Intentional Side-Effects of Action*

Robin Scaife

Department of Philosophy, University of Sheffield
45 Victoria Street, Sheffield S3 7QB, UK
r.scaife@sheffield.ac.uk

Jonathan Webber

School of English, Communication, and Philosophy, Cardiff University
Humanities Building, Colum Drive, Cardiff CF10 3EU, UK
webberj1@cardiff.ac.uk

Abstract

Certain recent experiments are often taken to show that people are far more likely to classify a foreseen side-effect of an action as intentional when that side-effect has some negative normative valence. While there is some disagreement over the details, there is broad consensus among experimental philosophers that this is the finding. We challenge this consensus by presenting an alternative interpretation of the experiments, according to which they show that a side-effect is classified as intentional only if the agent considered its relative importance when deciding on the action. We present two new experiments whose results can be explained by our hypothesis but not by any version of the consensus view. In the course of doing so, we develop a methodological critique of the previous literature on this topic and draw from it lessons for future experimental philosophy research.

Keywords

Intentional action; Side-effects; Experimental philosophy; Joshua Knobe; Moral psychology

Recent empirical research into the folk classification of the outcomes of actions as intentional is usually taken to show that such classification has an irreducibly normative dimension. Various interpretations of the experimental data have in common the claim that whether the side-effect of an action counts as intentional depends on some normative valence of that

* Research for this paper was stimulated by discussions at the November 2007 workshop of the AHRC Culture and Mind project at the University of Sheffield. We are grateful to the organiser, Stephen Laurence, and to Joshua Knobe for encouraging this line of criticism. We are also grateful to the participants of Cardiff Philosophy Work-in-progress Seminar, at which a version of this paper was presented, and to Clea Rees and two anonymous referees for this journal for pointing out difficulties with the details in an earlier draft.

side-effect.¹ This is the way that Joshua Knobe, for example, whose experimental research started this debate, understands the data. Some critics of this view claim that the experiments indicate only a bias in the folk application of the concept rather than an aspect of the concept itself. A more radical criticism denies that we should explain the data with reference to the normative valence of the side-effect, claiming instead that whether an effect is classified as intentional depends on its role in the agent's reasoning. Edouard Machery has advanced a version of this view, although strong evidence has been presented against his position.

Our aim in this paper is to argue for a new version of the view that neither the folk concept of intentional action nor its usual application has a normative dimension: the data rather show that a side-effect is classified as intentional only if it is understood to have been taken into consideration in the deliberation culminating in that action. We present two new experiments and argue that only our hypothesis can explain their results as well as the existing experimental data.

In denying that there is a normative dimension to the folk concept of intentional action or to its application, we undermine the use of this literature to motivate wider philosophical and jurisprudential claims. It has been argued, for example, that this literature demonstrates just one aspect of the pervasive influence that moral assessment has on judgments about other people's mental states and that we should therefore reconsider whether folk psychology is primarily aimed at explaining and predicting behaviour (see Knobe 2006; Nichols and Knobe 2008; Pettit and Knobe 2009). One application of this concerns the role of juries in criminal trials: if people's judgments of whether an outcome was intentional reflect their moral assessment of that outcome rather than their assessment of the agent's state of mind, then perhaps we ought to abandon the practice of asking juries whether they consider the defendant to have acted intentionally (Nadelhoffer 2006a).

Our preferred explanation of the data implies that these concerns are misplaced. For our hypothesis is that the experiments discussed in this literature show that people generally classify a side-effect of an action as intentional only if they see the agent as having taken that side-effect into consideration, where this means that the agent assigned that side-effect

¹ Throughout this paper, we use 'side-effect' to mean any outcome of an action other than the agent's primary purpose, thus following the usual use in this debate, rather than following Nadelhoffer (2007) in distinguishing positive 'fringe benefits' from negative 'side effects'.

some level of importance relative to the importance they assigned to their primary objective. Our argument for this is in six stages. In the first section, we explain the consensus interpretation in terms of the normative valence of the side-effect. In the second, we develop our own interpretation by considering the evidence against Machery's similar interpretation. In the third, we formulate three hypotheses for explaining the existing data: one of these is our own; the other two are intended to capture the two different ways in which the normative valence of the side-effect could be central. The fourth introduces our first experiment and argues that only one of the two hypotheses opposed to ours can explain its results. The fifth shows how our hypothesis explains the results of our first experiment, and in so doing develops our hypothesis further. The sixth introduces our second experiment and shows that our hypothesis can explain our results whereas its remaining rival cannot.

Over the course of this argument for our claim, we also develop a methodological critique of previous literature in this debate. Discussions of these experiments, we claim, have implicitly assumed that switching one normative term in a story for a term with its opposing normative valence makes no difference to the reader's understanding of other sentences in that story. We argue that the design and interpretation of these kinds of experiment should take account of the fact that context influences the way a reader is likely to disambiguate an ambiguous statement. In the final section of this paper, we draw methodological lessons from the details of this critique.

1. The Normative Valence of the Side-Effect

Recent debate over the folk concept of intentional action revolves around a series of pen-and-paper experiments. In the classic form of the experiment, each participant is given one of two vignettes in which an action brings about a side-effect foreseen by the agent. The vignettes differ only in the moral valence which the participant is presumed to ascribe to the side-effect. The participant is asked whether the side-effect is intentional. For example, here is a vignette from Knobe's original (2003a, 191) version of this experiment:

The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, and it will also 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.

Participants given this vignette were asked whether the chairman of the board had harmed the environment intentionally and were required to answer simply 'yes' or 'no'. Other participants were given a vignette identical to this one except that the term 'harm' had been replaced with 'help'. They were asked whether the chairman had helped the environment intentionally. When asked about harming the environment 82% of participants judged it to be intentional, but when asked about helping the environment this figure fell significantly to just 23%. Knobe checked the results with a second experiment, which transposed the same questions into a different scenario, and concluded that people 'seem considerably more willing to say that a side-effect was brought about intentionally when they regard that side-effect as bad than when they regard it as good' (2003a, 193).

