. The game requires that Brown roll a six to win. So, hoping to get a six, Brown throws a die onto the table. Unluckily for the other players, the die lands six-up and Brown wins the game. Question: Did Brown intentionally *win the game*?

Faced with the first scenario, only 10% of participants answered that Brown intentionally rolled a six (a neutral event). Faced with the second case, 62.5% of participants answered that Brown intentionally won the game (a non-morally good event that we would expect the agent to desire). How can we account for this difference? Our hypothesis is that asking whether 'Brown intentionally won the game' redirects attention to Brown's desire to win the game (eliciting Meaning 1)

whereas asking whether 'Brown intentionally rolled a six' redirects attention to the fact that winning the game requires luck (eliciting Meaning 3).[18]

7. Other Empirical Data

Finally, in this section, we examine some empirical data that are not related to any particular meaning and try to account for them.

7.1 The Nazi Law Cases

Knobe (2007) designed a puzzling pair of cases. Here is the first case:

> *Nazi Law (Violation)*: In Nazi Germany, there was a law called the 'racial identification law.' The purpose of the law was to help identify people of certain races so that they could be rounded up and sent to concentration camps. Shortly after this law was passed, the CEO of a small corporation decided to make certain organizational changes. The Vice-President of the corporation said: 'By making those changes, you'll definitely be increasing our profits. But you'll also be *violating* the requirements of the racial identification law.' The CEO said: 'Look, I know that I'll be *violating* the requirements of the law, but I don't care one bit about that. All I care about is making as much profit as I can. Let's make those organizational changes!' As soon as the CEO gave this order, the corporation began making the organizational changes.

In the *Fulfillment* case, all occurrences of 'violating' were replaced by 'fulfilling.' In the *Violation* case, 81% of participants said the CEO intentionally violated the requirements of the law, while only 30% of participants said he intentionally fulfilled the requirements of the law in the *Fulfillment* case. At first sight, it seems as if our account cannot explain this asymmetry, since, from a normative point of view, we would expect the CEO to have the desire to violate the (Nazi) law and to be reluctant to fulfill its requirements.

However, these normative considerations turn out not to be very salient to participants. In Knobe's original experiments, participants had to judge how much blame or praise the CEO deserved, using a scale from −3 ('a lot of blame') to 3 ('a lot of praise'), with the 0 point marked 'no blame or praise'. The mean answer was −1.7 in the *Fulfillment* case and −0.9 in the *Violation* case, a difference that wasn't significant. Even when asked how much blame or praise the CEO

[18] This strongly suggests the following rule:

> *Principle of the priority of descriptive expectations over possibilities of failure: ceteris paribus*, descriptive expectations are more salient than possibilities of failure

Nevertheless, things seem to go in the exact opposite directions in the *Bull's-Eye (No Skill)* case. As suggested by Sripada's version of this case, it seems that there is no straightforward hierarchy in salience between descriptive expectations and possibilities of failure. It all depends on which is the more salient in a given context.

would have deserved if he had been specifically trying to violate the law, the mean answer was 0.3, showing that participants didn't consider the violation of the law as morally good or bad, but rather as morally neutral. These results suggest that, in the *Violation* case, there is no salient *normative* expectation. But, on the contrary, *descriptive* expectations are multiple. First, there is the descriptive expectation that the CEO will be reluctant to violate the law, because it will get him into trouble. Furthermore, there is a non-zero chance that he is a Nazi, and, if he is, then he is likely to have the desire that the law be respected. Thus, the CEO is descriptively expected to be significantly more *reluctant* to violate the law than to have the desire to do so, while there is no strong normative expectation. So, Meaning 2 is preferentially activated and, because the CEO doesn't care, we consider his violating the law as intentional.

In the *Fulfillment* case, the same reasons generate a salient *descriptive* expectation: we expect a reasonable man to prefer not to get into trouble with the Nazi authorities, for his own sake, and thus to have the desire to fulfill the requirements of the law. Surely, there is also the *normative* expectation that the agent will not have the desire to fulfill the law, but we have seen that this expectation is too weak to generate a normative expectation that the agent will have the desire *not to* fulfill the law (that is: to violate it). So, in the *Fulfillment* case, Meaning 1 is the most salient, and because the CEO does not have the desire to fulfill the requirements of the law, most participants consider its action as non-intentional.

