



ELSEVIER

Contents lists available at ScienceDirect

Cognition

journal homepage: [www.elsevier.com/locate/cognit](http://www.elsevier.com/locate/cognit)

## Original Articles

## Moral hindsight for good actions and the effects of imagined alternatives to reality



Ruth M.J. Byrne\*, Shane Timmons

Trinity College Dublin, University of Dublin, Ireland

## ARTICLE INFO

## Keywords:

Causal inference  
Counterfactuals  
Hindsight  
Moral judgment  
Reasoning

## ABSTRACT

Five experiments identify an asymmetric moral hindsight effect for judgments about whether a morally good action should have been taken, e.g., Ann should run into traffic to save Jill who fell before an oncoming truck. Judgments are increased when the outcome is good (Jill sustained minor bruises), as Experiment 1 shows; but they are not decreased when the outcome is bad (Jill sustained life-threatening injuries), as Experiment 2 shows. The hindsight effect is modified by imagined alternatives to the outcome: judgments are amplified by a counterfactual that if the good action had not been taken, the outcome would have been worse, and diminished by a semi-factual that if the good action had not been taken, the outcome would have been the same. Hindsight modification occurs when the alternative is presented with the outcome, and also when participants have already committed to a judgment based on the outcome, as Experiments 3A and 3B show. The hindsight effect occurs not only for judgments in life-and-death situations but also in other domains such as sports, as Experiment 4 shows. The results are consistent with a causal-inference explanation of moral judgment and go against an aversive-emotion one.

## 1. Introduction

In daily life, we sometimes hear about morally good actions, for example, recent media reports have highlighted the actions of Syria's 'white helmets', the civilian first responders who risk their lives in airstrikes to rescue survivors, and they are regarded as humanitarian heroes. Yet there has been little research on how people reason about such self-sacrificial acts of virtue. Research on morally good actions has focused on emotional responses and it has shown that observers of such actions are often emotionally uplifted, an experience termed 'moral elevation' (e.g., Algoe & Haidt, 2009; Aquino, McFerran, & Laven, 2011; Diessner, Iyer, Smith, & Haidt, 2013; Schnall & Roper, 2012). Observers of morally good actions often wish to emulate them, to engage in prosocial behavior to do something good to improve the welfare of others (e.g., Cox, 2010; Freeman, Aquino, & McFerran, 2009; Schnall, Roper, & Fessler, 2010). But there has been little research to examine the cognitive processes that underlie judgments about morally good actions. We address two issues: first, we test whether judgments that a morally good action should have been taken are affected by outcome knowledge that it succeeded or failed. For example, suppose you hear that a Syrian 'white helmet' risked his life to pull a child from the rubble of a bombed building. Is your judgment that he should have done so affected when you hear that his action succeeded and the

outcome was good, the child lived; or when you hear that he failed and the outcome was bad, the child died? Second, we test whether judgments that a morally good action should have been taken are affected by imagined alternatives to the outcome. For example, suppose you hear that the Syrian 'white helmet' pulled the child from the rubble and the child lived. Is your judgment that he should have pulled the child from the rubble amplified by a counterfactual about how the outcome could have been different, such as, 'if he hadn't rushed to pull the child from the rubble, she would have died'? A counterfactual is an imagined alternative to reality and the example illustrates a downward comparison to a worse world. Is your judgment that he should have pulled the child from the rubble diminished by a 'semi-factual' about how the outcome could have been the same, such as, 'even if he hadn't rushed to pull the child from the rubble, she would have lived anyway'? A semi-factual is an imagined alternative to reality that results in the same outcome as reality. We examine whether judgments that the action should have been taken are affected by these imagined alternatives. We report the results of five experiments to address these questions.