Further studies have produced the same asymmetry in cases where the vignettes differ not in the side-effect's presumed moral valence for the experimental participant, but in its practical valence for the agent in the story. Where the action of increasing overall sales has the side-effect of decreasing sales in one part of the country, for example, this side-effect is (considered purely in itself) practically bad for the agent and is generally classified as intentional (Knobe and Mendlow 2004). Similarly, where an action has the side-effect of violating some law that the agent does not care about but which is a law that the experimental participant can reasonably be expected to find abhorrent, the side-effect is generally classified as intentional (Knobe 2007, Appendix). The asymmetry in judgments of intentional action cannot be explained in terms of moral evaluation of the side-effect, therefore. Knobe has proposed that we instead understand it in terms of the immediate classification of the side-effect as violating some basic norm, even when participants' considered views might be that it is permissible or even best that the norm be violated in this instance (Knobe 2007, § 9). Such norms might be moral, but might equally be practical.

The central controversy over this account of the data concerns Knobe's claim that these norm-driven judgments are correct applications of the concept of intentional action. We should not think of the concept as purely descriptive, he argues, but as having an irreducible normative dimension (e.g. Knobe 2006, § 5). Opponents argue that the data should not be understood to track the contours of the concept of intentional action, but rather to reveal ways in which applications of that concept can be inappropriately

biased. One suggestion is that participants classify certain side-effects as intentional because they do not want to be taken to be exonerating the agent: denying that an effect is intentional is often a way of denying that the agent is responsible for it, even though it is generally agreed that one can be responsible for unintentional effects (Adams and Steadman 2004a, 2004b). A similar suggestion is that our negative emotional responses to certain side-effects lead us to classify them as intentional in order to blame the agent (Nadelhoffer 2006a, esp. § 4). This debate is sophisticated and showing no sign of imminent resolution (see Knobe 2006; Nichols and Ulatowski 2007; Machery 2008), but it presupposes that the asymmetry in the data is to be explained by the participants describing certain side-effects as intentional because of their negative normative valence.

2. The Deliberation Behind the Action

An alternative approach is to explain the data in terms of the role played by the relevant consideration in the agent's deliberation. Frank Hindriks offers an account of this kind. He argues that a side-effect is intentional if it is given some significance in the deliberation leading up to the action or if it ought to have been given some such significance but was not: people classify the chairman's helping the environment, in Knobe's original experiment, as unintentional because the chairman 'is not motivated to help the environment', but classify harming it as intentional because the chairman 'fails to be moved by a consideration to which he should attach negative significance, the harm that will be done to the environment' (2008, 635).² Hindriks explains the asymmetry partly in normative terms, therefore. Although this normativity is not directly a matter of the valence of the side-effect, it does seem to be so indirectly. For we are owed an explanation of just why some side-effects ought to be considered in deliberation whereas others need not be. In describing his view as providing a 'deeper

² Hindriks seems to vacillate, in fact, between this claim that the harm is intentional because the chairman ought to consider it in his deliberation (see also p. 631) and the claim that it is intentional because the chairman *believes that* he ought to take it into consideration but still does not do so (p. 634; statement DST on p. 635; pp. 637–8). Since the experimental vignette does not warrant the imputation of this belief to the chairman, or even suggest that he might have such a belief, this latter claim seems implausible. Moreover, the considerations presented in this paper against the former claim are equally considerations against the latter.

explanation' than Knobe provides and as explaining why 'the moral character of the effect can influence judgments about intentional action', Hindriks admits that his view is a version of the consensus view that a side-effect is classified as intentional only when it has some negative normative valence (2008, 640–1).

Our aim in this paper is to argue for an explanation of the classification of some but not all side-effects as intentional that does not rely on, or turn out to be a more sophisticated account of, the differing normative valences of those side-effects. In so doing, we hope to show that the concept of intentional action is a wholly descriptive concept after all, that its role in explanation and prediction of behaviour does exhaust its content. Edouard Machery (2008, § 3) has provided one such account, though this has met with a forceful objection. Our preferred view is distinct from Machery's, however, and reflection on the shortcomings of his account will help to explain and motivate our alternative. For our own claim is that a side-effect is intentional only if it was considered in the deliberation behind the action, whereas Machery proposed that a side-effect is intentional only if it is a cost. In the rest of this section, we hope to show only that the objection to Machery's 'trade-off hypothesis' is no objection to our view.

In the case of Knobe's original experiment, Machery's claim is that participants see harming the environment as a cost outweighed by the benefit of increased profits but see helping the environment as no cost at all. Since people generally consider costs incurred in pursuit of benefits to be incurred intentionally, the harm is intentional, whereas there is no parallel reason to classify the help as intentional. To test this hypothesis, Machery devised a vignette in which Joe orders a smoothie of the largest size available and is then told by the cashier either that the price had risen by one dollar or that smoothies of this size currently come with a free commemorative cup. Joe declares that he does not care about this and orders the smoothie. When asked whether Joe had intentionally paid an extra dollar, 95% answered affirmatively. But when asked whether he intentionally obtained a commemorative cup, only 45% answered affirmatively. Machery explicitly allows his account to be ambiguous over *who* sees the side-effect as a cost incurred for a benefit: it might be that participants classify the side-effect as intentional when they consider *the agent* to see the side-effect in this way, or it might be that they classify it as intentional when *they themselves* see the side-effect in this way (2008, 177 n10). But neither reading of Machery's data is compatible with the results of a further experiment conducted by Ron Mallon in response.

Mallon designed vignettes in which the side-effect is seen by the agent as good and occurs in pursuit of a goal that the agent sees as good, where the participants of the experiment are reasonably expected to see the goal as bad. The agent does not see the side-effect as a cost, therefore, and the participant does not see the goal as a benefit; neither agent nor participant sees the side-effect as a cost incurred for a benefit. In one example, the agent is a terrorist who aims to kill as many Americans as possible. After learning of a plan that will kill some Americans but also kill some Australians, in one version of the story, or will indirectly benefit a local orphanage by lowering property prices in the area, in the other version of the story, the terrorist replies that the side-effect is indeed good but 'I don't really care about that. I just want to kill as many Americans as possible'. When asked whether the terrorist intentionally killed Australians, 92% of participants responded affirmatively, whereas only 12% responded affirmatively when asked whether the terrorist had intentionally benefited the orphanage (2008, § 2).

