

# *The Effect of What We Think may Happen on our Judgments of Responsibility*

**Felipe De Brigard & William J. Brady**

**Review of Philosophy and  
Psychology**

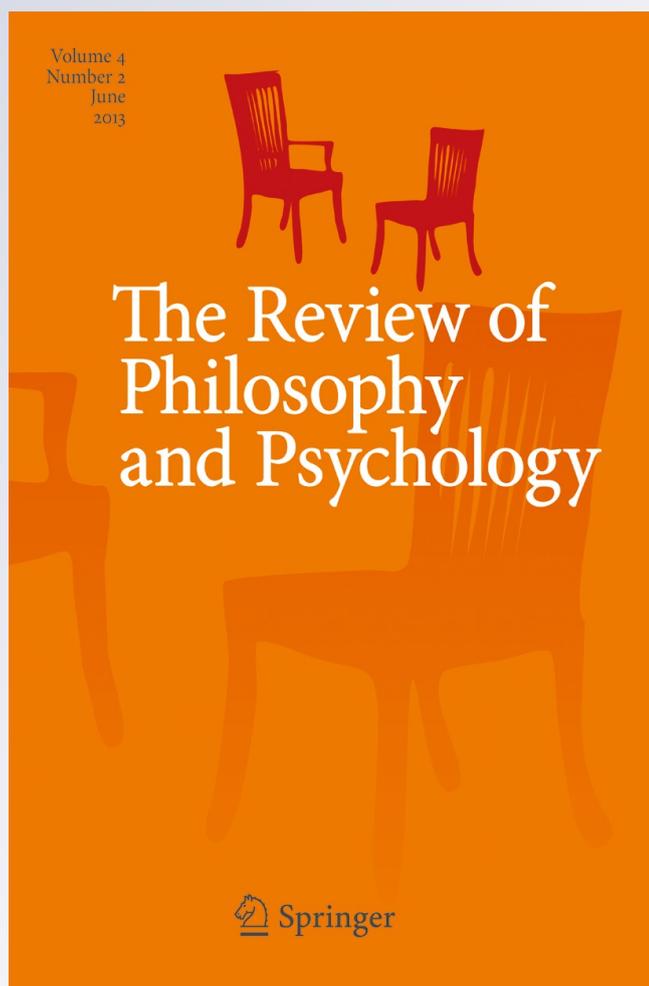
ISSN 1878-5158

Volume 4

Number 2

Rev.Phil.Psych. (2013) 4:259-269

DOI 10.1007/s13164-013-0133-8



**Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media Dordrecht. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# The Effect of What We Think may Happen on our Judgments of Responsibility

Felipe De Brigard · William J. Brady

Published online: 6 February 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** Recent evidence suggests that if a deterministic description of the events leading up to a morally questionable action is couched in mechanistic, reductionistic, concrete and/or emotionally salient terms, people are more inclined toward compatibilism than when those descriptions use non-mechanistic, non-reductionistic, abstract and/or emotionally neutral terms. To explain these results, it has been suggested that descriptions of the first kind are processed by a *concrete* cognitive system, while those of the second kind are processed by an *abstract* cognitive system. The current paper reports the results of three studies exploring whether or not considerations about possible future consequences of holding an agent responsible at a present time affect people's judgments of responsibility. The results obtained suggest first that the concrete system does not produce compatibilist judgments of responsibility *unconditionally*, even when facing appropriately mechanistic, reductionistic, emotionally loaded and concretely worded deterministic scenarios. Second, these results suggest that considerations about possible future consequences for innocent third parties that may follow as a result of holding an agent responsible affect people's judgment as to whether or not the agent is responsible for what she did. Finally, it is proposed that these results compliment extant evidence on the so-called "Side-effect effect", as they suggest that emotional reactions toward possible future side effects influence people's judgment of responsibility. The impact of these results for philosophy and moral psychology is discussed.