## 1.1. Outcome knowledge

The first question we examine is whether judgments that a morally good action should have been taken are affected by outcome knowledge

\* Corresponding author at: School of Psychology and Institute of Neuroscience, Trinity College Dublin, University of Dublin, Ireland.  
E-mail address: [rmbyrne@tcd.ie](mailto:rmbyrne@tcd.ie) (R.M.J. Byrne).

that it succeeded or failed. We examined judgments about whether Ann should have run into traffic to save Jill who fell before an oncoming truck. We compared judgments when Ann's action succeeded and hence the outcome was good, Jill sustained minor bruises, compared to judgments when Ann's action failed and hence the outcome was bad, Jill sustained life-threatening injuries. Outcome knowledge affects judgments about morally *bad* actions (e.g., Baron & Hershey, 1988; Fleischhut, Meder, & Gigerenzer, 2017; Oeberst & Goekenjan, 2016). Consider two protagonists who threw bricks over a wall from an overpass bridge; one person's brick hit a car and killed the driver and the other person's brick hit the pavement without accomplishing the intended harm. Both protagonists carried out the same action with the same knowledge and the same intention, yet participants' judgments of blame and punishment are harsher for the person whose brick hit a car and killed the driver compared to the one whose brick hit the pavement (e.g., Lench, Domskey, Smallman, & Darbor, 2015; Martin & Cushman, 2016a). But it is a matter of luck whether a car was passing underneath the bridge at the time each one threw a brick (e.g., Cushman, 2008; Young, Nichols, & Saxe, 2010). The effects of outcome knowledge on judgments about morally bad actions may arise either from emotional reactions to the aversive outcome, or from causal inferences about whether the outcome could have been foreseen.

The 'aversive-emotion' explanation suggests that the bad outcome leads people to experience a more negative reaction when the attempt to harm succeeds, a driver was killed by the brick, compared to when it fails, no-one was harmed by the brick. The emotional response evoked by the bad outcome overshadows any consideration of the protagonists' intentions. There has been extensive debate about negativity biases (e.g., Rozin & Royzman, 2001) and the relative contribution of emotional and cognitive processes to moral judgments about bad actions (e.g., Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Haidt, 2001; Mikhail, 2007; Gubbins & Byrne, 2014). The alternative 'causal-inference' explanation suggests that people consider whether the individual could have known what the outcome would be. They judge the brick-thrower who killed a driver harshly because they decide, with hindsight, that he could have known that his bad action of brick-throwing would succeed in harming a person (e.g., Baron & Hershey, 1988; Martin & Cushman, 2016b). People construct a causal model that links the action to its intended and expected outcome (e.g., Cushman, 2013; Timmons & Byrne, 2018). When there is a clear causal link between the action and the outcome, they consider that an individual who carried out the action could have known what the outcome would be. People with outcome knowledge tend to believe they would have predicted the outcome all along, perhaps because of mistaken beliefs about its predictability before it occurred (e.g., Fischhoff, 1975; Roese & Vohs, 2012). Hence people allocate blame and responsibility to the brick-thrower whose brick kills a driver. They judge the lucky brick-thrower who did not harm anyone leniently because they decide, with hindsight, that he would have known that his bad action of brick-throwing would fail to harm anyone. The aversive-emotion and causal-inference explanations make similar predictions for outcomes from morally bad actions, but they make different predictions for outcomes from morally good actions.

We test the two explanations for judgments about whether an agent should have carried out a morally *good* action, for example, judgments about whether Ann should have run into traffic to save Jill who fell before an oncoming truck. We examined judgments when Ann's action succeeded and hence the outcome was good, Jill sustained only minor bruises, and judgments when Ann's action failed and hence the outcome was bad, Jill sustained life-threatening injuries. The two explanations make different predictions, as Table 1 shows. The aversive-emotion hypothesis predicts an effect of outcome knowledge for morally good actions that fail and result in a *bad* outcome, such as when Jill sustains life-threatening injuries. People will judge that the morally good action should have been taken less often when they know it failed, compared to when they do not know the outcome, because they experience a

**Table 1**

Predictions of two competing hypotheses for judgments that a morally good action should have been taken, after information is provided about the outcome or imagined alternatives (relative to judgments made prior to receiving such information).