There is something disingenuous, however, about the terrorist's claim that benefiting the orphanage would indeed be good. It is difficult to square a concern for the welfare of orphans with a plan to kill adults, especially since the plan will probably orphan more children. Consider a variation on Mallon's story, in which the plan to bomb a nightclub in order to kill one hundred Americans is about to go ahead when it becomes apparent that if a different nightclub were bombed instead one hundred Americans would still die but there would also be an effect either of killing some Australians or of benefiting an orphanage by depressing the property market. All other things being equal, would the additional effect plausibly give the terrorist reason to change plan? If the terrorist thinks that killing Australians is good, then it seems that it would. But it is far less plausible that a terrorist would change the plan in order to indirectly benefit an orphanage. The same can be said about Mallon's second experiment, in which the story is different but the goal of the experiment, and indeed the results, remain the same (2008, § 2). In this case, a gang leader is considering whether to flood the local area with cheap cocaine, the relevant side-effects being that more police will die in drug-related violence or that the local addicts will have more money for food and housing. In both cases, the gang leader is said to consider these side-effects good, but it is far more plausible that the gang leader hates the police than that the gang leader is concerned about the welfare of the very addicts exploited by the gang.

This criticism of Mallon's vignettes does not impugn his argument against Machery. For all that argument requires is data showing that a

side-effect is considered intentional when the agent considers it good and when the participant considers the overall goal of the action bad. Such results are obtained where the side-effect of the terrorist's action is killing Australians and where the side-effect of the gang leader's action is that more police are killed. Neither of these cases can be cast as the intentional incurring of a cost in pursuit of a benefit. But what these concerns about the vignettes do undermine is any claim that might be made on their basis against our thesis that a side-effect is intentional only if it is taken into consideration in the deliberation behind the action. For when the terrorist says that killing Australians is good, or when the gang leader says that the increase in deaths of police officers would be good, this indicates that they accord these considerations some weight, though they go on to say that this weight is minor compared to that of the central goal. But when the terrorist says that helping the orphanage is good, or the gang leader says it is good that addicts have more money, this sounds insincere or even ridiculous: it certainly does not sound as though these characters genuinely consider these factors to carry any weight in their deliberations.

Herein lies a general lesson for the construction of experimental vignettes: the same words spoken by a character in the story do not necessarily have the same meaning, or give the reader the same impression, when the surrounding story has changed. It is generally assumed in this debate that when Knobe switches the term 'harm' for 'help' in his original experiment, for example, he has preserved the rest of the vignette precisely. But this need not be so. For the chairman's declaration that 'I don't care at all about *x*. I just want to make as much profit as I can' is ambiguous. It can be read as a rhetorical declaration that profit is so overwhelmingly important a consideration that *x* can make no significant difference to the decision. Or it can be read more literally as a statement that the chairman is not even going to consider the importance of *x*.

Altering the context in which the statement occurs can alter the way in which the reader is likely to disambiguate it. Where *x* is a consideration the reader assumes is generally considered to oppose the other consideration in play, this statement is likely to be disambiguated to mean that this countervailing consideration is trumped by that other consideration. But where the reader assumes *x* is generally considered to be neutral or in harmony with the other consideration in play, they are more likely to take the claim not to care about it literally. So in switching 'harm' for 'help', Knobe has altered the probable disambiguation of the chairman's words, from a declaration that environmental harm is outweighed to a declaration that environmental benefit is no consideration at all.

3. Normative and Descriptive Hypotheses

We intend to show that the data in the current experimental debate over the concept of intentional action is best explained by what we will call the Consideration Hypothesis (CH). Rather than displaying any normative dimension of the folk concept itself, or of the way in which it is usually applied, we argue, the data is best interpreted as manifesting the role of deliberation and decision in intentional action. CH is the view that an outcome of an action is intentional only if it was taken into consideration in the deliberation leading to the action. By ‘taken into consideration’, we mean that the agent has weighed the value of the outcome against the values of other considerations in deciding what to do. According to CH, people see Knobe’s chairman *either* as according little importance to environmental harm in comparison to profit *or* as refusing to even consider the relative importance of environmental benefit. Similarly, on this view, people see Machery’s Joe as deciding that the smoothie is worth the extra dollar on this occasion, but as not even considering the value of the free cup. And CH claims that people read Mallon’s terrorist as according some value to killing Australians but not as seeing the impact on the local orphanage as a serious consideration at all.

Rather than survey all of the published variations on these experiments, we will argue for our hypothesis by presenting new experimental results which cannot be explained by the opposing interpretations. Our hypothesis, CH, holds that being taken into consideration in the deliberation behind the action in the way we have explained is a necessary condition for an outcome being classified as intentional. We will not address here the further question of whether it is sufficient. That view faces a challenge from research that suggests that the roles played by luck and skill in bringing about the outcome also affect whether people classify it as intentional (Malle and Knobe 1997; Knobe 2003b). We leave this question for another time. Our claim here is just that the data so far adduced in favour of the claim that there is a normative dimension to the folk classification of an outcome as intentional in fact shows only that CH is true.

Interpretations of the data in terms of the normative valence of the side-effect can be divided into two kinds. The more common kind holds the data to show that an outcome is more likely to be classified as intentional if it has some negative normative valence. We will call this the Effect Evaluation Hypothesis (EEH). We intend our definition of EEH to be sufficiently broad to encompass both the view that this negative normative valence is *necessary* for such classification and the weaker view that it *often influences* such

classification. It also encompasses both the view that this condition or influence reflects part of the concept of intentional outcomes and the opposing view that it reflects a bias affecting our application of the concept. We have therefore already seen variations of EEH voiced by Knobe (2003a, 2007), Knobe and Mendlow (2004), Adams and Steadman (2004a, 2004b), and Hindriks (2008). Various others working in this area subscribe to a form of EEH, including Nichols and Ulatowski (2007).

An alternative view holds that it is not the normative valence of the side-effect itself that has this influence, but rather some relation between this and the normative valence of the action as a whole. This view seems to be implied by descriptions of the data in terms of blame (e.g. Knobe 2003a, Nadelhoffer 2004a). For if the side-effect renders the agent blameworthy, then it must render the whole action bad. Otherwise, people could be blameworthy for actions that are not bad. One form of this view would see the action as a whole rendered bad by the badness of the bad outcome outweighing the combined goodness of any good ones, though we are unaware of anyone propounding this interpretation of the data. Another form, suggested by Nadelhoffer (2004b, 268), holds the action to be bad because it had a bad outcome and was motivated by mental states for which the agent is blameworthy, such as not caring about harming the environment.³ We will treat such actual and possible views of the data together, calling their common thought the Action Evaluation Hypothesis (AEH).

