Intentional action and the semantics of gradable expressions
(On the Knobe Effect)

Paul Égré[*]

Abstract

This paper is about the Knobe Effect, an asymmetry identified by J. Knobe concerning attributions of intentional action in ordinary language. What the effect indicates is that judgments about intention and causation are influenced by moral considerations. The goal of this paper is to put the effect in a broader semantic perspective, with an aim to clarifying the workings of the adjective "intentional". We buttress and extend a hypothesis put forward by D. Pettit and J. Knobe ("The Pervasive Impact of Moral Judgment", *Mind and Language*, 24, (2009), 586-604) to account for the effect. According to Pettit and Knobe, one should look at the semantics of the adjective "intentional" on a par with that of other gradable adjectives such as "warm", "rich" or "expensive". What this analogy suggests is that the Knobe effect might be an instance of a much broader phenomenon, which concerns the context-dependence of normative standards relevant for the application of gradable expressions. The present contribution adduces further evidence for this view and goes on to examine the predictions one obtains by assuming that the adjective "intentional" involves a two-dimensional scale of comparison, which implies evaluating how much an action or outcome is *desired* on the one hand, and how much it can be *foreseen* as a consequence of one's actions on the other.

1 Introduction

In recent years the philosopher and moral psychologist Joshua Knobe has conducted a series of original and influential studies concerning people's judgments about intentional action. What Knobe has found is that confronted with scenarios with basically the same causal structure, subjects are more likely to judge an action intentional when the outcome is morally reprehensible or detrimental than when the outcome is morally good or desirable. This asymmetry in people's evaluation of intentional action, now commonly referred to as the "Knobe Effect", is currently an object of intense study and different interpretations of the phenomenon have been proposed.[1] According to Knobe, in particular, what the asymmetry suggests is that moral considerations play an essential role in the folk psychology of intentionality and agency.

In this paper I propose to look at Knobe's results from the point of view of the semantics of natural language. More specifically, my aim is to examine the semantics of the adjective "intentional" in the light of Knobe's findings. One motivation for that, worth emphasizing, is that Knobe's experiments belong as much to the field of psycholinguistics as to the domain of moral psychology, since for the most part Knobe's experiments provide subjects with the linguistic description of a particular scenario, and ask them to judge whether a given sentence is acceptable or not in the context of that scenario. In that sense, one can legitimately view Knobe's work as probing for people's truth-conditional intuitions concerning that part of the lexicon that deals with agency and intentional action.

More fundamentally, in a recent paper, Pettit and Knobe have outlined a new and convincing line of explanation for the phenomenon brought to light by Knobe. According to Pettit and Knobe, one should look at the semantics of the adjective "intentional" on a par with that of other gradable adjectives such as "warm", "rich" or "expensive". What Pettit and Knobe suggest, based on this analogy, is that from a semantic point of view the Knobe effect itself could be an instance of a much more general phenomenon which concerns the context-dependence of gradable adjectives and the sensitivity of the normative standards relevant for their application to comparison classes. As I see it, this hypothesis is not only plausible, but presently the most promising we have in order to account for the asymmetry found in people's judgments about intentional action. In this paper I therefore propose to adduce further evidence in favor of their hypothesis, and then to go on to discuss in greater detail the dimensions along which the context-dependence of the adjective "intentional" is typically understood.

In section 2 I give a presentation of the Knobe effect intended mostly for readers unfamiliar with the effect. The core of the paper is section 3, where Pettit and Knobe's main explanatory hypothesis is reviewed, and the connection with the semantics of gradable expressions examined more closely. I discuss further examples of expressions involving grading, in particular concerning judgments about probabilities and quantities, and argue that Knobe-type asymmetries are even more pervasive than suggested by Pettit and

---

[1]See in particular Adams and Steadman 2004, Nichols and Ulatowski 2007, and Machery 2008 for variations and concurrent explanations.

Knobe. The particular lexical semantics of the adjective "intentional" is discussed more specifically in section 4. I argue that judgments about "intentional" involve at least two dimensions of comparison, a dimension pertaining to desire, another pertaining to knowledge. This assumption allows us to account for a particular Knobe-type asymmetry proposed by Machery along the general lines of Pettit and Knobe.

2 The Knobe Effect

2.1 The "side-effect" effect

The contrast found by Knobe in people's judgment about intentional action has been evidenced in a variety of experiments and scenarios, both by Knobe and by colleagues. In order to be sufficiently precise on the phenomenon, and to qualify its scope, it will be useful to recall one of Knobe's paradigmatic cases first. In a series of studies (Knobe 2003b), Knobe presented the following two scenarios to two different groups of subjects:

> **Scenario A:** 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.

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

People confronted with Scenario A (the 'Harm condition') were then asked to answer the following question:

(1)   Did the chairman of the board intentionally harm the environment?

Similarly, people confronted with Scenario B (the 'Help condition') were asked to answer the following question:

(2)   Did the chairman of the board intentionally help the environment?

What Knobe found out is that 82% of subjects who got Scenario A answered positively to the question asked (18% negatively), while 77% of subjects who got Scenario B answered negatively to the corresponding question (23% positively), thereby giving a striking contrast between the two conditions. On top of that, Knobe asked subjects a second question,

namely to rank on a scale from 1 to 6 whether the chairman deserved praise or blame from what he did. The results obtained indicate that subjects are much more likely to ascribe blame in the harm condition (M=4.8) than they are likely to ascribe praise in the help condition (M=1.7). Finally, Knobe found judgments about praise or blame to be correlated to judgments about intentionality in each scenario.

What is so intriguing about Knobe's contrast is that while the chairman's *main goal* remains invariant from one condition to the other ("make as much profit as I can"), while he is *equally indifferent* regarding the side-consequences of his action at the time of his decision ("I don't care at all"), finally while he is *equally aware* of all the relevant potential consequences of his decision, the result of the chairman's action is widely judged intentional in the Harm condition, and widely judged not intentional in the Help condition. Because those two conditions vary essentially on the nature of the side-effect that is brought about, this contrast, which is often referred to as the Knobe effect, is also commonly called the "side-effect" effect.

Based in part on the correlation found between judgments of intentionality and judgments of praise and blame, what Knobe originally concluded from his data is that whether an action is judged intentional or not depends primarily on people's moral evaluation of the outcome of this action, rather than on the agent's goal or awareness of that goal.

2.2   Scope of the phenomenon

The asymmetry observed by Knobe raises a number of different issues. The first concerns the scope and robustness of the phenomenon. This issue divides into two: one is whether and how much the consideration of side-effects is essential to the contrast observed between judgments. This problem, clearly, has to do with the structure of the scenario. The other is whether and how much the use of the adverb "intentionally", or more fundamentally of the adjective "intentional", is essential to the contrast observed. This second problem concerns the nature of the question asked.

Both of these problems have been examined, whether directly or indirectly, by Knobe and by others. Regarding the first problem, one thing that should be pointed out is that the term "Knobe effect" has come to be used as a near-synonym for "side-effect effect", or conversely. However, earlier experiments by Knobe, which originally involved the consideration of *deviant causal chains*, and no side-effects at all, seemed to exemplify just the same contrast. In one of these studies (see Knobe 2003a, and also Nadelhoffer 2005 for further supporting data), Knobe presented the following two scenarios to different groups of subjects:

> **Scenario 1:** Jake desperately wants to win the rifle contest. He knows that he will only win the contest if he hits the bulls-eye. He raises the rifle, gets the bull's-eye in the sights, and presses the trigger. 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.

Scenario 2: 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. 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.

People confronted with Scenario 1 were then asked:

(3)   Did Jake intentionally hit the bull's eye?

People confronted with Scenario 2 were asked:

(4)   Did Jake intentionally kill his aunt?

76% percent of the subjects in Scenario 2 answered positively to the second question, while only 28% of the subjects in Scenario 1 answered positively to the first question. Prima facie, and disregarding here the controls used by Knobe in his experiment,[2] what appears is that subjects are more likely to judge the second action intentional because killing is morally reprehensible, and less likely to judge the first action intentional because hitting the bull's eye is morally neutral, or even is a matter of praise, rather than blame.

One observation we can make on this pair of scenarios, however, is that they involve more parameter variations than the scenarios about side-effects. In particular, the questions asked, arguably, do not quite constitue a minimal pair, since one might have expected either the first question to be: "did Jake intentionally win the rifle contest?" (keeping the second question fixed, and focussing on the main goal of the shooting), or rather the second to be: "did Jake intentionally hit his aunt in the heart?" (keeping the first question fixed, and assuming a corresponding change in the description of the scenario: "he gets his heart in the sights"). It seems plausible, in particular, that the percentage of positive answers should decrease in the latter case, because "hitting in the heart" is a more fine-grained event than simply "killing". These further predictions have been tested by B. Malle (see Malle 2006), who compared different choices of the verb in the aunt scenario, and found that indeed significantly more subjects assented to "did Jake kill his aunt intentionally?" than to "did Jake hit his aunt's heart intentionally?" (namely 100% against 49% in a between-subject design). Despite this, it remains quite possible that even taking into account such modifications, people would more readily assent to "Jake intentionally hit his aunt in the heart" than to "Jake intentionally hit the bull's eye", because the first action is considered morally reprehensible.[3] We shall say more about this below, but this is enough

---

[2]Knobe essentially crossed the immoral/non-immoral distinction with a luck/skill condition, namely compared the two cases we considered, in which Jake's hitting the target is lucky, to a situation in which "Jake is an expert marksman", whose "hands are steady", with the shot "aimed perfectly". See Knobe (2003a) for details.