7.2 Means are More Intentional than Side-Effects

Overall, means are judged more intentional than side-effects, even if, as we already mentioned, an asymmetry similar to the Knobe Effect can be found at the level of means (Cova and Naar, forthcoming a). How can we account for this difference?

The fact is that Meaning 1 and Meaning 2 are sensitive to how strong the agent's desire about the occurrence of an event is: the more the agent desires an event to occur, the more intentional his bringing it about will be. Now, when an event is a means rather than a side-effect, it is more strongly desired, because in addition to the intrinsic desirability of the outcome, the means can also be the object of an instrumental desire. If and when it is, the strength of the instrumental desire for the means will depend on the strength of the desire for the goal. So, because (i) means are the objects of stronger desires than are side-effects (by definition) and (ii) the stronger the desire for an event, the more the causation of the event is judged intentional, according to Meanings 1 and 2, means will be considered more intentional than side-effects.

8. Conclusion: Implications for Folk Psychology

The Knobe and Skill Effects have been taken by many as a proof that folk psychology is deeply influenced by moral considerations. Some have taken this influence to be

a bias (Nadelhoffer, 2006) while others, such as Knobe, have concluded that the purpose of folk psychology is not only to provide a quasi 'scientific' description of the world but that it also has an evaluative function (Knobe, 2006, 2010). Some have rejected this view, arguing that these effects could be explained without postulating an impact of moral considerations on ascriptions of mental states (Machery, 2008; Sripada, 2010).

Where do our hypotheses stand in this debate? We grant that moral considerations and normative expectations do influence participants' answers. However, we do not think that this proves that moral considerations have an impact on folk psychology and mental state attributions. Moral considerations have an impact on what speakers mean when they use the English adverb 'intentionally', i.e. on the choice of the concept that they link to this particular word. In the experiments under discussion, the task of a participant is to judge whether an agent whose action is described in a scenario caused an effect intentionally. Processing information about the agent's attitudes conveyed by the scenario is likely to cause a participant to produce a moral evaluation of the agent, which in turn will affect the participant's selection of a particular meaning of the adverb 'intentionally'. But it is one thing to account for the conditions under which English speakers are willing to use this adverb when asked to judge whether the agent caused some effect intentionally. Another thing is the speaker's ability to ascribe beliefs, desires, intentions and emotions to the agent. For example, two persons may share the very same beliefs about an agent's mental states, objective control over his action in the *Aunt (No-Skill)* case, but they could still give very different answers to the question 'did Jake intentionally kill his aunt?' One could be using Meaning 2, the other Meaning 3. This clearly shows that a speaker's mindreading ability (i.e. her ability to ascribe beliefs, desires and intentions to an agent) is one of the necessary conditions explaining her selection of one or another meaning of the English adverb 'intentionally' in response to the question. But it is far from being sufficient. The fact that the moral evaluation of the agent affects the selection of one of the two meanings of 'intentionally' fails to establish that the speaker's mindreading ability itself is shaped by moral considerations.

So, according to our theory, moral considerations may play a role at a linguistic, expressive level, but not at the level of folk psychological attribution of mental states to others. It would be a mistake not to recognize the distinction between how people *think* (about others' thoughts) and how they verbally *report* their thoughts (about others' thoughts) (Cova, Dupoux and Jacob, 2010; Egré, 2010).

If our hypothesis is correct, then this also means that there is not a single, irreducible, folk psychological concept of INTENTIONAL ACTION. We have concepts of pro-attitudes (e.g. desires), beliefs and goal-directed actions. Some combinations of these concepts can be expressed using the word 'intentionally' in some particular contexts. But studying the use of the word 'intentionally' with the hope of directly probing a key component of our folk psychology is highly problematic. According to our account, studying the Knobe Effect can teach us a lot about our linguistic use

of the word 'intentionally' and indirectly, but not directly, about theory-of-mind and/or folk psychology.