It might seem that AEH is already under pressure from an experiment described earlier. Where the action of increasing overall sales has the side-effect of decreasing sales in one place, as we have seen, this side-effect is generally classified as intentional even though the overall action is neither morally bad nor practically bad for the agent. One way to accommodate this example would be to construe AEH as allowing outcomes to be intentional when they were incurred at a *risk* of the overall action turning out bad: through bad luck, this action might have decreased sales in one area without increasing them elsewhere. Another would be to construe AEH as holding not that the relation between the valence of the outcome and the overall valence of the action is *necessary* for the classification of the outcome as intentional, but as holding only that it *can lead or always leads to*

³ Nadelhoffer does not restrict his proposal to cases of a bad outcome and an agent blameworthy for their motivating mental states, but suggests that the same effect occurs when the outcome is good and the agent praiseworthy (2004b, 268). We ignore this aspect of his position for the sake of simplicity, since it makes no difference to our argument.

this classification (see Nadelhoffer 2004b, 259–6). We intend AEH to be understood sufficiently loosely as to encompass both these possibilities, parallel to the loose specification of EEH.

Our strategy is to argue that CH is the preferable of the three hypotheses on grounds of its greater explanatory power. In the next section, we present a novel variation on Knobe's experiment: each vignette involves not one but two side-effects, but each participant is asked only about one of them. We argue that our results generate two explanatory questions that cannot be answered by EEH. These questions can be answered clearly by each of AEH and CH, though AEH can provide this explanation only by making two contentious assumptions. Rather than address the plausibility of these assumptions, the next stage of our argument involves a variation on Knobe's experiment in which the negative side-effect was necessary to the pursuit of a good goal. The results of this experiment can be explained by CH but not by AEH.

4. Adding Another Side-Effect

If the classification of an outcome depends on its normative valence without reference to the overall valence of the action, as EEH claims, then we should expect the same classifications even when the action also has a second side-effect of the opposite valence. Our first experiment therefore aimed to test this prediction. In order to control for any influence of the two side-effects being of importantly different kinds, we used two versions of the vignette: in one, the side-effects were increased pollution and improvement in safety; in the other, decreased pollution and lowering of safety. Here is one version of our vignette:

The vice president of the company went to the chairman of the board and said: 'We're thinking of changing the way the factory works. There are three factors to consider: it will increase profits, it will improve safety, but it will increase pollution'.

The chairman of the board answered: 'All I care about is increasing profits, so let's do it'.

So they altered the factory and, sure enough, this had the effects the vice-president had predicted.

In the other version, the three factors for the chairman to consider were: 'it will increase profits, it will lower pollution, but it will make the factory less safe'. In both vignettes, the chairman claimed to care only about increasing profits. Each participant was asked one of the following questions: whether

the chairman had intentionally increased pollution, had intentionally improved safety, had intentionally decreased pollution or had intentionally made the factory less safe. Participants were undergraduate students in the first year of a Philosophy degree at the University of Sheffield or in the second year of a Geography degree at the University of Bristol. The results are given in Table 1.

In the context of the existing debate, the most striking result in this table is the top row: the difference between the proportion of those classifying the increase in pollution as intentional and those not classifying it as intentional is not significant, according to the chi-squared test. This is precisely the negative side-effect that Knobe asked about in his original experiment. Where he found that participants were significantly likely to judge it to be intentional, we have found that introducing a positive side-effect alongside it eliminates this likelihood. To ensure that this result was robust, we ran the experiment again with this vignette and question only. Participants were 56 undergraduate students in the first year at Cardiff University and at the University of Sheffield. Of the respondents 51.8% classified the increase in pollution as intentional and the remaining 48.2% did not; this difference was again not significant, $\chi^2=(1, N=56)$ 0.07, $P>0.1$. Combining this with our first result gives a sample of 98, of whom 49% classified the increased pollution as intentional and 51% did not, again an insignificant difference $\chi^2=(1, N=98)$ 0.04, $P>0.1$.

This is a finding that stands in need of explanation: why does the introduction of a positive side-effect alongside the negative one eliminate the significant likelihood of participants classifying the negative one as intentional? Notice that this question cannot be answered by any version of EEH. For what they have in common, what makes them versions of EEH, is that they explain the experimental results solely in terms of the normative valence of the side-effect that the participant is asked about. It is therefore

Table 1 Numbers and Percentages of Yes/No responses in our Chairman vignettes.

Outcome	Sample	Yes	No	% Yes	% No	χ^2 Test
increase pollution	42	19	23	45.2	54.8	$\chi^2=0.38$, $P>0.1$
improve safety	44	10	34	22.7	77.3	$\chi^2=13.09$, $P<0.001$
decrease pollution	46	7	39	15.2	84.8	$\chi^2=22.26$, $P<0.001$
make factory less safe	44	32	12	72.7	27.3	$\chi^2=9.09$, $P<0.005$

impossible for them to explain the impact of introducing a second side-effect alongside the one which the participant is asked about. Any hypothesis that can explain this impact will, all other things being equal, be explanatorily superior, and hence preferable, to any version of EEH.

AEH seems well suited to this task. For the introduction of a side-effect with a normative valence opposing that of the side-effect asked about makes a difference to the way the participants are likely to see the overall action. When we ask about a negative side-effect, the participants are less likely to see the action that has this side-effect as bad when it also has a good side-effect than they are when it has no other known side-effect. When we asked about positive side-effects in our experiments, whether improving safety or decreasing pollution, we found that participants were significantly unlikely to classify this positive side-effect as intentional. These results are in the second and third rows of Table 1. But this does not present a problem for AEH, since AEH is not the claim that side-effects in general are likely to be classified as intentional when their actions overall are judged to be bad, but the claim that a *bad side-effect* is likely to be classified as intentional when it is judged *to render* the action bad overall.

If we compare the top and bottom rows of Table 1, we find a divergence that also stands in need of explanation. For the bottom row is consonant with Knobe's original finding where the top row is not. This difference in response patterns is statistically significant ($\chi^2=(1, N=86) 6.73, p<.01$). The negative valence of the side-effect which the participant is asked about cannot explain this finding, for this is something that the two cases have in common. So, again, no version of EEH has the resources to provide this explanation, for EEH just is the hypothesis that evaluation of that side-effect is doing all the work. Any hypothesis which can offer a unified explanation of the existing data and the divergence between these two cases, therefore, will be preferable to EEH, other things being equal.