## 1 Introduction

Philosophical and psychological research on free will and moral responsibility has increased rapidly (Woolfolk et al. 2006; Nichols 2011). One such prominent area of research pertains to people's judgments of responsibility when considering

---

F. De Brigard (✉)  
Department of Psychology, Harvard University, 860 William James Hall 33 Kirkland St.,  
Cambridge, MA 02139, USA  
e-mail: brigard@wjh.harvard.edu

W. J. Brady  
Department of Psychology, New York University, New York, NY, USA

deterministic versus non-deterministic scenarios (Nahmias et al. 2007). According to many philosophers (e.g., Kane 1996; Pereboom 2001), when confronted with the possibility of determinism—roughly, the view according to which every event is entailed by past events and the laws of nature—people's intuitive reaction is to side with incompatibilism about moral-responsibility: the view according to which determinism and moral responsibility are incompatible (Fischer 1994). The thought is that, by and large, people's default belief is that an agent should not be held responsible for an action she was bound to do. However, results from recent studies cast doubt upon this claim. For example, Nahmias and collaborators (Nahmias et al. 2005; Nahmias et al. 2006) asked participants to make judgments of moral responsibility following the presentation of deterministic scenarios, and they found that, contrary to the received philosophical view, people's judgments were actually in line with compatibilism.

Nonetheless, these results alone do not establish that people are naturally compatibilists. In fact, further studies have shown that if events that preceded the action in a given case are couched in concrete and emotionally salient terms, people tend to produce assessments in accordance with compatibilism. However, when scenarios are abstract and emotionally void people's intuitions tend to be incompatibilist (Nichols and Knobe 2007; Sarkissian et al. 2010). Similarly, when the events that preceded the action are described as occurring in a possible but non-actual world, people's intuitions are more incompatibilist than when events preceding the action are described as occurring in the actual world (Roskies and Nichols 2008). Complementary results have been reported in studies assessing people's judgments of responsibility when presented with reductionistic versus non-reductionistic scenarios (Nahmias et al. 2007). These results suggest that when people face cases in which the events preceding the action are described in purely neurological and non-psychological terms, people's assessments of responsibility vary as a function of how concrete and/or emotionally salient the case is described. When the action in question is described in purely abstract terms, people's responses accord with incompatibilism, whereas if the action is described in concrete and emotionally charged terms, people's responses tend toward compatibilism (De Brigard et al. 2009).

Taken together, the results of these and similar studies suggest that if the deterministic description of the events leading up to a morally questionable action is couched in mechanistic, reductionistic, concrete and emotionally salient terms, people are more likely to hold the agent responsible than when those descriptions use non-mechanistic, non-reductionistic, abstract and emotionally neutral terms. For some theorists, these results are indicative of a more basic psychological asymmetry that cuts across the incompatibilism / compatibilism chiasm. This asymmetric pattern of responses is said to be the result of two putatively different underlying cognitive systems driving our judgments of moral responsibility (e.g., Nichols and Knobe 2007; Sinnott-Armstrong 2008). One such system—call it the *concrete system*—would be sensitive to producing judgments of moral responsibility for mechanistic, reductionistic, concrete, and emotionally loaded deterministic scenarios, whereas the other—the *abstract system*—would be sensitive to producing judgments of moral responsibility for general, abstract and emotionally void deterministic descriptions.

However, several questions about the precise behavior of such systems remain unanswered. One such question is whether the concrete system produces compatibilist judgments of responsibility in every mechanistic, reductionistic, emotionally

loaded and concrete deterministic (MRECD) description *unconditionally*, or whether there are conditions under which MRECD descriptions can nonetheless lead to incompatibilist judgments of responsibility. The studies we present here attempt to shed light on this issue by examining whether considerations about possible future consequences of holding an agent responsible<sup>1</sup> may affect people's judgments of responsibility.