Département d'Etudes Cognitives,
Ecole Normale Supérieure

Institut Jean Nicod,
Centre National pour la Recherche Scientifique

Laboratoire de Sciences Cognitives et Psycholinguistique,
Centre National pour la Recherche Scientifique

References

Adams, F. and Steadman, A. 2004a: Intentional action in ordinary language: core concept or pragmatic understanding. *Analysis*, 64, 173–81.

Adams, F. and Steadman, A. 2004b: Intentional action and moral considerations: still pragmatic. *Analysis*, 64, 264–67.

Adams, F. and Steadman, A. 2007: Folk concepts, surveys and intentional action. In C. Lumer and S. Nannini (eds), *Intentionality, Deliberation and Autonomy: The Action-Theoretic Basis of Practical Philosophy*. Aldershot: Ashgate Publishers.

Biro, S., Csibra, G. and Gergely, G. 2007: The role of behavioral cues in understanding animacy, agency and goal-directed actions in infancy. In C. von Hofsten and K. Rosander (eds), *Progress in Brain Research: From Action to Cognition*, *Vol. 164*, 303–22. Amsterdam: Elsevier.

Biro, S. and Leslie, A. 2007: Infants' perception of goal-directed actions: development through cue-based bootstrapping. *Developmental Science*, 10, 379–98.

Bratman, M. E. 1987: *Intention, Plans, and Practical Reason*. Cambridge, MA: Cambridge University Press.

Cova, F., Dupoux, E. and Jacob, P. 2010: Moral evaluation shapes linguistic report of others' psychological states, not theory-of-mind judgments. *Behavioral and Brain Sciences*, 33, 334–35.

Cova, F. and Naar, H. forthcoming a: Side-effect effect without side effect: the pervasive impact of moral considerations on judgments of intentionality. *Philosophical Psychology*.

Cova, F. and Naar, H. forthcoming b: Testing Sripada's deep self-model. *Philosophical Psychology*.

Csibra, G. 2008: Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, 107, 705–717.

Cushman, F. and Mele, A. 2008: Intentional action: two and half folk concepts. In J. Knobe and S. Nichols (eds), *Experimental Philosophy*. New York: Oxford University Press.

Egré, P. 2010: Qualitative judgments, quantitative judgments and norm-sensitivity. *Behavioral and Brain Sciences*, 33, 335–6.

Feltz, A. 2007: Knowledge, moral praise, and moral side effects. *Journal of Theoretical and Philosophical Psychology*.

Guglielmo, S. and Malle, B. F. 2010: Can unintended side effects be intentional? Resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin*, 36, 1635–47.

Knobe, J. 2003a: Intentional action and side-effects in ordinary language. *Analysis*, 63, 190–93.

Knobe, J. 2003b: Intentional action in folk psychology: an experimental investigation. *Philosophical Psychology*, 16, 309–24.

Knobe, J. 2006: The concept of intentional action: a case study in the uses of folk psychology. *Philosophical Studies*, 130, 2, 203–31.

Knobe, J. 2007: Reason explanation in folk psychology. *Midwest Studies in Philosophy*, 31, 90–107.

Knobe, J. 2010: Person as scientist, person as moralist. *Behavioral and Brain Sciences*, 33, 315–29.

Knobe, J. and Burra, A. 2006: Intention and intentional action: a cross-cultural study. *Journal of Culture and Cognition*, 1–2, 113–32.

Lanteri, A. 2009: Judgments of intentionality and moral worth: experimental challenges to Hindriks. *The Philosophical Quarterly*, 59, 713–20.

Leslie, A., Knobe, J. and Cohen, A. 2006: Acting intentionally and the side-effect effect: 'theory of mind' and moral judgment. *Psychological Science*, 17, 421–27.

Machery, E. 2008: The folk concept of intentional action: philosophical and experimental issues. *Mind & Language*, 23, 165–89.