[3]Malle's report does not allow us to conclude on that, as far as I can see. Malle's experiment compared subjects' answers depending on the verb, and the part of the body that was hit, and contrasts the luck and skill conditions in the aunt case, namely for the immoral case. But Malle does not say whether he

for us to conclude that consideration of side-effects is most likely not an essential feature of the phenomenon examined by Knobe.[4]

Turning now to the second issue, the question concerns how much the contrast observed by Knobe, whether in the side-effect scenarios or in the deviant causal chains scenarios, has to do with the particular lexical semantics of the adverb "intentionally", or even with the fact that it is an adverb rather than any other construction. This question has received close attention from Knobe himself. In a recent paper, Pettit and Knobe survey a number of variations on the side-effect scenario in which the question asked was changed, and other words or constructions used to qualify the main agent's attitude. They point out, in particular, a study by Tannenbaum, Ditto and Pizzaro (2007) in which the question asked is: "did the chairman have a desire to help [harm] the environment?", for which the same contrast was observed. Knobe himself compared different formulations of the question "did the chairman intentionally help [harm] the environment?". Among the questions he asked was: "was it the chairman's intention to help [harm] the environment?". Interestingly, the percentage of positive answers dramatically decreased in comparison to the questions "did the chairman intentionally help [harm] the environment?", but just the same contrast was observed, namely significantly more positive answers were given for the harm condition than for the help condition. Analogous scenarios involving target expressions such as "be opposed", "be in favor of", "advocated" exemplified the same contrast. Based on these facts, Knobe and Pettit's conclusion is that:

> "the surprising results obtained for intuitions about intentional action do not really have anything to do with the distinctive features of the concept of intentional action in particular. Rather there is a perfectly general process whereby moral judgments serve as input to folk psychology, and the effects observed for intentional action should be understood as just one manifestation of this broader phenomenon."

The upshot of these considerations should be that what the "Knobe effect" refers to is not primarily an asymmetry having to do with side-effects, nor essentially an asymmetry having to with the notion of intentional action, but a broader phenomenon, which also concerns ascriptions of desire, wishes, and which involves a contrast between our attitudes pro or contra an outcome.

2.3   Discussion

Pettit and Knobe's qualification of the phenomenon is quite convincing, but it raises two further issues. The first is whether Pettit and Knobe's conclusion could be even further generalized. In their paper, Pettit and Knobe invoke the pervasiveness of moral considerations, but they essentially consider scenarios involving predicates that serve to qualify a

---

replicated the experiment for the rifle contest, namely in the non-immoral case.

[4]Note that this is not an objection to talking about "side-effect effects" in order to refer to Knobe's paradigmatic case. This is merely an objection to limiting the definition of the asymmetries investigated by Knobe to that range of scenarios.

particular agent's attitude toward an action or outcome (viz. *intending, opposing, desiring, being in favor of*, and so on).[5] However we may wonder if the asymmetry observed by Pettit and Knobe is limited to a particular repertoire of attitudes such as these. Could the Knobe asymmetry manifest itself in scenarios that do not explicitly refer to the description of a specific agent's attitude, desire or intention?

The second issue, which relates to the previous one, concerns the scope of what Pettit and Knobe call "moral judgments". In his first studies on the topic, Knobe originally suggested that the asymmetry found in people's judgments may have to do with the particular level of *praise* or *blame* attached to an action (see Knobe 2003). In their recent account however, Pettit and Knobe take "moral judgments" in a broader sense. They suggest that "moral judgments" essentially involve whether an outcome is judged favorably or unfavorably by people. This suggests that in some cases, Knobe's asymmetry may simply attach to whether a particular outcome is judged desirable or undesirable, irrespective of whether it intrinsically involves blame or praise. Following this line, the question can be made even more radical: is Knobe's asymmetry specific to matters which concern the desirability of an outcome, or is the asymmetry not even specific to moral issues in that broad sense?

Pettit and Knobe's account directly predicts that the answer to the first question should be positive, and indeed in the next section we will find other examples suggesting that Knobe's asymmetry extends to an even wider class of lexical items than those they consider in their paper. For example, as stressed by Knobe, moral considerations appear to influence not only our judgments about intention, but our judgments about plain causation (judgments of truth and falsity for sentences of the form "agent *A* caused event *e*").[6]

The answer to the second question, on other hand, is not as straightforward (though a positive answer to it would obviously entail a positive answer to the first question). In their paper, however, Pettit and Knobe formulate a central hypothesis, namely that the asymmetry found in judgments about intentional action and kindred concepts is exactly of the same kind as asymmetries found in judgments concerning the application of gradable adjectives such as "warm" or "expensive". Because these adjectives do not obviously involve a moral component, we may wonder if the Knobe effect says something specific about moral considerations. In the next section, we address these two issues in greater detail. In order to do that, we first need to look at Pettit and Knobe's analogy with gradable predicates.

---

[5]Note that these are taken to be moral attitudes by Pettit and Knobe, not causal actions. See however Wolff (this volume) for the abstract representation of the underlying concepts.

[6]See Alicke (1992) and Alicke (2000) for evidence on this phenomenon, and the further results and discussion in Knobe and Fraser (2008) and Knobe (forthcoming). A separate analysis of those data would be highly relevant in the context of this volume, but would detract us from our main target, which is the analysis of judgments about intentional action proper.

3 Context-dependence and gradability

How do moral considerations come to interfere with ascriptions of intention to particular events? The answer examined in this section, originally proposed by Pettit and Knobe, is that normative considerations can contextually influence the perception of causal and intentional processes that are identical in all other respects.[7] I introduce further evidence for this hypothesis and use their analogy with gradable expressions to show how Knobe-type asymmetries can be semantically represented.

3.1 Pettit and Knobe's analogy

In their paper, Pettit and Knobe propose an illuminating analogy in order to account for the Knobe effect. The analogy concerns the attribution of gradable predicates such as "warm", "cold", or "hot" and their semantics. They point out that two liquids, such as beer and coffee, can be at exactly the same temperature, but this temperature be such that one is judged cold and the other not cold. Consider, for instance, a bottle of beer at 20°C, and a cup of coffee also at 20°C. Then the following two sentences would easily get opposite verdicts in truth-value. (5) will typically be judged false, while (6) will be judged true:

(5)   The beer is cold.

(6)   The coffee is cold.

The reason for this contrast, as Pettit and Knobe explain, is that:

> "people are rating each liquid relative to a *default* that specifies what it is supposed to be like. Coffee is supposed to be at a higher temperature, beer at a lower temperature. Hence, when both are at room temperature, the coffee falls below the default (and is classified as "cold"), while the beer falls above the default (and is therefore classified as "warm")."

Another way to put it is to say that beer and coffee set distinct standards of comparison for adjectives such as "warm" or "cold" (see Kennedy 1999, 2007). More precisely, we may say that beer and coffee constitute distinct *comparison classes* for the applicability of these predicates. Within each of these comparison classes, the *threshold* for the applicability of "warm" and "cold" is placed differently on the temperature scale.[8] For instance, in order

---

[7]For the sake of clarity, note that normativity in that sense differs from what for instance Thomason (this volume) calls *normalcy* about causation. Normalcy, as I understand it, is essentially a *causal expectation* of regularity concerning the mechanism of an event. Normativity qualifies a *moral expectation* regarding alternative ways in which an event could have happened. Below I discuss the sense in which normative expectations could relate to statistical regularities, but the basic intuition is that moral expectations are more fundamental for the kinds of asymmetries under discussion.

[8]Often, comparison classes are referred to by means of *for*-Prepositional Phrases, as in "Mary is tall for a fifth-grader". Note that we can say in the same way: "this is cold for coffee", but "this is not cold for beer". *For*-PP are not the only way to make comparison classes explicit, however. In (5) and (6), the nouns "beer" and "coffee" in subject position play the same role.

for beer to count as cold, its temperature would typically have to be less than a threshold of around 4°C, while in order for coffee to count as cold, it is sufficient for its temperature to be less than around 50°C.

Based on this analogy, Pettit and Knobe point out that a similar explanation can be put forward for the contrast found in Knobe's side-effect scenario and others of the same kind. Let us consider the pair:

(7)   The harm to the environment caused by the chairman's decision was intentional.

(8)   The help to the environment caused by the chairman's decision was intentional.

In this case, harm and help can be viewed as representing distinct comparison classes for the attribution of the adjective "intentional", in the same way in which beer and coffee represent distinct comparison classes for the attribution of the adjective "cold".[9]

One thing about adjectives such as "cold" and "warm", however, is that the scale is obvious and concerns temperature. The structure of the underlying scale for the adjective "intentional" is certainly more complex, a point to which we shall return below. In their paper, Pettit and Knobe assume that the scale for attributions of intention, desire, and related attitudes, is a scale that concerns the valence of the attitude the agent is supposed to have toward the outcome of her action, namely from con to pro, with various degrees. Thus, assuming a one-dimensional scale of the same kind we have for temperature, what matters is that, the threshold for "intentional" in the case of harm is much lower on that scale than the threshold for help. In other words, keeping identical the chairman's main motivations, awareness of the consequences, and indifference to side-effects, but changing the status of the outcome from harm to help, the agent's pro-attitude toward the harm outcome is high enough on the relevant scale to count as intentional, while the agent's pro-attitude toward the help outcome is not high enough to count as intentional.

Pettit and Knobe's explanation is quite convincing in our view. Before trying to refine it, I propose to adduce further evidence in support of their main hypothesis, on the basis of two further examples. These two examples, drawn from the area of probability and quantity evaluations, also involve grading in an essential way. Besides, they suggest that the Knobe asymmetry is even more pervasive than might have been thought at first.

3.2   Further contrasts: probability and quantity judgments

Judgments on probabilities.   The first example I wish to consider is due to Stefania Pighin and concerns people's evaluation of risk and probability. In a recent study, S. Pighin, J-F. Bonnefon and L. Savadori measured the evaluation of specific probabilities in a population of pregnant women.[10] They presented participants with a scenario in which a

---

[9]As pointed out by N. Hansen, a more natural formulation for (8) might be: "the benefit to the environment caused by the chairman's decision". I keep "help" for the paraphrase, to match Knobe's initial sentence.