This particular manipulation was chosen for three reasons. First, since the question we wanted to explore was whether MRECD descriptions of the causal history leading up to an action unconditionally generate compatibilist judgments of responsibility, we needed a way to keep the MRECD description fixed between conditions while still looking for differences in judgments. We were able to do so by manipulating considerations about possible future events. Second, the question of whether possible future events that may occur as a result of holding an agent responsible affect people's judgment of responsibility remains largely unexplored. In fact, almost all studies on free will and moral responsibility have focused exclusively on the way in which people's assessment of the *causal history* of events preceding the to-be-evaluated action plays a role in their judgments of responsibility (Nichols 2011). Specifically, researchers have been interested in discerning which features of the description of the causal history preceding an agent's action affect people's judgment of his or her responsibility. A cursory look at the vignettes employed in studies such Nahmias et al. (2005, 2006) and Nichols and Knobe (2007), to name a few, reveals that only events preceding the action upon which the protagonist is judged are ever mentioned, and that there is no reference about possible future events. But there is no reason to think that ordinary people *only* care about the events immediately preceding the agent's action, other than philosophers' insistence that possible consequences that might ensue as a result of holding or failing to hold an agent responsible are irrelevant when it comes to judgments of responsibility. However, it is unclear whether ordinary people only keep in mind the considerations philosophers assume they ought to keep in mind when making judgments of responsibility. As such, the experiments reported here are designed to explore the effect of considering possible future consequences of holding an agent responsible on people's judgment of responsibility.

Finally, a third motivation for choosing this particular manipulation comes from recent evidence suggesting that moral considerations on the side effects of an agent's action affect people's judgments on whether or not such actions were done intentionally (Knobe 2003), whether or not agents performed such actions knowingly (Beebe and Buckwalter 2010), and even whether or not the agent was responsible for such actions (Björnsson and Persson *In press*). However, in virtually all the studies on this so-called "side-effect effect" or "Knobe effect", (Feltz 2007) such side effects are always part of the *history* that preceded the action upon which the agent's responsibility is evaluated. Whether or not considerations of possible future side effects also affect judgments of responsibility remains an open question. The studies we present here attempt to shed light on this issue as well.

<sup>1</sup> To avoid excessive verbiage throughout the paper, whenever we refer to the consequences of holding an agent responsible, we mean holding an agent responsible *at the present time*.

## 2 Study 1

### 2.1 Procedure

Fifty-four undergraduates from UNC-Chapel Hill (Age  $M=20.8$ ,  $SD=4.1$ ) voluntarily participated in this study. None of the participants were philosophy or psychology majors. All participants read a vignette (see below) describing a situation in which an agent performed a morally questionable action as a result of a deterministic chain of events. The description was couched in MRECD terms, akin to those used in many of the studies described above (e.g., Nahmias et al. 2007; Nichols 2011). Half of the participants were randomly assigned to the *Better* condition, in which the vignette stated that the status of an innocent third party may *improve* if the agent is found responsible. The other half were assigned to the *Worse* condition, where it was stated that the status of an innocent third party may *worsen* if the agent is found responsible. The specific vignette used was the following (*Better* condition in parenthesis and italics):

Mary is the single mother of two: Mark, 7, Sally, 4. Mary works most of the day, and although she is known for being fairly patient and good natured, over the last year she has exhibited some unusually aggressive behavior toward her neighbor. Last week, when she came back from work late at night, she couldn't drive into her garage because her neighbor had blocked her driveway with his new BMW. Enraged, she stepped on the gas pedal and crashed her car into her neighbor's. Unfortunately, her neighbor was still inside the car (it was too dark for anyone to see him), and both his legs were seriously broken in several places. Now he is not only suing her for several thousand dollars, but he's also pressing charges. However, a neurologist examined her brain and discovered that, in the last year, Mary has been developing a rare tumor in her frontal lobe. Since the frontal lobe is necessary for emotional suppression—that is, the capacity to control one's emotions—the neurologist claims that, unlike a healthy person, Mary was completely unable to control her rage and her desire to smash the car. "In fact", he says, "any person with this kind of tumor", facing the exact same situation, would have done exactly what Mary did. She couldn't have done otherwise. "If Mary is found responsible for her actions, she may be sent to a federal medical facility for the next 6 months". There she could receive medical treatment, but she won't be able to see her children. Unfortunately [*Fortunately*], during that time, they would be living with Social Services [*Aunt Elizabeth*], in what might be a much worse [*better*] environment for them.