Malle, B. F. 2006: Intentionality, morality, and their relationship in human judgment. *Journal of Cognition and Culture*, 6, 87–112.

Malle, B. F. and Knobe, J. 1997: The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, 101–21.

Mele, A. and Cushman, F. 2007: Intentional action, folk judgments and stories: sorting things out. *Midwest Studies in Philosophy*, 31, 184–201.

Nadelhoffer, T. 2004a: Praise, side effects and intentional action. *Journal of Theoretical and Philosophical Psychology*, 24, 196–213.

Nadelhoffer, T. 2004b: Blame, badness and intentional action: a reply to Knobe and Mendlow. *Journal of Theoretical and Philosophical Psychology*, 24, 259–69.

Nadelhoffer, T. 2005: Skill, luck, control and intentional action. *Philosophical Psychology*, 18, 5, 343–54.

Nadelhoffer, T. 2006: Bad acts, blameworthy agents and intentional actions: some problems for jury impartiality. *Philosophical Explorations*, 9, 2, 203–20.

Nadelhoffer, T. 2006c: Foresight, moral considerations and intentional actions. *Journal of Cognition and Culture*, 6, 1, 133–58.

Nadelhoffer, T. 2007: Fringe benefits, side effects and intentional actions: a reply to Feltz. *Journal of Theoretical and Philosophical Psychology*, 27, 801–09.

Nanay, B. 2010: Morality of modality? What does the attribution of intentionality depend on? *Canadian Journal of Philosophy*, 40, 28–40.

Nichols, S. and Ulatowski, J. 2007: Intuitions and individual differences: the Knobe effect revisited. *Mind & Language*, 22, 346–65.

Pacherie, E. 2000: The content of intentions. *Mind & Language*, 15, 400–32.

Pellizzoni, S., Siegal, M. and Surian, L. 2009: Foreknowledge, caring and the side-effect effect in young children. *Developmental Psychology*, 45, 289–95.

Pellizzoni, S., Girotto, V. and Surian, L. 2010: Beliefs and moral valence affect intentionality attributions: the case of side effects. *Review of Philosophy and Psychology*, 1,201 -09.

Phelan, M. and Sarkissian, H. 2008: The folk strike back: or, why you didn't do it intentionally, though it was bad and you knew it. *Philosophical Studies*, 138, 291–98.

Phelan, M. and Sarkissian, H. 2009: Is the trade-off hypothesis worth trading for? *Mind & Language*, 24, 164–80.

Searle, J. 1983: *Intentionality*. Cambridge: Cambridge University Press.

Sousa, P. and Holbrook, C. 2010: Folk concepts of intentional action in the contexts of amoral and immoral luck. *Review of Philosophy and Psychology*, 1, 3, 351–70.

Sripada, C. 2010: The Deep Self model and asymmetries in folk judgments about intentional action. *Philosophical Studies*, 151, 2, 159–76.

Sargent, M.J., Tannenbaum, D., Ditto, P. H. and Pizarro, D. A. ms: Motivated reasoning and the assessment of intentionality for outgroup members.

Sripada, C. and Konrath, S. 2011: Telling more that we can know about intentional action. *Mind & Language*, 26, 353–80.

Sverdlik, S. 2004: Intentionality and moral judgments in commonsense thoughts about action. *Journal of Theoretical and Philosophical Psychology*, 24, 224–36.

Tannenbaum, D., Ditto, P. H. and Pizarro, D. A. ms: Different moral values produce different judgments of intentional action.

Wible, A. 2009: Knobe, side effects, and the morally good business. *Journal of Business Ethics*, 85, 173–78.

Wright, J. and Bengson, J. 2009: Asymmetries in folk judgments of responsibility and intentional action. *Mind & Language*, 24, 237–51.

Young, L., Cushman, F., Adolphs, R., Tranel, D. and Hauser, M. 2006: Does emotion mediate the effect of an action's moral status on its intentional status? *Journal of Cognition and Culture*, 1–2, 291–304.

Zalla, T. and Machery, E. ms: The concept of intentional action in Asperger Syndrome.