[10]This study, as far as I can tell, was run without any acquaintance of Knobe's results, and with a practical application in mind, namely improving risk communication in medical diagnosis. This makes the

30 year-old pregnant woman named Elisa is communicated the numerical probability that her child might have various diseases. Their first experiment, based on a 2 by 2 between-subject design, was one in which the gynecologist says to Elisa: "there is a risk of [1 in 28/1 in 307] that your child will be affected by [Down's syndrome / insomnia]". Subjects then were asked to complete the following sentence: "In your opinion, the probability of [1 in 28; 1 in 307] that her child will be affected by [Down syndrome, insomnia] is:...", by selecting a response on a 7-point scale, ranging from "extremely low" to "extremely high".

What Pighin and colleagues found is that subjects mostly evaluate the probability of 1/28 as higher on that scale than 1/307 within each disease, in agreement with the numerical ordering of these numbers. However, they evaluated the same probability of 1/28 [1/307] to be higher in the case of Down's syndrome than the same probability in the case of insomnia. Even more spectacularly, on average subjects evaluated the probability of 1/307 as higher when pertaining to Down's syndrome than the probability 1/28 when pertaining to insomnia. Independently, Pighin and colleagues tested for people's judgment concerning the severity of each disease, and found a massive consensus on people judging Down's syndrome to be severer than insomnia. What they conclude from their experiment is that:

> "Coherently with the literature about subjective interpretation of probability words and graphical representations, probability evaluations of the same numerical risk level were greater when associated with a severe outcome (Down syndrome) than when associated with a not severe outcome (insomnia)".

These results obviously bear a direct and striking analogy with Knobe's findings. Indeed, what Pighin's data suggest is that the same probability value is judged "higher" when associated to an outcome that is less desirable (what Pighin and colleagues call a "severity bias", following Bonnefon and Villejoubert 2006, and Weber and Hilton 1990). By way of consequence, if subjects had been asked to answer by yes or no to specific sentences, rather than to give graded judgments on a scale, presumably the same contrast would have been observed for the following two sentences:

(9)   A probability of 1 in 28 for a child to have Down's syndrome is high.

(10)   A probability of 1 in 28 for a child to have insomnia is high.

Consistently with the explanation put forward by Pettit and Knobe, what this suggests is that the probabilistic threshold for what counts as "high" is not set on the same value for all diseases. This, however, leaves open the exact interpretation of "high". One hypothesis is that people judge whether the probability of some event is high or not in relation to the utility of that outcome. In other words, people would link probabilities to risk proper.[11] Clearly, the utility of one's child having Down's syndrome is negative and much more

---

results all the more interesting.

[11]This hypothesis was suggested by N. Bonini during questions to S. Pighin. See in particular Weber and Hilton (1990), who talk of "worry effect", for an account broadly compatible with this view.

so than the utility of one's child getting insomnia. Assuming that the utility of one's child not getting Down Syndrome is just the same as the utility of one's child not getting insomnia, then this entails that a lower probability of getting Down's syndrome than of getting insomnia is sufficient to produce the same expected utility.

As pointed out by S. Yalcin (p.c.), however, an explanation based merely on utility considerations will be inadequate if it does not take into account symmetric cases involving positive utilities. The examples Yalcin puts forward are variants of Pighin et al.'s scenario involving a fair coin flip. In Case 1, Heads you get a dollar, Tails you get nothing. In Case 2, Heads I burn down your house, Tails you get nothing. In Case 3, Heads you get 10 million dollars, Tails you get nothing. Yalcin's judgment, which is also mine, is that the chance that I get a dollar in case 1 would tend to be judged "average" on a scale ranging from "very low" to "very high", while the chances that my house gets burnt down in case 2, like the chance that I get 10 million dollars in case 3, would both be judged "very high", even though the utilities are exactly opposite. Because of that, a more adequate way of phrasing the generalization behind the idea of "severity bias" might be to say that for two events X and Y that have the same numerical probabilities of occurring, the probability of X is judged higher than the probability of Y in proportion to how significantly more positive, or significantly more negative the utility of X is relative to that of Y.

A forceful objection Yalcin makes to this line of explanation, however, is that two events of equal numerical probabilities, and of roughly neutral utilities, could still lead to different qualitative estimations. The case imagined by Yalcin to make this point is again a fair coin flip, such that Heads, all of your Department's members must wear yellow pants on the same day, while Tails, nothing happens. The chances that my Department's members wear yellow pants may be ranked higher in this case than even the chances that I get a dollar in Case 1, despite the fact that the utility of that outcome appears to be null. In other words, what really matters is that we expect the *ordinary* probability of an event such as the event that all members of a Department will wear yellow pants on the same day to be significantly lower than a half. More generally, our estimations of probabilities will fundamentally vary depending on what we expect to be normal for the kind of event under consideration. In what follows, this element will really be the nerve of our explanation. Importantly, however, this general line of explanation remains compatible with the idea that the more negative the utility of an uncertain and contingent outcome turns out to be, the lower we might normatively expect the probability of that outcome to be.

Before saying more about this case, another aspect worth pointing out concerning Pighin's data on probability judgments is that they do not *directly* pertain to agency, desire or intentions. In other words, whether a probability is judged as high or not does not obviously have to do with a particular agent's attitude in the scenario given. Because of that, the example should convince us that if moral considerations are indeed pervasive, as Pettit and Knobe put it, they are so in the broad sense of "moral" that we identified, namely in relation to what subjects who assess the sentence judge desirable or undesirable. Still, we emphasize that this example does not directly pertain to agency because, once again, subjects most likely understand "high" in the case of probability to mean: "higher than is desirable, given the severity of the outcome". In a number of cases, subjects

may consider the probability of some outcome to be "high" also because they consider that one could easily act in ways that would make it lower, and that would reduce risk. For instance, one may get indignant about the particular probability of dying in a car accident in a given country, based on the consideration that if road regulations were only slightly different from what they are, the frequency of accidents could be much lower. This hypothesis is compatible with Pettit and Knobe's explanation in terms of pro-attitude and con-attitude, in the sense that one would judge high a frequency of accidents which one estimates that responsible agents should strive to reduce to zero.

Judgments about "many".   The second example I wish to consider concerns the semantics of the determiner "many". Among quantifier words, "many", just like "few", is notoriously vague and context-dependent (see Sapir 1944, Partee 1990). Sapir (1944: 94) gives the following significant description and examples concerning our understanding of "many":

> *Many* merely means any number, definite or indefinite, which is *more than* some other number taken as point of departure. This point of departure obviously varies enormously according to context. For one observing the stars on a clear night thirty may be but "few," for a proof-reader correcting mistakes on a page of galley the same number may be not only "many" but "very many." Five pounds of meat may be embarrassingly "much" for a family of two but less than "little" from the standpoint of one ordering provisions for a regiment.

Moreover, as well as for the evaluation of probabilities to count as high, the evaluation of specific quantities to count as many is sensitive not only to the comparison class at hand, but to the expectations one might have in a given context (see Lappin 2000).

Let us imagine the following scenario: a boat just sank in the Atlantic Ocean with 80 passengers on board. 40 passengers perished, and 40 passengers survived. Now consider the following pair of sentences:

(11)   Many passengers perished in the sinking.

(12)   Many passengers survived the sinking.

Would the two sentences give rise to the same judgments? A natural prediction is that a Knobe-type contrast would easily arise for such pairs. In other words, it seems quite plausible that people would more readily judge the first sentence true than the second true in this particular context.[12]

This prediction, of course, calls for some qualifications. First of all, attributions of the form "Many As are Bs" are sensitive not only to the absolute number of As, but to the ratio

---

[12] Soon after I formulated this hypothesis, F. Cova and I tested for people's intuitions based on such pairs involving "many". The data obtained confirm the prediction, namely they show a significant contrast between judgments about the "perish" case and the "survive" case. The results of that study will be presented in a separate paper.

of As to Bs (see Partee 1989). For instance, if we now imagine that 1000 passengers were on board, 500 perished, and 500 survived, it may turn out that *ceteris paribus* people assent more or less equally to "Many passengers survived" and to "Many passengers perished", because 500 is in absolute value a large number, even as the mathematical proportion of 500 to 1000 remains the same as 40 to 80, namely 1 in 2. Conversely, if there were only 2 passengers, 1 perished, while the other survived, we might most likely disagree with both (11) and (12), simply because the numbers are too small in absolute value to simply count as "many".

In any case, the guess remains that for a range of appropriate situations and numerical values, one should find the same kind of contrast as the one found by Knobe and others regarding side-effects. To make the example even more dramatic, let us imagine that a house burnt down, in which it has been reported that there were six children: three of them perished in the fire, while the other three managed to escape. In such a case, I believe I would readily assent to:

(13)   Many children perished in the fire.

and much less readily accept:

(14)   Many children escaped the fire.

The point about these cases is that in both examples we put forward, the absolute number of individuals who perished is equal to the absolute number of people who survived, and likewise the proportion is the same. What essentially differs between the two sentences concerns only the nature and severity of the outcome. Presumably, in order to consider that "many people survived" from a dramatic event, one should typically be in a position to say that "few people perished". The number or proportion that fixes the threshold for "many", whatever its value in a given context, is thus typically set higher when the outcome is negatively valued than when the outcome is positively valued. Thus, I consider that 3 children dying out of 6 is many because I would have preferred or expected *fewer* children to die. Conversely, I consider that 3 children escaping out of 6 is not many because I would have expected or preferred *more* children to escape.[13]

As in the case of probability judgments, the expectation of what counts as "many" is relative not simply to the absolute or relative number of things that are under consideration, but to the utilities attached to the outcomes. Suppose that that the subjective value of a one child dying is negative and weighs more in absolute value than the subjective value of

---

[13]Compare with Sapir's early remarks on grading and affect: "grading and affect are intertwined, or to put it differently, *more than* and *less than* tend to have both an objective grading value and a subjective grading value dependent on a desired or undesired increase or decrease" (1944: 109). More recently, Geurts (2009) and Nouwen (2011) have discussed data bearing a connection to the same idea, in relation to *framing effects*. Geurts notes that in a Tversky-Kahneman type of scenario in which 200 people survived out of 600, and the other 400 died, one can say: "it is good that 200 people survived", but not "it is good that 400 people died", though the propositions embedded are truth-conditionally equivalent in context (namely the propositions that would be expressed by "200 out of 600 people survived" and "400 out of 600 people died"). Nouwen (2011) discusses related contrasts involving the quantifiers "a few" vs. "few".

one child getting rescued, which is positive. Suppose moreover that the value of n children dying (surviving) is computed as n times the value of one child dying (surviving). Then, the negative value of n children dying will simply outweigh the positive value of n children not dying.