Participants were asked to say whether they agree or disagree with the following statement: "Mary is morally responsible for crashing her car into her neighbor's". A Lickert scale running from 1 to 7 was used to record participants' answers, with 1 corresponding to "Disagree", 7 to "Agree" and 4 to "Undecided".

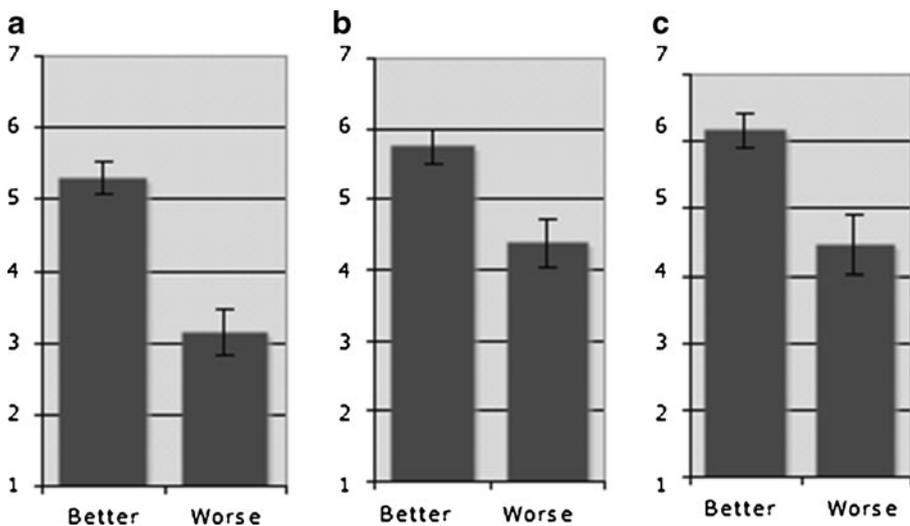
### 2.2 Results and Discussion

Participants were more likely to say that Mary was responsible for running her car into her neighbor's in the *Better* condition ( $M=5.30$ ;  $SD=1.2$ ) than they

were in the *Worse* condition ( $M=3.15$ ;  $SD=1.7$ ),  $t(26)=9.605$ ,  $p<.001$ ,  $r=.59$  (Fig. 1a). These results suggest that, when facing the possibility of an innocent third party's life *improving* as a result of holding the agent responsible, most participants tended to hold the agent responsible even when it was clear that her decision could be traced back to factors—like her tumor—which were outside of her control (compatibilist judgment). However, when the third party could be *worse off* as a result of holding the agent responsible, most people tended to not hold her responsible (incompatibilist judgment). These results suggest that the alleged concrete system, which supposedly elicits compatibilist judgments of responsibility for MRECD descriptions, does not operate unconditionally. Considerations of possible future consequences can sway such judgments toward incompatibilism.

Moreover, these results also suggest that factors other than the causal history leading up the agent's action can influence people's assessment of an agent's responsibility. In particular, these results suggest that the possible consequences that may follow as a result of holding a person responsible can be factored into people's judgments of responsibility. Finally, these results also suggest that moral considerations on side effects need not be confined to side effects that have already occurred, as possible future consequences that have not yet occurred can also sway people's judgment of responsibility in a manner that is consistent with extant results on the side-effect effect (Knobe et al. 2011).

A potential difficulty raised by the vignette employed in Study 1 is that it featured an agent with a neural pathology. Previous studies have shown that the use of cases featuring agents with neural pathologies yields mixed results (e.g., Greene and Cohen 2004; De Brigard et al. 2009). To address this worry, a second study was conducted.