Hypothesis	Aversive-emotion	Causal-inference
Outcome knowledge		
Good outcome		
<i>Jill sustains minor bruises</i>	No effect	Amplify judgments
Bad outcome		
<i>Jill sustains serious injuries</i>	Diminish judgments	No effect
Imagined alternatives		
Counterfactual, Good outcome		
<i>Jill's (minor) injuries would have been worse</i>	Amplify judgments	Amplify judgments
Counterfactual, Bad outcome		
<i>Jill's (serious) injuries would have been worse</i>	Amplify judgments	Amplify judgments
Semi-factual, Good outcome		
<i>Jill's (minor) injuries would have been the same</i>	No effect	Diminish judgments
Semi-factual, Bad outcome		
<i>Jill's (serious) injuries would have been the same</i>	No effect	Diminish judgments

negative reaction to the bad outcome. The aversive-emotion hypothesis predicts no effect of outcome knowledge for morally good actions that succeed and result in a good outcome, such as when Jill sustains minor bruises. People will judge that the morally good action should have been taken as often when they know it succeeded, compared to when they do not know the outcome, because the aversive-emotion hypothesis is based on negative emotional reactions. The focus on negative emotions rather than positive ones can be considered akin to the proposal that people weight losses more than gains (e.g., Kahneman, 2011; Tversky & Kahneman, 1981).

The predictions of the causal-inference hypothesis are different. It predicts a moral hindsight effect for *good* outcomes, because the good outcome provides confirmation of its causal link to the action, Ann rushed to save Jill and her action resulted in the outcome that Jill was saved. People can construct a causal model in which the action is clearly linked to the outcome. They decide, with hindsight, that Ann could have known that her action would save Jill. People will judge that the morally good action should have been taken more often when they know it succeeded, compared to when they do not know the outcome. The causal-inference hypothesis predicts no effect of outcome knowledge for morally good actions that fail and result in a bad outcome. Participants attribute a high degree of responsibility for a bad outcome to a protagonist despite their good intentions (e.g., Kominsky, Phillips, Gerstenberg, Lagnado, & Knobe, 2015; Sarin, Lagnado, & Burgess, 2017). A bad outcome may provide disconfirmation of the causal link between the good action and its expected good outcome, Ann rushed to save Jill but her action did not result in the outcome that Jill was saved, although people have difficulty reasoning about disconfirmation (e.g., Cherubini, Castelvecchio, & Cherubini, 2005). But a bad outcome may not disconfirm the causal link between the good action and the expected good outcome, and instead may indicate that other disabling factors intervened to prevent the outcome, for example Ann rushed to save Jill and held her down as the truck drove over them and her action would have succeeded but for other factors, such as attempts by Jill to escape, last-minute manoeuvres by the truck-driver, and so on (e.g., Byrne, 1989; Oaksford & Chater, 2017). Hence, people will not decide, with hindsight, that Ann could have known that her action would not save Jill. They will judge that the morally good action should have been taken as often when they know it failed, compared to when they do not know the outcome. We test whether a moral hindsight effect occurs for good actions when they succeed and lead to good outcomes (in

Experiment 1) and when they fail and lead to bad outcomes (in Experiment 2).