Similarly to what we said about Pighin et al.'s experiment, however, an explanation of the data based purely on utility considerations would be insufficient to account for the full range of judgments one may find concerning "many". In particular, supposing more contextual information to be given, such that the fire was so fierce that anyone escaping it would be deemed miraculous, we can expect (14) to get more assent, and (13) less assent, and the effect we found to even be reversed.[14] More generally, in each of the examples we reviewed, the context can be so manipulated that considerations about the objective chances of an event may prevail over considerations about the negative vs. positive utility of that event. Again, this objection to an account based purely on utility considerations is correct. Nevertheless, I consider that the force of the examples we reviewed is that, precisely because so little is said explicitly in the scenario about risk and chances, the asymmetry in judgments appears to depend essentially what subjects consider as a *desirable* outcome.[15] In other words, normative expectations certainly depend on *normality* in the sense of the expected chance of some actual event, but what matters more from our present perspective is that they also depend on the *desirability* of some alternative outcome.

The idea that our judgments about quantities is relative to utilities and values is not new. As Fara points out:

> "Sometimes 'a lot' can mean 'significantly more than is typical'. But other times it can mean 'significantly more than is wanted or needed'. And also it can mean 'significantly more than is expected'."

Quite generally, as argued by Fara, the evaluation of what counts as "a lot" or as "many" is therefore relative in large part to our practical interests depending on the kinds of things or outcomes under consideration. That asymmetries can be found in our judgments about whether the same absolute or relative quantities count as "many" is predicted on such a view. Precisely because of that, however, the "pervasiveness of moral judgments" invoked by Pettit and Knobe may be seen as a particular case of the "pervasiveness of practical interests" in all of our graded judgments about a property or quality.

3.3   Standards of comparison and normative judgments

The example of "many" indicates that what differs between its applicability to the verb "perish" and its applicability to the verb "survive" cannot be purely extensional. As

---

[14]I am indebted to M. Kneer, E. Machery and an anonymous referee for calling my attention to this important point.

[15]Interestingly, Pighin et al. (2009) includes a second study in which the gynecologist communicates that each of the probabilities communicated is "above average". Their data indicate that subjects become sensitive to this reference point, but even so, the severity bias persists, that is the lower probability of the more severe disease is still judged to be higher than the greater probability of the less severe disease.

argued by Lappin, the evaluation of "many" depends on a "set of normative situations" (Lappin 2000: 602). When we say: "Many As are Bs", we mean that the number of As that are Bs is above some standard, namely is greater than the expected number of As that are Bs, where this number depends on the nature both of As and Bs in relation to our interests. Likewise, when we say that the probability of an event is "high", we mean that this probability is higher than some expected threshold *for the kind of event under consideration.*

Judgments about "many" and about how "high" a probability is are exactly in line with the paradigmatic case of "cold": to say that a liquid is cold means that it is colder than the expected temperature for that kind of liquid. By way of consequence, the two cases I put forward fall in place with Pettit and Knobe's analysis of "intentional": to say that the outcome of an action is intentional means that the degree of intention attached to the action is greater than the expected degree for the kind of action in question.

In order to make fully clear the parallel between these various cases, it will be useful to regiment them within the appropriate semantic framework. Let us focus on the adjectives "cold", "high", and "intentional" first. All adjectives can be viewed as mapping individuals to degrees on the appropriate scale. Building on the analyses of Bartsch & Vennemann 1972, Fara 2000 and Kennedy 2007, we may say that "x is cold" is true in a context if $\mathbf{cold}(x) \succ \mathbf{norm}(A)(cold)$, namely if the degree to which x is cold is (significantly) greater than a normative degree dependent on the context, relative to a given comparison class A. Here, $\mathbf{norm}(A)(cold)$ denotes a function specifying the minimum degree to count as cold within the class of things A, while $\mathbf{cold}(x)$ denotes the effective degree of coldness ascribed to x on the scale.[16] In this case, the comparison class is given by the category of the subject of the sentence, but it need not be so in general.[17] Basically, "this coffee is cold" is true if $\mathbf{cold}(c) \succ \mathbf{norm}(coffee)(cold)$, where c stands for the demonstrative "this coffee".

Similarly, in the case of Pighin's probability judgments, we may represent the target sentences as follows, where p stands for "The probability 1/307 to have Down's syndrome" and p' for "the probability 1/28 to have insomnia":

(15)   a.   The probability 1/307 to have Down's syndrome is high.
            b. $\mathbf{high}(p) \succ \mathbf{norm}(DownSyndrom)(high)$

(16)   a.   The probability 1/28 to have insomnia is not high.
            b. $\mathbf{high}(p') \preceq \mathbf{norm}(Insomnia)(high)$

Following Pettit and Knobe, the same analysis can be given for the harm vs. help case. Let h [h'] stand for "the harm [help] to the environment caused by the chairman's

---

[16]I follow Kennedy's 2007 convention here, and let $\mathbf{cold}$ represent the denotation of the adjective *cold* in the context of evaluation.

[17]See Kennedy 2007: ex. 24 sqq. for discussion of this point. Note that the analysis we are proposing here can be seen as a compromise between Bartsch and Venneman's original analysis and Kennedy's: like Kennedy 2007, we do not suppose that $\mathbf{norm}$ will always select the average within a given class, but we preserve the idea that $\mathbf{norm}$ has an argument for a comparison class.

decision":

(17) a. The harm to the environment caused by the chairman's decision is intentional.
   b. $\mathbf{intentional}(h) \succ \mathbf{norm}(Harm)(intentional)$

(18) a. The help to the environment caused by the chairman's decision is not intentional.
   b. $\mathbf{intentional}(h') \preceq \mathbf{norm}(Help)(intentional)$

As the semantic representations make clear, the joint truth of both (17) and (18) is semantically compatible with $\mathbf{intentional}(h)$ and $\mathbf{intentional}(h')$ being equal, so long as the thresholds $\mathbf{norm}(Harm)(intentional)$ and $\mathbf{norm}(Help)(intentional)$ are not identical themselves.[18]

Likewise, on a simplified analysis, "many As are Bs" can be taken to mean that $|A \cap B| > \mathbf{norm}(A, B)$, namely the cardinal of individuals that are both A and B is above some context-dependent normative standard that depends on As and Bs taken as intensional parameters.[19] Let us imagine that in the given context: $\mathbf{norm}(child, die) = 2$, and $\mathbf{norm}(child, not\ die) = 4$. Then one can consistently have $|child \cap die| = |child \cap \overline{die}| = 3$, so that "many children died" will count as true, and "many children did not die" will count as false.

An issue that remains open in all these cases, however, concerns the determination of the norm. Obviously, the position of the norm will vary between individuals. For instance, someone may consider that below 4 out of 8 people, not many people perished, but someone else may consider that if already 2 people perished out of 8, already many people perished. Such interpersonal variations are expected on any reasonable account, however, due to the vagueness of the expressions involved. What matters, for example, is that irrespective of how people draw the line individually in their judgments about "many", they should tend to leave a significant gap between what counts as "many" relative to "die" and the same word "many" relative to "not die", that is they should set the norm lower for "die" than for "not die".

The second issue we must consider concerns the number of contextual parameters we need to postulate for the analysis to come out right. In our formalization of (17) and (18), we introduced the expressions $Harm$ and $Help$ as semantic parameters, but we left unspecified the fact that these conditions concern environment-related harm and environment-related help. Obviously, we may find even more variations depending on whether one is talking of harm caused to the environment, to an animal, to a human adult, or indeed to a criminal, and so on.

We grant this point. Obviously, the normative standards for what counts as high,

---

[18]C. Piñon asks whether h and h' represent individual actions or actions types. As I see it, the harm or help caused to the environment is the particular result of the individual action taken by the chairman to carry out out a specific program. As such, I consider that h and h' refer to individual actions. However, the comparison class arguments $Help$ and $Harm$ contribute the relevant action types.

[19]More precisely, following Lappin, we may say that "many As are Bs" is true in a context w iff $|A_w \cap B_w|$ is greater than $|A_{w'} \cap B_{w'}|$ for every world w' that is normatively preferred to w.

intentional, or cold are likely to vary very finely depending on the nature of the predicates that these expressions qualify, and on additional features of the context. What matters for the explanation in this regard is first and foremost the idea that the applicability of each of the vague expressions we discussed depends on a standard of comparison that varies with the kind of action, event or outcome qualified by that expression.

4   The semantics of "intentional"

The purpose of the foregoing section was to buttress the analogy proposed by Pettit and Knobe for their account of the Knobe asymmetries. What we saw is that the Knobe asymmetries can be viewed as a particular case of a much broader phenomenon of context-dependence, of the kind we find in judgments involving grading more generally.[20] In extending this analogy, however, we have mostly disregarded the details of the lexical semantics of the adjective "intentional". In this section, our aim will be to look more specifically at the semantics of that adjective, in relation to some of Knobe's examples. We close the section with a discussion of a variant on the Knobe effect due to Machery (2008).