**Fig. 1** Mean responses for both *Better* and *Worse* conditions. **a** displays mean responses for Study 1. **b** displays mean responses for Study 2. **c** displays mean responses for Study 3. Error bars depict standard errors of the mean

### 3 Study 2

#### 3.1 Procedure

Fifty undergraduates from UNC-Chapel Hill (Age  $M=21.7$ ,  $SD=5.2$ ) participated in this study. The methodology was identical to that in [Study 1](#), except that this time around the vignettes depicted scenarios that did not rely on a pathological disorder. Instead, the MRECD description of the events leading up to the action was captured by the use of a brain monitoring system. The specific vignette used was the following (*Better* condition in parenthesis and italics):

Mary is the single mother of two: Mark, 7, Sally, 4. Mary works most of the day, and although she is known for being fairly patient and good-natured, lately she has been under a lot of stress. Due to financial trouble, many of her co-workers are being laid off, while the remaining ones are getting salary cuts. As a result, she has been taking on many more responsibilities and working extra-hours. Yesterday, when she came back from work late at night, she couldn't drive into her garage because her neighbor had blocked her driveway with his new BMW. Enraged, she stepped on the gas pedal and smashed her car into her neighbor's. Unfortunately, her neighbor was still inside the car (it was too dark for anyone to see him), and both his legs were seriously broken in several places. Now he is not only suing her for several thousand dollars, but he's also pressing charges. Coincidentally, while all this happened, Mary was wearing a BrainCap©: a new device that works like a heart monitor but instead of recording the activity of your heart, it records the activity of your brain. In fact, it records all electrical and chemical activity going on in your brain during a particular period of time. Mary was wearing the BrainCap© because she was a participant in a research study at the university—it was a good way of getting some extra cash. The neuroscientist in charge of the study was then called to read the brain activity which occurred during the event, and the result was rather clear: right after seeing the parked BMW, and right when she was about to press the gas pedal after seeing the BMW, Mary's motor cortex (i.e. the area of the brain controlling muscular movements) was active while her pre-frontal cortex was inactive. "This indicates", said the neuroscientist, "that her frontal lobes were unable to suppress the emotions she was feeling, and as a result they could not inhibit the impulse to press the pedal. The pre-motor cortex what then left uninhibited, and it sent the instruction to her foot which led to Mary's pedal pressing. Since there was no brain activity in her frontal lobe, which is required to control emotions, at that moment Mary was completely unable to control her rage and her desire to crash the car. In fact, any person with this kind of brain activity would have done exactly the same thing. The set of brain events that led to the pedal pressing was completely determined; Mary just couldn't have done otherwise". If Mary is found responsible for her actions, she may be sent to a federal medical facility for the next 6 months. There she could receive medical treatment, but she won't be able to see her children. Unfortunately [*Fortunately*], during that time, they would be living with Social Services [*Aunt Elizabeth*], in what might be a much worse (*better*) environment for them.

### 3.2 Results and Discussion

The results of [Study 2](#) showed that participants were more likely to say that Mary was responsible for running her car into her neighbor's in the *Better* condition ( $M=5.75$ ;  $SD=1.26$ ) than they were in the *Worse* condition ( $M=4.38$ ;  $SD=1.76$ ),  $t(23)=12.145$ ,  $p<.001$ ,  $r=.41$  (Fig. 1b). These results were consistent with those obtained in [Study 1](#): participants were more likely to attribute responsibility in the case of a third party possibly being better off as a result of holding the agent responsible than they were if the third party could end up worse off. However, unlike [Study 1](#), participants assigned to the *Worse* condition were no more likely to exculpate the agent, as their answers sat on the mid-point. Nonetheless, these results still suggest that MRECD descriptions of deterministic scenarios do not necessarily elicit compatibilist judgments of responsibility. After all, when participants were confronted with the possibility of an innocent third party being worse off, their responses did not differ from the mid-point, indicating uncertainty as to whether or not responsibility should be ascribed.