The most obvious difference between the cases concerns the victims of the resulting harm. Making the factory less safe inflicts a risk of harm (a risk that would itself seem to be a harm) on specific identifiable individuals: those who work in the factory. Contributing to overall pollution, however, is usually a harm inflicted on nobody in particular. There may well be harm for people resulting from it, but the victims of such harm are epistemically indeterminate at the time of the chairman's decision and may always remain so, given that it is often impossible to trace a result of pollution back to a specific act of pollution. The two cases therefore involve harms of distinct moral kinds. They do not differ in degree, since examples of either sort

can be given that range from the mild to the extreme. This difference is not, moreover, some subtle philosophical consideration that it might seem implausible to suggest as underlying folk classifications: it is manifest in the clear possibility of victims of lowered safety levels taking legal action against the chairman whereas victims of increased pollution would be hard pressed to even identify the person against whom to take such action, never mind succeed in establishing culpability.

Proponents of AEH could therefore explain the divergence between the top and bottom rows of Table 1 in terms of such risk of legal action. We have already seen, in the example of decreasing profits in one area to increase them overall, that the badness of the side-effect classified as intentional need not be moral badness but can be practical badness for the agent concerned. In order to make this move, proponents of AEH need to claim that participants classify the risk of a practically bad outcome as itself something practically bad for the agent. With this commitment in place, proponents of AEH could argue that lowering safety is something with a negative moral valence and, since it includes the risk of suffering legal action, a negative practical valence for the agent, which together are sufficient for many participants to consider the overall action to have a negative normative valence. Since this practical valence is lacking in the case of increased pollution, participants are less likely to classify that outcome as intentional.

Making this move leaves AEH vulnerable to two objections. It might be argued that moral and practical valences simply cannot be aggregated in the way this explanation requires. Or it might be argued that it is implausible to suggest that participants tend to classify taking a risk upon oneself as itself having a negative normative valence irrespective of whether or not that risk in fact results in a bad outcome. At the very least, it might be argued that proponents of AEH should provide some independent empirical motivation for these claims. Rather than pursue these lines of thought, however, we will argue against AEH by presenting data from an altogether different experiment, whose results can be explained by CH but not AEH. But before doing so, we will explain in the next section how CH accommodates the data from our first experiment.

5. Taking Outcomes into Consideration

Our vignettes discussed so far preserve perfectly the ambiguity of the chairman's announcement in Knobe's original experiment. 'All I care about is increasing profits', our chairman declares, which might mean either that he

has not taken the other outcomes into consideration or that he considers them outweighed by profit. This might seem an odd way to test our hypothesis. Why not eliminate the ambiguity and simply state whether the chairman takes the side-effect into consideration? This suggestion overlooks the fact that our technical sense of ‘taken into consideration’ is not the only sense that phrase can carry in English: one could describe the decision not to even think about the value of an outcome as a minimal form of taking that outcome into consideration. Conversely, the sentence ‘the chairman did not take the pollution into consideration’ could be read as denying that the chairman even heard or understood that the plan would have this side-effect. The only way to avoid such ambiguity would be to state explicitly whether the chairman weighed up the relative importance of the relevant outcome. But the technical nature of this aspect of the story would reveal our interest in the experiment, so we would no longer be measuring how participants *actually* classify outcomes as intentional, but would instead get *their opinion* about whether their classifications are based on whether the agent took the outcome into consideration in our sense. For reasons that will become clear in the rest of this section, we should not be seeking their opinion on this matter.

Rather than attempt to eliminate the ambiguity involved in Knobe’s original experiment, therefore, we exploit that very ambiguity in explaining the data from our experiments in terms of CH, just as we did with Knobe’s experiment. Drawing attention to a positive side-effect such as improving safety or decreasing pollution, side-effects whose value the chairman need not consider since they speak in favour of the policy he will adopt anyway on other grounds, tends to lead the participants to disambiguate the chairman’s remark as indicating that he does not take safety or pollution into consideration but thinks only about profit. Participants tend to classify these positive side-effects as unintentional, that is to say, because when asked about these side-effects they tend to see the chairman as not deliberating about the various effects at all but simply as acting on his desire to increase profits.

When the participant is asked about a negative effect, on the other hand, the results are more complicated. Given that making the factory less safe involves running the risk that a worker might be injured as a result, which might have negative consequences for the company and for the chairman personally, it would seem unlikely that the chairman would not even take this factor into consideration when deciding to pursue the plan. It seems safe to assume that the participants focusing on this factor believe that the chairman would take it into consideration, that is to say, even though

the vignette does not explicitly say that he does so. Participants would here be relying on either the view that most people would take it into consideration or the view that they themselves would do so (or perhaps both). In this way, the question leads the participants to disambiguate the chairman's statement to mean that the risk is worth taking for the increased profits.

In the case of increasing pollution, however, there is no parallel risk. According to CH, this means that the chairman's declaration remains ambiguous and the participants are evenly divided over whether or not the chairman takes the increased pollution into account. But the mere absence of personal risk is not sufficient to explain why participants do not resolve this ambiguity one way or the other. For we know from Knobe's original experiment that participants do consider this outcome intentional when there is no countervailing side-effect. CH is committed, therefore, to the view that most participants think of increased pollution as something one would take into consideration, and hence that the chairman probably did take into consideration, at least in the absence of any implication that he did not take it into consideration. In our version, the chairman's failure to mention the beneficial side-effect of improving safety, moreover, can be taken to have precisely this implication. For this beneficial side-effect is presented to him alongside the increase in profits, which are together opposed by the increase in pollution, but the chairman cites only the increase in profits in his decision. This might, but need not, be taken to imply that he has ignored or dismissed the side-effects completely, without considering their relative importance. So his declaration remains more evenly ambiguous than it would do in the absence of the beneficial side-effect.

Describing the processes underlying the participants' classifications of the side-effects in this way does not imply that the participants can or would defend their classifications in these terms. We intend only to be tracking the contours of the concept of an intentional outcome by showing which factors determine its application. This does not require that the participants go through a process of inferential reasoning in order to apply the concept. Indeed, it could not imply that, since no chain of inferential reasoning can begin without some concept or other already having been applied. Competence in the employment of a concept does not imply that one can accurately report the processes involved in that employment, therefore, so we cannot simply test CH directly by asking participants whether being taken into consideration in deliberation is the key factor in whether an outcome counts as intentional. The same holds for both EEH and AEH, which owe at least some of their notoriety to the fact that neither

normative valence nor blameworthiness are factors which people typically report as underlying their judgments of whether an effect is intentional.