4.1   Desiring and foreseeing

As emphasized in the previous section, it is one thing to envisage that an action can be more or less intentional, and thus that "intentional" is susceptible to degrees, and another thing to specify the structure of the associated scale. Importantly, like most expressions susceptible of grading, an adjective like "intentional" may be multidimensional. Consider the adjective "bald". As is well known, our use of the word "bald" is not exclusively dependent on the number of hairs on someone's head, but it depends on the distribution of hairs, on their density on the scalp, and so on. Most likely, when we judge whether someone is bald, we rely on a measure that integrates these various aspects. The same may hold in the case of "intentional". In order to judge whether an action is intentional, we certainly rely on some salient dimensions, but these dimensions may be numerous.

   The first thing to note about the adjective "intentional" is that it attaches as much to an action as to the outcome of that action. This is clearer when we consider the adverb

---

[20]Note that, like Sapir, we do not essentially restrict the perspective on grading to gradable adjectives. As Sapir writes (1944: 94), "every quantifiable, whether existent (say *house*) or occurrent (say *run*) or quality of existent (say *red*) or quality of occurrent (say *gracefully*) is intrinsically gradable". Thus, Pettit and Knobe's considerations on scales extend readily to verbs or verbal expressions, such as "have a desire to", "have the intention to", and so on. And indeed, we can say things like: "Jake very much desired to...", "Jake did not particularly have the intention to...", etc. Some readers have asked whether "intentional" really falls in the category of gradable adjectives. The answer is positive, based on the usual tests of grammaticality for gradable adjectives. As pointed by N. Hansen based on a Google search, comparative constructions such as "more intentional than/less intentional than" are relatively frequent. Constructions with modifiers such as "very/absolutely/extremely intentional" seem to be marginal, but "fully intentional" is fairly common.

"intentionally". We can ask for instance:

(19)   Did the chairman intentionally start the program?

(20)   Did the chairman intentionally harm the environment?

Clearly, "starting the program" and "harming the environment" correspond to distinct events. Starting a program may be accomplished by simply saying: "let us start the program!". Harming the environment, on the other hand, is the outcome of a number of intermediate events following the chairman's order to start the program.

Bearing in mind this distinction, the hypothesis I would like to consider is that our judgments about whether a particular event (action or outcome) is intentional depend on two main dimensions. These dimensions concern the *desire* to bring about a particular outcome on the one hand, and the knowledge or capacity to *foresee* that a particular plan or action taken will lead to that outcome on the other. Following Nichols and Ulatowski (2007), we may call these two dimensions *motive* and *foreknowledge*, but bearing in mind that motive pertains to how much one wants or desires a particular outcome, and that foreknowledge pertains to how much one can anticipate or control that one's actions will have a particular outcome.[21] In particular, the notion of desire is slightly broader than the notion of motive, since it will also help us to qualify the attitude an agent has toward subsidiary outcomes that are not his or her main motive. Likewise, it may be more appropriate to talk of control rather than foreknowledge, to factor in not only the ability to foresee an outcome, but also the element of agency involved in bringing about the outcome. However we talk of knowledge primarily to emphasize the epistemic element of rational anticipation involved in acting in a particular way.

The relevance of these two aspects to the concept of intentional action is not new. To a large extent, it can be traced to Aristotle's analysis of voluntary and involuntary action.[22] Fundamentally, knowledge attaches to the relation between means and ends. It concerns one's ability to anticipate that a particular action or process will bring about a particular outcome. Desire, on the other hand, attaches first and foremost to ends. For Aristotle, if I am compelled to make an action, the outcome of which I find undesirable, this action is not voluntary, even if I can foresee the outcome as a consequence of that action. Similarly, if I make an action whose outcome escapes my control entirely, the outcome is accidental on Aristotle's account, and can no longer be considered voluntary.

---

[21] According to Machery (p.c.), foreknowledge and control should be more sharply distinguished, since an agent can desire an outcome, adequately foresee it as a result of some action, but simply lack control over the process linking action and outcome. Cova too in his work on the Knobe effect originally distinguished control as a separate dimension. If Machery is right, to say that the semantic of "intentional" involves two dimensions of comparison remains an oversimplification, it should involve at least three dimensions, and possibly more.

[22] "By the voluntary I mean, as has been said before, any of the things in a man's own power which he does with knowledge, i.e. not in ignorance either of the person acted on or of the instrument used or of the end that will be attained (e.g. whom he is striking, with what, and to what end), each such act being done not incidentally nor under compulsion (e.g. if A takes B's hand and therewith strikes C, B does not act voluntarily; for the act was not in his own power)." Aristotle (*Nich. Ethics*, V, 10, 1135a-b)

The distinction also plays a role in the design of Knobe's scenarios. Indeed, Knobe's two kinds of scenarios, side-effects on the one hand, deviant causal chains on the other, can be seen roughly as ways of manipulating these two parameters. Side-effects are outcomes that are not intrinsically desired by an agent, but that she can yet foresee as consequences of her actions. Knobe's deviant causal chains on the other hand correspond to outcomes that are intrinsically desired by an agent, but brought about in ways that are no longer foreseeable by the agent as the outcomes of her action at the time the action is performed.[23]

Thirdly, the distinction between these two aspects has been argued to play a role, quite convincingly, by Nichols and Ulatowski (2007), in their analysis of Knobe's side-effects proper. Indeed, Nichols and Ulatowski point out that in the side-effect scenarios, there is a minority of people who consistently deny that the chairman intentionally harmed the environment. In debriefing, these subjects typically emphasize that the chairman's "original intent was not to harm the environment". Conversely, the majority of people who consider that the chairman did intentionally harm the environment generally stress the fact that the chairman "knew that he would harm the environment". Therefore, the difference between these two kinds of subjects seems to be that they do not weigh the two dimensions of the chairman's action in the same manner.

Most likely, these minority judgments are just flipped in the case of Knobe's deviant causal chain scenarios. If I consider that Jake intentionally killed his aunt, I discount the randomness of the process, and put stress on Jake's initial desire to bring about that outcome. But I can also consider that Jake did not intentionally kill his aunt, in which case I put stress on the randomness of the process that brought about the action, and on Jake's inability to secure that his action would lead .

Finally, Knobe's study on the comparison between:

(21)   Did the chairman have an intention to help [harm] the environment?

(22)   Did the chairman intentionally help [harm] the environment?

suggests that subjects on average will be less likely to accept that the chairman had an intention to harm or help than to accept that the chairman intentionally harmed or helped the environment. In the help case, 0% the subjects probed by Knobe accepted "the chairman had an intention to help the environment". This suggests that they rely on the fact that help is not the chairman's motive for her initial action, even though the chairman can foresee help in the same way in which he can foresee harm as a consequence of his action.

---

[23]Fiction is rife with similar cases of dissociation between one's desire to bring about a particular outcome, and one's control over the process leading to that outcome. J. Cleese and C. Crichton's *A Fish named Wanda* stages a character who eventually succeeds at causing the death of an old lady by unexpectedly killing her dog instead of her (I am indebted to E. Dupoux for reminding me of that scene). L. Buñuel's film *Ensayo de un Crimen*, based on R. Usigli's novel, tells the story of a character named Archibald de la Cruz who strongly desires the death of several women during his lifetime: all of them die soon after that, but never according to Archibald's plans to kill them, and almost always without any causal link to those plans (except in one case in my opinion). Archibald ends up accusing himself of several murders, but is immediately relaxed by the officer who points out that mere desires to take actions legally cannot be blamed.

Though the asymmetry between harm and help is still observed by Knobe between the two conditions, what appears is that when judging of help, subjects focus primarily on whether help was intrinsically desired, rather than simply foreseen.[24]

4.2 Scales for desire and foreknowledge

Granting the distinction between these two main dimensions along which an action may be intentional, what we would like to see is how they are integrated in the semantics of the adjective "intentional" proper. In the previous section, we proposed that an action x of type A is intentional if its degree of intention along the relevant scale is above some contextually determined threshold, dependent on A, namely if:

(23) $\mathbf{intentional}(x) \succ \mathbf{norm}(A)(intentional)$

As explained previously, we do not need more than these truth conditions to account for Knobe's asymmetry, provided we assume as available an integrated scale of comparison. What I would like to suggest however is that we get a more fine-grained view of judgments about intentional action by distinguishing between at least those two dimensions.

Consider what happens if we assume that intention (or "intentionalness") is a factor of two dimensions, namely desire and foreknowledge. Subjects who have to judge whether a person's action or outcome is intentional should check whether that action or outcome is sufficiently desired by the agent on the one hand, and whether it is sufficiently foreseen or foreseeable by the agent as a consequence of his actions on the other.[25] Because desire and foreknowledge are also susceptible to degrees, this means that one needs to check whether:

---

[24]As pointed out to me by N. Hansen, Austin (1966) draws an interesting distinction between doing an action *intentionally* and doing it *deliberately*. Austin gives the example of a hurried driver who has to drive fast and who happens to drive over a gocart "deliberately, but not intentionally". The setting of Austin's scenario indicates that the driver can foresee and control that he will drive over the gocart, though he would rather do otherwise if he did not have to speed up. Austin writes: "at no time did I intend to drive over it. It was incidental to anything I intended to do". Interestingly, however, Austin writes: "I did drive over the gocart deliberately, but not intentionally – *nor of course, unintentionally either*" (emphasis mine). Thus, Austin's judgment seems to indicate that for him to say the action was done intentionally, it had to be properly desired by the agent, or to be his main motive. In terms of our approach, the fact that Austin accepts that the action was not intentional proper, but not unintentional either, is compatible with two explanations. One explanation is in terms of the two dimensions we postulate: possibly, the action is "not intentional" because it falls below the standard on the *desire* scale, but still it fails to be "unintentional" because it falls above the standard on the *knowledge* scale. The other explanation is that "intentional" and its antonym "unintentional" set two distinct standards on the same scale. Consider for instance the *desire* scale. In Austin's scenario, the degree of inclination of the driver to drive over the gocart is arguably not high enough to be deemed "intentional", but is still too high on the scale to be considered "unintentional" (after all, the driver is described as free to stop and move the gocart, so he is not opposed to driving over it). The latter explanation strikes me as more plausible than the former in this case. A fuller discussion of this example would lead us to discuss the status of antonyms in greater detail.