In addition, these results lend further support to the suggestion that factors other than the causal history preceding the agent's action can influence people's judgment of responsibility. Whether or not future consequences of holding the agent responsible may make an innocent third party better or worse appear to differentially affect people's judgment of responsibility. Finally, these results also provide additional support to the idea that moral considerations on possible future side effects that have not yet occurred (as opposed to past side effects that have already occurred) can affect people's judgment of responsibility.

At this point it is reasonable to wonder whether the effect of possible future consequences is a more pervasive feature of our judgments of responsibility in general, regardless of whether or not the description of the causal history leading up to the to-be-evaluated action is described in MRECD terms. Moreover, it is reasonable to wonder whether this effect influences our judgments of responsibility even in the absence of a description of the events preceding the action upon which the agent is evaluated. To explore this question a third study was conducted.

## 4 Study 3

### 4.1 Procedure

Forty-eight undergraduates from UNC-Chapel Hill (Age  $M=20.9$ ,  $SD=3.9$ ) participated in this study. The methodology was identical to that in [Studies 1 and 2](#), except that this time all references to the deterministic character of the events prior to the agent's action were left out. The specific vignette used was the following (*Better* condition in parenthesis and italics):

Mary is the single mother of two: Mark, 7, Sally, 4. Mary works most of the day, and although she is known for being fairly patient and good-natured, lately she has been under a lot of stress. Due to financial trouble, many of her co-workers are being laid off, while the remaining ones are getting salary cuts. As a result, she has been taking on many more responsibilities and working extra-

hours. Yesterday, when she came back from work late at night, she couldn't drive into her garage because her neighbor had blocked her driveway with his new BMW. It was cold, very dark, and she was exhausted. Enraged, she stepped on the gas pedal and smashed her car into her neighbor's. Unfortunately, her neighbor was still inside the car (it was too dark for anyone to see him), and his leg was broken. Now he is not only suing her for several thousand dollars, but he's also pressing charges. If Mary is found responsible for her actions, she may be sent to a correctional facility for the next 6 months, and she won't be able to see her children. Unfortunately [*Fortunately*], during that time, they would be living with their alcoholic father [*Aunt Elizabeth*], in what might be a much worse [*better*] environment for them.<sup>2</sup>

## 4.2 Results and Discussion

The results of **Study 3** revealed that participants were more likely to say that Mary was responsible for running her car into her neighbor's in the *Better* condition ( $M=6.17$ ;  $SD=1.24$ ) than they were in the *Worse* condition ( $M=4.46$ ;  $SD=2.21$ ),  $t(22)=9.900$ ,  $p<.001$ ,  $r=.43$  (Fig. 1c). These results are consistent with those from Studies 1 and 2, as participants were more likely to attribute responsibility to Mary if a third party could be better off than if a third party could be worse off in the future. However, unlike study 1, and like study 2, participants were no more likely to exculpate Mary when a third party could be worse off in the future, as their responses did not differ from the mid-point.

Consistent with the results in Studies 1 and 2, the results of **Study 3** suggest that MRECD descriptions of deterministic scenarios do not necessarily elicit compatibilist judgments of responsibility. Once again, when participants were confronted with the possibility of an innocent third party being worse off, their responses indicated uncertainty as to whether or not responsibility should be assigned. Moreover, these results strengthen the suggestion that factors other than the causal history preceding the agent's action influence people's judgment of responsibility. Finally, these results lend further support to the claim that moral considerations on possible future side effects—as opposed to past side effects that have already occurred—can affect people's judgment of responsibility.

## 5 General Discussion

The current studies manipulated considerations about possible future consequences that may follow as a result of holding an agent responsible in order to investigate three inter-related issues. First, we wanted to know whether MRECD descriptions of

<sup>2</sup> This study was initially conducted using the same wording for the *Worse* condition as in the previous two vignettes, so the last line read "Unfortunately, during that time, they would be living with Social Services, in what might be a much worse environment for them". This study, however, did not yield a significant effect ( $p=.08$ ) with an equivalent number of participants. Although the non-significant result in this first study could have been a power issue, we decided to conduct the study anew changing the wording of the *Worse* condition, so instead of "Social Services" it read, as shown above, "alcoholic father". This variation allows us to verify that the effect translates into other possible future negative consequences.

the causal history preceding an action generate compatibilist judgments of responsibility *unconditionally*. The results from the studies reported above suggest that the alleged concrete system does not operate unconditionally, as even in the face of MRECD descriptions, which previous studies have shown to elicit compatibilist judgments of responsibility (e.g., Nichols and Knobe 2007; Nichols 2011), considerations about possible future consequences affect people's responsibility judgments.