This point is borne out, moreover, in the comments we gathered in our experiments. We allowed participants to add comments justifying their answer of ‘yes’ or ‘no’ to the question posed. Out of 232 participants, 157 offered comments, though 6 of those were judged not to be attempts to explain their preferred classification of the outcome. It is clear from the patterns in the classifications themselves that the comments that did attempt to explain the participants’ preferred classifications did not correctly report the reasons for those classifications. The rest of this section of the paper is concerned with showing this to be the case. In the following section, we will consider another example of this disparity between participants’ classifications and the reasons they give for them. We will go on to draw out substantial philosophical and methodological implications of this disparity. But one thing is immediately clear: the fact that very few comments advert explicitly to whether or not the agent took the outcome into consideration cannot be cited as evidence against our Consideration Hypothesis.

The comments were coded into the following five categories: ‘K’ means that the classification was explained by the agent’s knowing, being aware, foreseeing or having information that the outcome would occur; ‘SE’ means that the comment just pointed out the agent’s main aim or pointed out that the effect asked about was not the agent’s main aim or reason for action; ‘NC’ means that the comment explicitly referred to the effect asked about not being ‘considered’ or ‘taken into consideration’; ‘NE’ means that the comment did not offer an explanation of the classification; ‘OE’ means that the comment gives an explanation that is not otherwise covered in this coding. The comments were coded by Robin Scaife in consultation with Jonathan Webber. Reliability was assessed by a second coder, whose coding yielded a Cohen’s Kappa of 0.88. This level of agreement was considered sufficiently high for the codings to be used without modification.⁴ A breakdown of the results of these codings can be seen in Table 2.

Of the 64 participants who offered a comment to justify classifying an outcome as intentional, 85.9% claimed that the important fact was that the chairman knew that the outcome would occur. Of the 93 who offered a

⁴ Our first attempt at coding used ten categories. This coding was found to be unreliable, yielding a Cohen’s Kappa of 0.58. We attributed this to overcomplicated categorisation and ambiguities between the SE and NC categories. To improve reliability, we reduced the number of categories and amended the definitions of the SE and NC categories.

Table 2 Categorisation of justifications for classifications in our Chairman vignettes.

	K	SE	NC	NE	OE
Total (157)	55 (35%)	79 (50.3%)	6 (3.8%)	6 (3.8%)	11 (7%)
Intentional (64)	55 (85.9%)	2 (3.1%)	0 (0%)	3 (4.7%)	4 (6.3%)
Not Intentional (93)	0 (0%)	77 (82.8%)	6 (6.5 %)	3 (3.2%)	7 (7.5%)

comment to justify classifying a side-effect as not intentional, 82.8% claimed that the important fact was that the chairman only cared about, was only concerned with or only aimed at increasing profits. This is the pattern regardless of the normative valence of the side-effect classified. Six people explicitly replied that an outcome was not intentional because it was not taken into consideration. These comments cannot be taken as evidence of CH, however, not only because they represent a very small proportion of the comments received, but also because there were three comments relating the classification of an outcome to a moral judgment of the effect, the action overall, or the agent, which might seem to support EEH or AEH.

What matters for our purposes, moreover, is that the participants' comments do not report accurately the processes underlying their classifications. This can be shown by closer examination of our data. Of the 143 participants asked about a negative side-effect, the proportion that classified it as intentional and justified this classification with reference to the chairman's state of knowledge was 31.5%. However, of the 90 participants asked about a positive side-effect only 18.9% (with or without comment) classified it as intentional even though the vignettes were the same as those given to the participants asked about negative side-effects. Many of the participants who claimed that the negative outcome was intentional because the agent knew about it, therefore, would not have said the same had the outcome been positive. This need not be taken as evidence that participants' explanations of their classifications are nothing more than confabulations, however, since there are other ways in which the comments could fail to report accurately the process underlying the classifications, a point which we will consider in more detail in the next two sections of this paper.

Since the comments do not provide an accurate guide to the processes underlying the classifications, the fact that very few participants mentioned whether or not the agent took the side-effect into consideration cannot be cited as evidence against CH. The evidence in favour of CH, on the other

hand, is that it can explain the patterns in those classifications. We have seen that EEH cannot explain our data. But we have also seen that AEH can explain the data, so long as we allow that participants aggregate moral and practical valences and tend to classify risk as having negative normative valence. In the next section, we will see that a further experiment shows CH to be preferable to AEH.

6. Bad Side-Effects of Good Actions

We tested CH against AEH using a vignette in which an action as a whole has a positive normative valence but involves a negative side-effect which the agent clearly took into consideration. Since the side-effect is taken into consideration, CH predicts that participants will tend to classify it as intentional. The opposite prediction will be made by any form of AEH which claims that a necessary condition of the side-effect being classified as intentional is that the action overall has a negative normative valence, since in this case that purported necessary condition is not met. As we saw in section 3, however, AEH can take the weaker forms of claiming only that the side-effect is always or even just usually classified as intentional when it has a negative normative valence that renders the overall action similarly valenced. Such weaker forms of AEH make no prediction at all in experiments where the action overall is positive. But for that very reason, they cannot provide an explanation of the results of any such experiment, whereas CH can explain the result if its prediction is correct.

In order to test the prediction made by CH, we gave the following vignette to undergraduate Philosophy students at Cardiff University and at the University of Sheffield, and asked them whether the parent intentionally inflicted pain on the child:

The doctor said to the parent: ‘although your daughter is no longer showing any symptoms, we could run some tests to ensure that it won’t recur; but the tests are painful, so it’s up to you.’

After some consideration, the parent said: ‘the tests should be run, to be on the safe side.’ And so the tests were run.

Of the 83 responses we received, 66.3% replied that the parent had intentionally inflicted pain and the remaining 33.7% replied that the parent had not intentionally inflicted pain.⁵

⁵ In fact, we received 84 responses, but one was discounted for writing ‘don’t know’ rather than deciding Yes or No.

This difference in responses was statistically significant $\chi^2=1$, $N=83$) 8.78, $P<0.005$. This shows that participants were inclined to classify a negative side-effect of an overall positive action, where the agent had clearly taken that side-effect into consideration, as intentional. CH can explain this result, whereas weak forms of AEH cannot and strong forms of AEH predict the opposite result. CH therefore has greater explanatory power than any form of AEH.