[25]Here and in what follows, I tend to use "is foreseen" and "is foreseeable" indistinctly, the question being how much an agent *can* foresee a particular consequence of his action at the time it is taken.

(24)   a.   **desired**$(x) \succ$ **norm**$(A)(desired)$
         b.   **foreseen**$(x) \succ$ **norm**$(A)(foreseen)$

An important caveat at this point is that each of the notions may in turn correspond to the integration of further dimensions. For instance, desire for an outcome may itself be evaluated along several dimensions. Likewise, how much an outcome is foreseeable involves how much it is planned, but also how much it is caused by the agent, two aspects which may need separating on a more fine-grained analysis. We shall disregard these complexities here, however, in order to focus on intentional action as a function of the two parameters we distinguished.

Assuming that agents evaluate the intentional character of an action along those two dimensions, several questions immediately arise. The first concerns the structure of each of the scales for desire and for foreknowledge. The second concerns the integration of the two dimensions in order to evaluate whether an action is intentional.

Consider first the question of the structure of the desire scale. Let us focus on desire in relation to Knobe's scenarios about the environment. In the harm as in the help condition, the chairman of the board is described as *indifferent* between harming or not harming the environment, and likewise as indifferent between helping or not helping the environment. As Tannenbaum & al.'s figures on "had a desire" indicate, that same degree of desire, which can take to lie on the middle of the scale in each case, falls much below the expected degree for help. By contrast, it certainly lies above the expected degree for harm. This means that one would have expected the chairman to have more propensity to help, and less propensity to harm. Pettit and Knobe sum up the situation in this way:
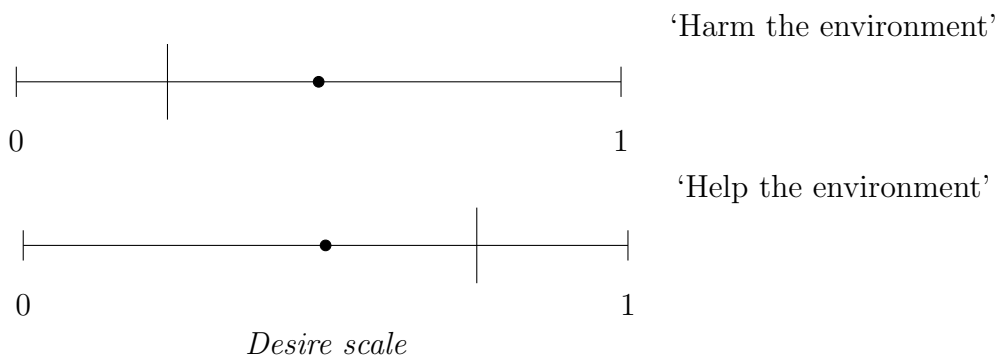
> In the help condition, one is inclined to think: 'How callous! Surely, any reasonable person would be at least a little bit more in favor of this outcome.' But in the harm condition, one has exactly the opposite reaction: 'How blasé! It seems like anyone should be at least a little bit more opposed to this outcome.'

This means that it would be a mistake to consider that indifference corresponds to the lowest degree on the scale of desire. Obviously, harming or helping the environment is definitely not the main motive for the chairman's action. However, from the fact that the chairman desires to make profit, knows that making profit through the program will involve harming or helping the environment, and finally decides to start the program, we can at least infer that the chairman *is not opposed* to either harming or helping, as Pettit and Knobe adequately put it. In other words, the scale associated to what we call motive or desire really is a scale of bouletic inclination, such that the lowest degree on that scale corresponds to an outcome that is *fully undesired* by the agent, while the maximum degree on that scale corresponds to outcomes the agent *fully desires*.[26] Outcomes that are indifferent should naturally fall on the middle of the scale. On the other hand,

---

[26]As pointed out by a reviewer, we make the assumption here that scales for desire and knowledge are closed scales rather than open scales. This assumption is not substantial to our account, but it is partly motivated by the fact that one can say that an agent *fully* believes or *perfectly* knows a proposition, and that he *absolutely* desires some outcome, or is *completely* opposed to it. If belief and knowledge are

the normative standard for whether an agent desires a particular outcome or not will be positioned above or below that middle point depending on the kind of outcome under consideration.

In the figure below, the black dot on the middle of the scale thus marks the effective degree of desire of the chairman toward either help or harm, namely indifference in both cases; the vertical bar marks the normative threshold in each comparison class, such that values above the threshold determine a 'yes' answer, and values below the threshold a 'no' answer regarding the ascription of the relevant property (desire to harm / help); the scale is taken to go from 0 to 1 where 0 marks an attitude against the event, and 1 an attitude for the event.[27] Although the effective degree is set at the same $1/2$ or indifference position in each case, the variation of the threshold is here compatible with the idea that the chairman "had a desire to harm the environment", but "did not have a desire to help the environment".

'Harm the environment'

0   1

'Help the environment'

0   1
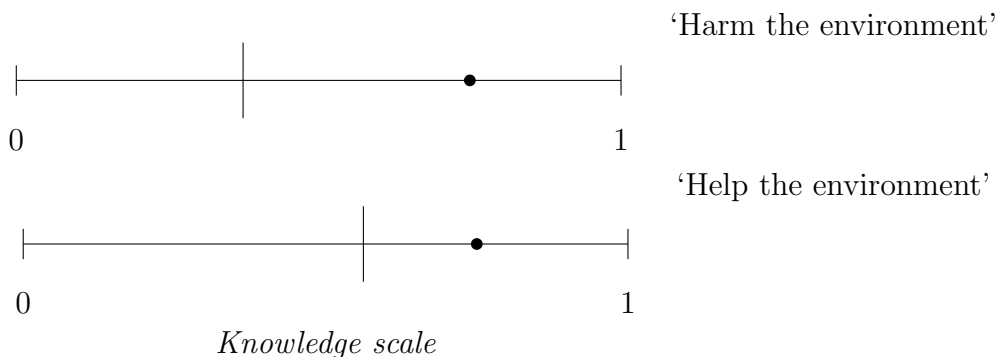
Desire scale

Let us now turn to the scale relevant for foreknowledge. In this case, assuming the scale also ranges from 0 to 1, let 0 denote total ignorance or unawareness of the event being a possible consequence of one's action, and 1 mark full knowledge or awareness of the event being a possible consequence of one's action. By stipulation, let the middle degree $1/2$ to correspond to the agent being aware but uncertain as to whether the outcome will happen. Consider Knobe's scenarios again. By construction of the scenario, the chairman is equally informed or aware of help and harm being consequences of his action in each scenario. Arguably, the degree to which the chairman can foresee help and harm would be set typically above $1/2$, considering that the vice-president of the CEO makes a reliable prognosis. In the help case as in the harm case, furthermore, that degree is each time presumably above the normative threshold.

_____

referred to probabilities, the assumption of closeness will appear as a standard assumption. In the case of the bouletic scale, measurement is obviously a more involved issue. Either way, I believe that the choice between a closed scale and an open scale is tangential to the account presented here.

[27]To mark valence, it may be more natural to assign value 0 to the midpoint, and values -1 and +1 to the opposite ends. Our choice of 0 and 1 is mostly to get a notion of distance that is automatically normalized for all scales in order to allow inter-attribute comparison (see below). Nothing hinges on that, and the reader may choose other units depending on the attribute, so long as all scales are eventually normalized to permit scale comparison.

A question that naturally arises is: could it be that even within the dimension of fore-knowledge the threshold relevant for help is set higher than the threshold for harm? This question was investigated recently by Beebe and Buckwalter, who indeed found a contrast, which they call "epistemic side-effect effect". Beebe and Buckwalter's data indicate that subjects tend to assent more readily to: "the chairman knew he would harm the environment" than to "the chairman knew he would help the environment". One way to account for this would be to say that the knowledge threshold for harm is positioned slightly lower for help than for harm, as in the figure below:[28]

'Harm the environment'

'Help the environment'

Knowledge scale

Here we arbitrarily assigned the threshold for Help above $1/2$, and the threshold for Harm to below $1/2$, to mark that it is easier to ascribe knowledge of harm than knowledge of help. An important point is that the position of the effective degree of foreknowledge above the threshold for Harm and for Help would support both ascriptions that "the chairman knew that he would harm [help] the environment". In Beebe and Buckwalter's study, however, subjects were asked graded judgments about the acceptability of the sentence "the chairman knew he would harm [help] the environment". They found that on average, subjects gave significantly higher scores for harm than for help. Based on our semantic framework, we may conceive that subjects give a higher score as the distance of the effective degree to the normative threshold is itself greater. This observation could be generalized to judgments of desire and of other attributes, for any situation in which graded judgments are elicited.

4.3   Weighing desire and foreknowledge

So far we have only shown that by distinguishing a scale for desire and a scale for knowledge, we could in principle account for Knobe contrasts relative to each of these dimensions taken separately. What we need to examine is how these two dimensions are integrated in judgments about intentional action. Imagine a subject whose semantic representations

---

[28]As pointed out to me by J. Knobe, one might wish to relate the contrast found by Beebe and Buckwalter to Pighin et al.'s data about probability judgments. Even when the objective chances of harm and help are the same, in other words, that probability is perceived as "higher" for harm, and correlatively it should become easier to ascribe foreknowledge of the outcome to the chairman.

for the chairman's degree of desire, of foreknowledge, and for the normative thresholds are as in the above figures with regard to Harm. Let x denote the event of "harming the environment". For our subject, the following two conditions are true together:

(25)   a. $\textbf{desired}(x) \succ \textbf{norm}(Harm)(desired)$
           b. $\textbf{foreseen}(x) \succ \textbf{norm}(Harm)(foreseen)$

In this case, we should expect that the subject will judge the sentence "the harm caused to the environment by the chairman was intentional" to be true. More generally, when both conditions $\textbf{desired}(x) \succ \textbf{norm}(A)(desired)$ and $\textbf{foreseen}(x) \succ \textbf{norm}(A)(foreseen)$ are true together, one should expect the action x of type A to be judged intentional. Similarly, when both conditions are false together, the action should be evaluated to be not intentional. The problem concerns mixed cases, namely situations in which one condition is satisfied and the other is not. What we should we say of such cases?