Second, and more generally, the current studies investigated whether ordinary people are affected by other considerations besides the causal history of the events preceding the action evaluated. The three studies suggest that people may take into account not only the causal history, but also what could happen in the future if the agent is held responsible at a present time. The particular manipulation employed here showed that if holding an agent responsible could bring about a possible future *undesirable* consequence for an innocent third party, participants were *less inclined* to attribute responsibility to that agent than if the consequence was desirable.

Finally, by specifically manipulating possible future side effects of holding an agent responsible, these studies attempted to shed light on whether the side-effect effect (Knobe 2003; Feltz 2007) only applies to side-effects that have occurred prior to the moment in which the agent is judged. The results suggest that people factor in possible future consequences of holding an agent responsible in a way that accords with extant evidence on the side-effect effect. That is, people are more inclined to say that the agent is responsible for the action when the future side effect is negatively valued than when it is positively valued.

There was one important difference that emerged between the results in [Study 1](#), and [Studies 2](#) and [3](#). While in all three studies participants gave higher ratings of responsibility when the possible future side effects were positively valued, only in [Study 1](#) were negatively valued possible future side-effects associated with participants' exculpatory judgments. What could explain this asymmetric pattern of results? One possibility is that, when participants read about the possible future consequences on the innocent third party (i.e., Mary's children), the perceived valence of such a possible consequence determined whether or not it was to be factored in their judgment of responsibility. Positively valued future side effects appeared not to be factored in, and so they did not affect the output of the alleged concrete system, which led to compatibilist judgments of responsibility. On the other hand, negatively valued outcomes appear to be sometimes factored in, and thus are capable of overriding the concrete system's response by allowing participants to reassess their judgments of responsibility.

If this account is on the right track, then the results of [Study 1](#) may be explained in the following manner: positively valued possible future side effects did not affect the performance of the concrete system, and thus judgments of responsibility in line with compatibilism were elicited. However, negatively valued possible side effects did override the operations of the system, allowing the subject to re-focus her attention on other critical features of the vignette—like its MRECD features—prior to making the judgment. Consequently, in the *Worse* condition, participants' judgments may have been switched from compatibilism to incompatibilism either because the pathological component of the description elicited an intense affective reaction (Greene 2007), or because it lessened the impulse to blame (Knobe 2005) or punish (Nadelhoffer 2006) the agent. In [Studies 2](#) and [3](#), however, the perceived negative value of the possible

future consequences was only strong enough to override the normal operation of the concrete system, but none of the MRECD features of the description was sufficient to swing the judgment toward incompatibilism. As such, participants' responses hovered around the mid-point. At any rate, even if the suggested explanation is incomplete or inaccurate, we believe that the results reported here suggest that the behavior of the alleged concrete system for judgments of responsibility is more complex than originally thought. Whether or not similar manipulations may affect the abstract system is an interesting question for future research.

However, it is important to acknowledge a number of interpretative limitations to the results of the studies reported here. First, the way in which participants are asked to register their responses in each vignette may be subject to more than one potentially diverging reading. For instance, although participants were invited to read the question as asking them to judge the degree to which the protagonist is morally responsible, they may have read it instead as asking them to judge whether or not they think it is *appropriate* to hold her responsible for the relevant action. We believe that further research is needed to adjudicate between these two readings in order to clarify the precise nature of the judgment upon which considerations about what may happen in the future actually have an impact.