We again allowed the participants to add comments explaining their classifications. As with our previous experiment, the comments given by the participants do not seem to report the relevant factors underlying those classifications. Most participants (80.1%) took the opportunity to write something. These responses were coded into the following five categories: 'K' means that the classification was explained by reference to the agent's knowledge or awareness of the outcome of their decision; 'SE' means that the comment pointed out the main aim of helping the child or explained that the pain was a side-effect of the main aim or intention; 'E' means that the classification was explained by pain being the effect of the agent's choice or decision; 'D' is the claim that the pain was inflicted by the doctor not the parent; 'NE' means that the comment does not give an explanation of the classification, but rather clarifies it or aims to exonerate the parent. The comments were coded by Robin Scaife in consultation with Jonathan Webber. Reliability was assessed by a second coder, whose coding yielded a Cohen's Kappa of 0.86. This level of agreement was considered sufficiently high for the codings to be used without modification. A breakdown of these codings can be seen in Table 3.

The table shows that the participants' comments do not capture their actual practice. 85.7% of the comments attached to classifications of the pain as not intentional made reference to it being a side-effect rather than the agent's central aim or intention, whereas 52.1% of the explanatory comments attached to classifications of the pain as intentional made reference

Table 3 Categorisation of justifications for classifications in our Parent vignette.

	K	SE	E	NE	D
Total (67)	24 (35.8%)	29 (29.8%)	6 (8.9%)	16 (23.8%)	1 (1.5%)
Intentional (46)	24 (52.1%)	2 (4.3%)	7 (15.2%)	13 (28.3%)	0 (0%)
Not Intentional (21)	0 (0%)	18 (85.7%)	0 (0%)	2 (9.5%)	1 (4.8%)

to the agent's knowledge or awareness that the pain would occur. Were these comments accurate reports of thought processes underlying the classifications, we should expect roughly the same proportions of participants classifying known side-effects of actions as intentional or as not intentional regardless of variations in the nature of those side-effects. But we know, from all of the data in this debate, that this expectation is not met.

One participant wrote that the pain was intentional because the parent 'decided that the pain was worth the gain'. This participant does seem to be saying that the classification as intentional is dependent on the agent having taken it into consideration. In light of this, the comments that seem to exonerate the agent rather than explain why the outcome is intentional can look as though they express the view that the agent did take the pain into consideration but thought it a price worth paying. These comments, which make up a large proportion of the comments in the NE category, might therefore be seen as imprecise expressions of the thought that the outcome is intentional because it was taken into consideration. Once such imprecision is allowed, moreover, we might equally see the comments adverting to the agent's knowledge as expressions of the thought that the outcome was taken into consideration. These comments make up the whole K category, and so account for 52.1% of the comments attached to the classification of the outcome as intentional. It seems that the majority of those who classified the pain as intentional made a comment that can be taken to imply that the agent took the pain into consideration.

We do not intend this somewhat speculative and imprecise point about the comments as evidence of CH. Rather, we merely note that such comments cannot be taken at face value as accurate explanations of the participants' classifications of the outcomes. They might be manifestations of a vague or inchoate awareness of the underlying reasons for those classifications or they might be wholly misleading. The classifications themselves, on the other hand, are those predicted by CH. They cannot be explained by any form of AEH since the action was not negatively evaluated by the participants. So the results of this experiment, in conjunction with the results of our previous experiment, show that an outcome of an action is classified as intentional only if it was taken into consideration in the deliberation leading to the action, where this means that it was assigned an importance relative to the other factors under consideration. This hypothesis can also explain all of the experiments reviewed in this paper. We consider it, moreover, to explain all of the experiments discussed in the recent debate over the folk concept of intentional action, though these are too numerous to

review here. These experiments, therefore, should not be taken, as they are usually taken, to show that the normative valence of a side-effect itself influences its classification as intentional or otherwise. Neither AEH nor EEH provides the best explanation of the data.

7. Research Implications

The experiments discussed in this paper, therefore, should be taken to have shown that a necessary condition for an effect to be classified as intentional is that the agent took it into consideration when deciding on the action. We have seen CH explain aspects of our experimental findings that either contradict or at least cannot be explained by its rival hypotheses. Since an agent cannot take an effect into consideration without knowing or believing that the effect is likely, CH also explains why such knowledge or belief is necessary for the effect to be classified as intentional. This necessity is demonstrated by people's responses to a story in which the manager of a company upset an employee by asking how she and her husband were: when the story stated that the manager was unaware that this employee's husband had just left her for another woman, no participants classified the manager's upsetting her as being intentional; whereas when the story made clear that the manager was well aware of this and knew his question would upset her, 64% of participants classified upsetting her as intentional (Nadelhoffer 2004b, 266–7).

On our view, this requirement of knowledge or belief is not independent of the necessary condition that explains the difference between the harm and help versions of Knobe's original experiment. It is already implicit within the condition that the effect has been taken into consideration. Thus, an advantage of CH is that it can explain why we do not classify unforeseen side-effects as intentional: if they are unforeseen, then they cannot have been taken into consideration. The view that the experiments show that negative normative valence is involved in our classification of side-effects as intentional, on the other hand, leaves this requirement entirely unexplained.

Our view can explain two further experimental findings seldom emphasised in this debate, moreover. One is that people are more likely to classify an effect as intentional when they think that the agent brought it about in order to achieve a goal. Knobe gave his original chairman vignettes to a sample of people and asked not whether the side-effect was intentional,

but how strongly it ‘sounds right’ to say that the chairman harmed (or helped) the environment in order to maximise profits. He found that harming the environment is strongly classified as done in order to maximise profits, but helping the environment is not (Knobe 2004, 183–4). The other pertinent finding is that normatively neutral side-effects can be classified as intentional. Nadelhoffer tested a vignette in which a sniper in position to carry out his orders ‘realizes that the gunfire will definitely cause the barrel of his gun to get hot’. But ‘the sniper doesn’t care at all whether the barrel of the gun is hot, he doesn’t have to touch it anyway’. So he fires the gun and the barrel heats up. Of those asked whether he heated up the barrel intentionally, 68% answered affirmatively (Nadelhoffer 2006b, 20).