Let us consider a hypothetical subject whose semantic representations for the chairman's degree of desire, of foreknowledge, and for the normative thresholds are as in the above figures, this time with regard to Help. By hypothesis, the subject would assent to "the chairman knew that he would help the environment", but not to "the chairman had a desire to help the environment". What would such a subject say of: "the chairman helped the environment intentionally"? Several possibilities are open from a theoretical point of view: one option could be that the rule for such cases is to go for the *worst case*, namely falsity. A second possibility could be that one of the two dimensions is systematically given more weight over the other. For instance, it may happen that desire counts as systematically more relevant than knowledge to evaluate whether an action is intentional. A third possibility might be *freedom* regarding the choice of which dimension counts as most relevant. This option might be the most plausible if situations of this kind are situations in which one can say:

(26)   In a sense, the chairman intentionally helped the environment, because he knew he would help the environment. But in another sense, the chairman did not intentionally help the environment, because he had no desire to help the environment.
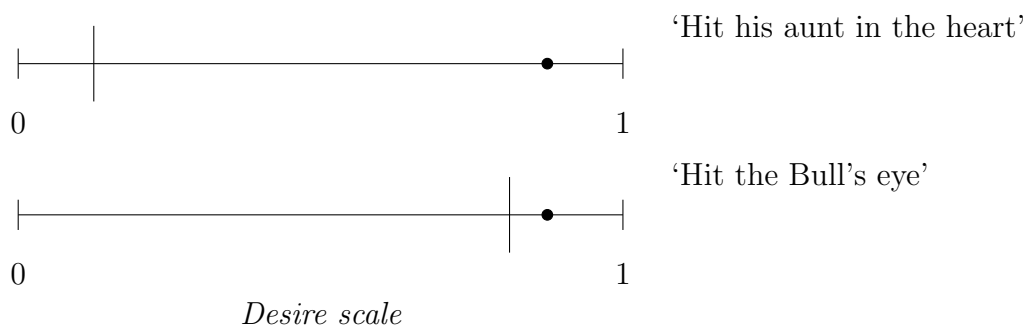
Even more options are conceivable. In particular, we can imagine a more fine-grained rule, whereby we actually compare normalized distances below and above the normative thresholds in each dimension. For instance, using d and f are abbreviations for *desired* and *foreseen*, we could imagine that if $\textbf{d}(x) > \textbf{norm}(A)(d)$ and $\textbf{f}(x) < \textbf{norm}(A)(f)$, then if $|\textbf{d}(x) - \textbf{norm}(A)(d)| > |\textbf{f}(x) - \textbf{norm}(A)(f)|$, the action is judged intentional, and conversely, if $|\textbf{f}(x) - \textbf{norm}(A)(f)| > |\textbf{d}(x) - \textbf{norm}(A)(d)|$, then the action is judged not intentional. What the rule says is: 'if the degree to which one attribute is positively above the norm comparatively exceeds the degree to which the other attribute is negatively below the norm, assert intentional action with the positive; conversely, deny intentional action with the negative.' Such a rule may seem complex at first, but what it means is that our judgments about graded properties may not simply rely on whether the target degree is below or above some standard, but also on how much below or above the standard that
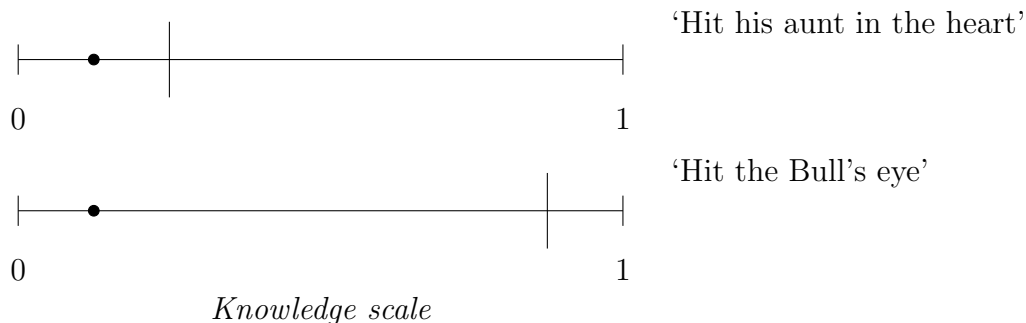
degree is.

Which of these options is the most adequate in judgments of intentionality remains a matter of empirical investigation. In principle, the best method to try and see how judgments about desire, foreknowledge, and finally intention interact would be to systematically compare graded answers to the three questions:

(27) a. (How much) Did the agent intentionally bring about x?
   b. (How much) Could the agent foresee that he would bring about x?
   c. (How much) Did the agent desire to bring about x?

A particularly interesting case to try and tease apart the hypotheses we distinguished concerns Knobe's rifle scenarios. Based on the data we have, it seems to me that at least the second hypothesis, namely the hypothesis that desire would systematically prevail over foreknowledge, could not be an adequate hypothesis in general. In the Bull's eye case, Jake's degree of desire is certainly high on the scale and above the standard, though it should count as normally high for a situation of contest in which winning is the purpose, so we can represent it as not too much above the standard. By contrast, Jake's degree of foreknowledge of the outcome is comparatively well below the expected threshold for a situation of rifle contest, leaving a much wider gap. Because of that, I believe that one's judgment about whether Jake intentionally hit the Bull's eye would tend to follow the negative judgment one would typically issue regarding whether "Jake could foresee that he would hit the Bull's eye". In other words, in the rifle contest context the dimension of foreknowledge is more discriminant than the dimension of desire, in agreement with our fourth hypothesis.

'Hit his aunt in the heart'

0   1

'Hit the Bull's eye'

0   1

Desire scale

'Hit his aunt in the heart'

'Hit the Bull's eye'

Knowledge scale

Consider now the question: "did Jake intentionally hit his aunt in the heart?". Here, Jake's actual desire to hit his aunt in the heart is certainly by itself very significantly above the expected threshold, simply because one does not normally expect people to want to kill their aunts. The position of Jake's degree of foreknowledge of the outcome of his action is less clear. Let us focus, however, on subjects who would consider that Jake could not know that he would hit his aunt in the heart at the time he shot, given his very bad control of the gun. Comparatively to the rifle contest, Jake's lack of control is obviously less significant in the situation, and therefore falls closer to the standard relevant to ascribe knowledge of the outcome. Because of that, my hunch is that a significant proportion of the subjects who would represent Jake's action as specified on the figure above would tend to align their judgment that "Jake intentionally hit his aunt in the heart" with his Jake's abnormal degree of desire. Nevertheless, I believe some of the same subjects may still discount the desire dimension in this case, and deny that Jake intentionally hit his aunt in the heart, if indeed they have a reason to focus on the way in which the action is performed, rather than on motivations.

Based on these considerations, I am tempted to conclude that in cases in which the evaluation of an action should yield opposite verdicts along the two dimensions we distinguished, the evaluation of whether the action is intentional will tend to go with the dimension for which the departure below or above the norm is judged the most significant. In some cases, however, subjects may still focus exclusively on one dimension, discounting the other, depending on what aspect of the action is more salient to them.

4.4   Machery's cases

Another test case for our articulation of Pettit and Knobe's hypothesis concerns a pair of scenarios originally put forward by E. Machery to criticize Knobe's main hypothesis, namely that moral considerations involving praise and blame play an essential role in our evaluation of intentional action. Machery compared answers given by subjects to the following two cases, in which subjects were asked to answer by yes or no:

Free cup case: Joe was feeling quite dehydrated, so he stopped by the local smoothie shop to buy the largest sized drink available. Before ordering, the

26

cashier told him that if he bought a Mega-Sized Smoothie he would get it in a special commemorative cup. Joe replied, 'I don't care about a commemorative cup, I just want the biggest smoothie you have.' Sure enough, Joe received the Mega-Sized Smoothie in a commemorative cup. Did Joe intentionally obtain the commemorative cup?

Extra dollar case: Joe was feeling quite dehydrated, so he stopped by the local smoothie shop to buy the largest sized drink available. Before ordering, the cashier told him that the Mega-Sized Smoothies were now one dollar more than they used to be. Joe replied, 'I don't care if I have to pay one dollar more, I just want the biggest smoothie you have.' Sure enough, Joe received the Mega-Sized Smoothie and paid one dollar more for it. Did Joe intentionally pay one dollar more?

What Machery found is that 95% of the subjects responded "yes" in the extra dollar case, against only 45% in the free cup case. Moreover, Machery found an overwhelming majority of subjects judging the agent's action to be morally neutral (as opposed to blameworth or praiseworthy) in the two cases. Based on these data and further evidence, Machery surmises that subjects, in this scenario as in the other Knobe-type scenarios, evaluate the negative side-effects in the same way they evaluate the extra-dollar case here. Because the agent incurs a cost (the harm caused) in order to reap a benefit, the action is judged to be more intentional.

Machery's suggestion that moral consideration about praise and blame do not seem to play a role in the contrast found in his scenarios seems entirely correct. However, one should examine whether Machery's scenario might directly be accounted for along the lines suggested by Pettit and Knobe. We basically agree with Machery's assumption that consideration of a tradeoff is relevant in this case. But our point will be that the tradeoff in question only affects the way standards of comparison are positioned for the two kinds of events under consideration.