Another potential difficulty is that participants may have interpreted the question as asking them to judge whether or not the protagonist deserved to be held responsible, or whether or not the protagonist was blameworthy, or even whether or not the protagonist should be punished.<sup>3</sup> Given the many purposes for which people normally employ the concept of moral responsibility, and the variety of situations in which it gets to be deployed, the rather exploratory character of the studies reported here is insufficient to adjudicate between all these potentially divergent readings of the vignettes' questions. Clearly, further research is needed to shed light on these issues. Nonetheless, we believe that the results reported here are interesting in their own right, and that they offer evidence to the effect that, when it comes to studying the folk's concept of moral responsibility, participant's considerations about possible future consequences should not be overlooked.

## References

- Beebe, J., and W. Buckwalter. 2010. The epistemic side-effect effect. *Mind & Language* 25(4): 474–498.
- Björnsson, G., and Persson, K. In press. A unified empirical account of responsibility judgments. *Philosophy and Phenomenological Research*. doi:10.1111/j.1933-1592.2012.00603.x
- De Brigard, F., E. Mandelbaum, and D. Ripley. 2009. Responsibility and the brain sciences. *Ethical Theory and Moral Practice* 12(5): 511–524.
- Feltz, A. 2007. The knobe effect: A brief overview. *Journal of Mind and Behavior* 28: 265–277.
- Fischer, J.M. 1994. *The metaphysics of free will*. Oxford: Blackwell Publishers.
- Greene, J. D. 2007. The secret joke of Kant's soul. In *Moral psychology, Vol. 3: The neuroscience of morality: Emotion, disease, and development*, ed. W. Sinnott-Armstrong. Cambridge, MA: MIT Press.
- Greene, J.D., and J.D. Cohen. 2004. For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society B* 359: 1775–17785. Special Issue on Law and the Brain.

<sup>3</sup> We thank an anonymous reviewer for pointing out these potential confounds.

- Kane, R. 1996. *The significance of free will*. New York: Oxford University Press.
- Knobe, J. 2003. Intentional action and side effects in ordinary language. *Analysis* 63(3): 190–194.
- Knobe, J. 2005. Cognitive processes shaped by the impulse to blame. *Brooklyn Law Review* 71: 929–937.
- Knobe, J., W. Buckwalter, P. Robbins, H. Sarkissian, T. Sommers, and S. Nichols. 2011. Experimental philosophy. *Annual Review of Psychology* 63(50): 72–73.
- Nadelhoffer, T. 2006. Bad acts, blameworthy agents, and intentional actions: Some problems for jury impartiality. *Philosophical Explorations* 9(2): 203–220.
- Nahmias, E.A., S.G. Morris, T. Nadelhoffer, and J. Turner. 2005. Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology* 18(5): 561–584.
- Nahmias, E., S.G. Morris, T. Nadelhoffer, and J. Turner. 2006. Is incompatibilism intuitive? *Philosophy and Phenomenological Research* 73(1): 28–53.
- Nahmias, E., D.J. Coates, and T. Kvaran. 2007. Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest Studies in Philosophy* V. XXXI: 214–242.
- Nichols, S. 2011. Experimental philosophy and the problem of free will. *Science* 331: 1401–1403.
- Nichols, S., and J. Knobe. 2007. Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous* XLI(4): 663–685.
- Pereboom, D. 2001. *Living without free will*. Cambridge: Cambridge University Press.
- Roskies, A., and S. Nichols. 2008. Bringing responsibility down to earth. *Journal of Philosophy* CV(7): 371–388.
- Sarkissian, H., A. Chatterjee, F. De Brigard, J. Knobe, S. Nichols, and S. Sirker. 2010. Is belief in free will a cultural universal? *Mind & Language* 25(3): 346–358.
- Sinnott-Armstrong, W. 2008. Abstract + Concrete = Paradox. In *Experimental philosophy*, ed. J. Knobe and S. Nichols, 209–230. New York: Oxford University Press.
- Woolfolk, R.L., J.M. Doris, and J.M. Darley. 2006. Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition* 100: 283–301.