CH can explain the link between an effect being classified as intentional and the agent seeing it as worth incurring in order to achieve a goal. CH can also explain why the side-effect of heating the gun barrel is classified as intentional despite not having a negative valence. In both cases, the effect is taken into consideration by the agent. Neither EEH nor AEH explain either of these findings. As well as explaining these, CH can explain the relation between an effect being foreseen and it being intentional and can explain all of the experimental findings discussed in this article. Unless some other hypothesis can be shown to have even greater explanatory power, we should accept that it has been demonstrated not that the classification of an effect as intentional depends on normative assessment of the action or effect, but rather that an effect is classified as intentional only if the agent is taken to have assigned that effect some relative importance in deciding what to do.

We have not attempted to argue for the further claim that having been taken into consideration (in our sense) is sufficient for an effect of an action to count as intentional. Such a claim would explain why the central aim of an action, what the agent primarily intended to do, always counts as intentional. We know that the converse is not the case: studies have shown that people classify some outcomes as *intentional* but not *intended* (Knobe 2004; McCann 2005). If having been taken into consideration is sufficient for an effect to be intentional, then this would explain why intended outcomes are always intentional: nothing can be intended without being assigned some importance relative to other outcomes. But this faces a serious experimental challenge: as we saw in section 3, experiments seem to show that classification of an outcome as intentional depends partly on the roles of luck and skill in bringing it about. Further consideration of these

experiments, however, might show that this challenge can be met. This would allow us to conclude that the classification of an effect of an action as having been taken into consideration (in the right way) is necessary and sufficient for, and therefore equivalent to, classification of it as intentional. But this research is for another time.

Such research, and indeed all research involving vignettes, should take account of the methodological lessons developed across this paper. The construction of vignettes and interpretation of the data they yield should be sensitive to the fact that the meaning of a sentence, or at least the way a reader is likely to understand it, is dependent on its context in such a way that simple substitution of one term for another does not necessarily preserve the other sentences of the narrative intact. As we saw in section 5, moreover, the probable disambiguation of such a sentence might depend not only on what is present in the context, but also on which aspects of the context are made salient by the experimental probe. The question might draw attention to part of the context that disambiguates the sentence in one way, where a different question about the very same vignette might disambiguate the sentence another way. We have also seen, finally, that these might be ineliminable features of the vignettes required to test some hypotheses. None of this should be taken as a criticism of this experimental method itself. Indeed, we have seen that these aspects of it can be turned to our advantage.

Finally, this emphasis on ambiguity might help to explain why participants make comments that do not directly and explicitly confirm CH, yet might be interpreted as inchoate expressions of their underlying thought processes rather than as simple confabulations. Perhaps the context-sensitivity of much of our everyday language makes it unlikely that a non-specialist experimental participant can accurately report the reasons for their response in a few words and in a very short space of time. It would not follow from this that they were unaware of the reasons behind their classifications, just that they find those reasons difficult to articulate. The difference between them and professional philosophers might not be a difference in conceptual competence or in understanding their own concepts, but a difference in their ability to clearly articulate that understanding. This is a hypothesis that can itself be tested in future empirical research. If it turns out to be correct, moreover, then careful experimental philosophy should be seen as an aid rather than a rival to traditional philosophical projects. Perhaps, that is to say, philosophers should follow Aristotle in seeking the truth that underlies both the teachings of the sages and the reflections of the folk.

Bibliography

- Adams, Fred and Annie Steadman. 2004a. Intentional Action in Ordinary Language: Core Concept or Pragmatic Understanding? *Analysis* 64: 173–81.
- . 2004b. Intentional Action and Moral Considerations: Still Pragmatic. *Analysis* 64: 268–276.
- Hindriks, Frank. 2008. Intentional Action and the Praise-Blame Asymmetry. *The Philosophical Quarterly* 58: 630–641.
- Knobe, Joshua. 2003a. Intentional Action and Side Effects in Ordinary Language. *Analysis* 63: 190–193.
- . 2003b. Intentional Action in Folk Psychology: An Experimental Investigation. *Philosophical Psychology* 16: 309–324.
- . 2004. Intention, Intentional Action and Moral Considerations. *Analysis* 64: 181–187.
- . 2006. The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology. *Philosophical Studies* 130: 203–231.
- . 2007. Reason Explanation in Folk Psychology. *Midwest Studies in Philosophy* 31: 90–106.
- Knobe, Joshua and Gabriel Mendlow. 2004. The Good, the Bad and the Blameworthy: Understanding the Role of Evaluative Reasoning in Folk Psychology. *Journal of Theoretical and Philosophical Psychology* 24: 252–258.
- McCann, Hugh. 2005. Intentional Action and Intending: Recent Empirical Studies. *Philosophical Psychology* 18: 737–748.
- Machery, Edouard. 2008. The Folk Concept of Intentional Action: Philosophical and Experimental Issues. *Mind and Language* 23: 165–189.
- Malle, Bertram and Joshua Knobe. 1997. The Folk Concept of Intentionality, *Journal of Experimental Social Psychology* 33: 101–121.
- Mallon, Ron. 2008. Knobe vs. Machery: Testing the Trade-Off Hypothesis. *Mind and Language* 23, no. 2: 247–255.
- Nadelhoffer, Thomas. 2004a. On Praise, Side Effects, and Folk Ascriptions of Intentionality. *Journal of Theoretical and Philosophical Psychology* 24: 196–213.
- . 2004b. Blame, Badness, and Intentional Action: A Reply to Knobe and Mendlow. *Journal of Theoretical and Philosophical Psychology* 24: 259–269.
- . 2006a. Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Jury Impartiality. *Philosophical Explorations* 9: 203–220.
- . 2006b. Desire, Foresight, Intentions, and Intentional Actions: Probing Folk Intuitions. *The Journal of Cognition and Culture* 6: 133–158.
- . 2007. Fringe Benefits, Side Effects, and Indifference: A Reply to Feltz. *Journal of Theoretical and Philosophical Psychology* 27: 801–809.
- Nichols, Shaun and Joseph Ulatowski. 2007. Intuitions and Individual Differences: The Knobe Effect Revisited. *Mind and Language* 22: 346–365.
- Nichols, Shaun and Joshua Knobe. 2008. Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions. In *Experimental Philosophy*, edited by Joshua Knobe and Shaun Nichols (Oxford: Oxford University Press): 105–126.
- Pettit, Dean and Joshua Knobe. 2009. The Pervasive Impact of Moral Judgment. *Mind and Language* 24: 586–604.