Let us call s the event of getting a mega-sized smoothie, p the event of paying one extra dollar, and c the event of obtaining a commemorative cup. Whether these events are foreseen by Joe depends on the moment in the story. At the moment Joe gives his answer to the cashier, these events are clearly and equally foreseen by the agent *qua* outcomes of his actions. In that case, therefore, we predict that distinctions must essentially concern the dimension of desire. We say "essentially", however, because how much an outcome is foreseen can here too be sensitive to the kind of event under consideration. Something Machery points out himself is that paying an extra dollar is a *means* in order to achieve the main goal of getting a smoothie, whereas obtaining a commemorative cup is a *consequence* of that goal. More importantly, it seems to us that *paying* something or not is an action that is directly under the agent's control, whereas *obtaining* something or not is not as immediately under the agent's control (for instance if "obtaining a cup" were semantically analyzed as "being given a cup"). Because of that, it may be the case that the degree we attach to how much Joe can foresee that he will pay is set higher above the

corresponding standard than the degree we attach to how much Joe can foresee that he will obtain a commemorative cup. However, that possible contrast is arguably not prevalent in comparison to what concerns our evaluation of Joe's desire to pay an extra dollar, as opposed to his desire to get a free cup.

So let us turn to our evaluation of Joe's desires. Joe is presented as determined to get a smoothie, but as indifferent toward getting a free cup or not, and also indifferent toward paying an extra dollar or not. We may assume that $\mathbf{desired}(c) = \mathbf{desired}(p) = 1/2$ (on a scale from 0 to 1), namely that Joe's absolute degree of desire for the cup is equal to his absolute degree of desire to pay one dollar, and expresses indifference: Joe is neither opposed, nor unwilling. The main difference between the two events, however, is that in the one case, Joe gets a bonus for free, while in the other, he incurs a cost.

What happens in this case seems to be the following: at equal costs, one generally expects any increase in the overall utility obtained to be all the more desirable. Conversely, for utilities that are equal, one generally expects any increase in the cost to be all the less desirable. Each of these expectations is deceived in the scenarios of Machery, because in the free cup case, one would normally expect Joe's desire for the free cup to be higher than it is, and conversely, we should expect Joe's desire to pay one extra dollar to be lower than it is. Consequently, we have that: $\mathbf{desired}(c) \preceq \mathbf{norm}(free\_good)(desired)$ and $\mathbf{desired}(p) \succ \mathbf{norm}(extra\_cost)(desired)$. In that case, we therefore take the two parameters along which the expectations differ to concern the fact that getting the cup is a free bonus on the one hand, and that paying a dollar is an extra cost on the other.

In this example as in the previous one, we should stress that the configuration we are describing purports to account for the contrast between the two cases, and in particular, for the lesser propensity to consider that Joe intentionally obtained the cup, as opposed to Joe intentionally paying a dollar more. In this case as in the foregoing, the present account does not rule out the possibility that some subjects would consistently deny that Joe intentionally paid a dollar more, or consistently agree that Joe intentionally obtained a cup. Regarding the former, in particular, I believe various subjects could say:

(28)   Before the cashier told him, Joe expected to buy a smoothie, but he did not expect to pay one dollar more.

As we can see, however, this relates back to the foreknowledge dimension, but the sentence refers to a moment that is anterior to the cashier's informing Joe. At that moment, Joe already had formed an intention to buy a smoothie, but simply could not foresee the implications of his buying a smoothie. This issue, however, concerns essentially the temporal underspecification in the question asked, namely the point of time relevant for evaluation of desire and foreknowledge. As such, it does not call into question the relevance of those two dimensions.

5 Conclusion

In this paper we examined different ways in which Pettit and Knobe's main hypothesis can be refined in order to explain Knobe-type contrasts. Like Pettit and Knobe, we tend to see these asymmetries as pervasive, but primarily because we view them as a particular case of the context-sensitivity of gradable expressions, and of the sensitivity of the semantics of words like "intentional", "intend", "desire", to expectations and to normative standards that vary with the kind of argument these expressions are predicated of.

In particular, we saw that similar contrasts as the one evidenced by Knobe are found in judgments involving the assessment of probabilities and quantities. Thus, in the same way in which the normative standard for what counts as "intentional" in the case of harm is set lower on the relevant scale than the standard for what counts as "intentional" in the case of help, the normative standard for what counts as "many" in judging of casualties of a harmful event will typically be set lower than the standard for what counts as "many" in relation to survivors.

Two consequences of our analysis may be pointed out for the debate concerning the interpretation of the Knobe effect. The first is that, in agreement with Machery in particular, we do not consider that Knobe-type asymmetries primarily have to do with moral considerations pertaining to praise or blame. Nevertheless, we do agree with the general line of Pettit and Knobe on the idea that in a lot of cases, our judgments about intentional action, just like the judgments about quantities or probabilities we considered in this paper, are sensitive to utility parameters and to our normative expectations depending on how positively or negatively we value particular outcomes. A second consequence is that on our view Machery's own contrasts can in turn be referred to the same semantic mechanism whereby the norm for what counts as intentional can be shifted depending on the kind of event under consideration.

Regarding the particular semantics of the adjective "intentional", finally, we have argued that this predicate involves at least two main dimensions of comparison, one pertaining to the degree of desire an agent has to bring about an outcome, the other to the degree to which the agent can foresee the outcome as a consequence of her action. Both dimensions appear to be manipulated in several of the scenarios proposed by Knobe and others. In particular, we have suggested that within each dimension, Knobe-asymmetries can occur. The distinction between at least two main dimensions of comparison for "intentional" also allows us to account for more than the Knobe asymmetry, namely for the ambiguity felt in cases in which the evaluation of the action goes in opposite directions along each dimension. For such cases, we have argued that ascriptions of intentional action are more likely to follow the dimension for which the departure from the norm is comparatively the most salient.

References

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

[2] Alicke M. (1992). Culpable causation. *Journal of Personality and Social Psychology*, 63, 368-378.

[3] Alicke M. (2000). Culpable control and the psychology of blame. *Psychological Bulletin* 2000, Vol. 126, No. 4, 556-574

[4] Austin, J. L. (1966). Three Ways of Spilling Ink. *The Philosophical Review*, 75 (4), pp. 427-440.

[5] Bartsch, R., & Vennemann, T. (1972). The grammar of relative adjectives and comparison. *Linguistische Berichte*, 20, 19-32.

[6] Beebe, J. R. & Buckwalter, W. (2010). The Epistemic Side-Effect Effect. *Mind and Language*, 25 (4), pp. 474-498.

[7] Bonnefon J-F. & G. Villejoubert (2006). Tactful or Doubtful? Expectations of Politeness explain the Severity Bias in the Interpretation of Probability Phrases. *Psychological Science*, 17, 747-751.

[8] Fara D. (2000). Shifting Sands: an Interest-Relative Theory of Vagueness. *Philosophical Topics* 28 (1). Originally published under the name "Delia Graff".

[9] Geurts B. (2009). Goodness, in *Proceedings of the 17th Amsterdam Colloquium*, ed. by M. Aloni, Harald Bastiaanse, Tikitu de Jager, Peter van Ormondt and Katrin Schulz, pp. 277-85.

[10] Kennedy C. (2007). Vagueness and Grammar: the Semantics of Absolute and Relative Gradable Adjectives. *Linguistics and Philosophy* 3) (1): 1-45.

[11] Knobe, J. (2003a). Intentional Action in Folk Psychology: An Experimental Investigation. *Philosophical Psychology*, 16, 309-324.

[12] Knobe J. (2003b). Intentional Action and Side-Effects in Ordinary Language. *Analysis*, 63, 190-193

[13] Knobe, J. & Fraser, B. (2008). Causal Judgment and Moral Judgment: Two Experiments. In W. Sinnott-Armstrong, *Moral Psychology*, Cambridge, MA: MIT Press. 441-448.

[14] Knobe J. (forthcoming). Causal action trees. To appear in *Topics in Cognitive Science.*

[15] Lappin S. (2000). An intensional parametric semantics for vague quantifiers. *Linguistics and Philosophy* 23, 599-620.

[16] Machery E. (2008). The Folk Concept of Intentional Action: Philosophical and Experimental Issues, *Mind & Language*, 23, 165-189.

[17] Malle B.F. (2006). Intentionality, Morality, and Their Relationship in Human Judgment. *Journal of Cognition and Culture* 6.1-2

[18] Nadelhoffer T. (2005). Skill, Luck, Control, and Intentional Action. *Philosophical Psychology* 18 (3), 341-352

[19] Nichols S. & J. Ulatowski (2007). Intuitions and individual differences: The Knobe effect revisited. *Mind & Language*, 22, 346-365.

[20] Nouwen R. (2011). Degree modifiers and monotonicity, in *Vagueness and Language Use*, P. Egré and N. Klinedinst (eds.), Palgrave McMillan, pp. 146-164.

[21] Partee B. (1989). Many Quantifiers. Repr. in *Compositionality in Formal Semantics, Selected Papers by Barbara Partee*, pp. 241-258.

[22] Pettit D. & J. Knobe (2009). The Pervasive Impact of Moral Judgments. *Mind and Language* 24 (5), pp. 586-604.

[23] Pighin S., J-F. Bonnefon, L. Savadori (2009). Overcoming number numbness in prenatal risk communication. Submitted manuscript, University of Toulouse and Department of Cognitive Science and Education, University of Trento.

[24] Sapir E. (1944). Grading, a Study in Semantics. *Philosophy of Science*, Vol. 11, No. 2, pp. 93-116.

[25] Tannenbaum, D., Ditto, P.H., and Pizarro, D.A. 2007. Different moral values produce different judgments of intentional action. Unpublished manuscript. University of California-Irvine.

[26] Thomason R. (This volume). Formal Semantics for Causal Constructions.

[27] Weber E. U. & D. J. Hilton (1990). Contextual Effects in the Interpretation of Probability Words: Perceived Base Rate and Severity of Events. *Journal of Experimental Psychology: Human Perception and Performance*, vol. 16 (4), 781-789.

[28] Wolff P. (this volume). Causal pluralism and force dynamics.