### 1.2. Imagined alternatives

The second question we address is whether judgments that a morally good action should have been taken are affected by imagined counterfactual and semi-factual alternatives to the outcome. We examined judgments about whether Ann should have run into traffic to save Jill who fell before an oncoming truck, and we compared judgments when participants heard a counterfactual about how the outcome could have been worse, ‘if Ann had not rushed to help, Jill’s injuries would have been worse’ compared to judgments when they heard a semi-factual about how the outcome could have been the same, ‘even if Ann had not rushed to help, Jill’s injuries would have been the same’. An imagined alternative to an outcome affects judgments about a morally *bad* action, such as throwing a brick from an overpass bridge. For example, participants’ condemnation of the ‘lucky’ brick-thrower who failed to harm is amplified when they are told a counterfactual about how things would have turned out worse, e.g., if he had thrown the brick from another section of the bridge he would have hit a car (e.g., Parkinson & Byrne, 2017a; see also Lench et al., 2015). When participants hear the counterfactual, they judge the brick thrower who did not harm anyone just as harshly as they judge the one who killed a driver. Counterfactuals amplify moral judgments about many sorts of bad outcomes (e.g., Branscombe, Owen, Garstka, & Coleman, 1996; see also Alicke, Buckingham, Zell, & Davis, 2008; Markman, Mizoguchi, & McMullen, 2008). Notably however, participants’ condemnation of the failed attempt to harm is not *diminished* when they are told a semi-factual alternative about how things would have turned out the same, e.g., even if he had thrown the brick from another section of the bridge he would not have hit a car (e.g., Parkinson & Byrne, 2017a). When participants hear the semi-factual, their judgments of the brick thrower who did not harm anyone are lenient, but no more so than when they have not heard any imagined alternative.

The effects are consistent with the aversive-emotion hypothesis: the imagined alternative re-allocates participants’ focus onto a different outcome – a counterfactual about how things could have been worse shifts the focus onto a worse outcome, the brick could have hit a car, and so it amplifies judgments of punishment and blame; a semi-factual places the focus on the same outcome, the brick did not hit a car, and so it does not affect judgments. The effects appear at first sight to be *inconsistent* with the causal-inference hypothesis: a counterfactual such as, if he had thrown the brick from another section of the bridge he would have hit a car, amplifies judgments because it makes explicit the causal relation between the action and its outcome, that throwing a brick over the bridge will result in a car being hit. A semi-factual such as, even if he had thrown the brick from another section of the bridge he would not have hit a car, should *diminish* judgments, because it denies the causal relation between the action and its outcome, that throwing a brick over the bridge will result in a car being hit (e.g., Byrne, 2016; Lagnado, Gerstenberg, & Zultan, 2013; McEleney & Byrne, 2006; Rasga, Quelhas, & Byrne, 2016; Roese & Epstude, 2017; Walsh & Byrne, 2007). Counterfactuals and semi-factuals have clear effects on causal judgments. For example, people judged that a painkiller caused an athlete to lose a race due to side effects of fatigue; their causal judgments were *amplified* when they knew about a counterfactual alternative, a painkiller with no side effects, and they were *diminished* when they knew about a semi-factual alternative, a painkiller with the same side effects (e.g., McCloy & Byrne, 2002). Hence if judgments about whether a protagonist should have taken an action depend on causal inferences about whether the protagonist could have known what the outcome would be, counterfactuals should amplify judgments that the protagonist should have taken the action, and semi-factuals should diminish them. The observation of a counterfactual amplification effect for judgments of morally bad actions, e.g., about the lucky

brick thrower, supports the causal-inference hypothesis, but the absence of a semi-factual diminishment effect for judgments of morally bad actions appears to go against the causal-inference hypothesis. However, it is important to note that the absence of a semi-factual diminishment effect for morally bad actions may merely reflect a floor effect: judgments are already lenient for the lucky brick-thrower’s failed attempt to harm and participants may be unable or unwilling to be any more lenient when they hear that he would have failed to harm even if he had thrown his brick from another part of the bridge.

A stronger test of the competing hypotheses is provided by imagined alternatives to the outcomes of morally *good* actions, because a floor effect for semi-factuals is ruled out. On the causal-inference hypothesis, judgments that a morally good action should have been taken will be amplified by a counterfactual, e.g., if Ann had not rushed into the traffic, Jill’s injuries would have been worse. The counterfactual makes a downward comparison to a worse world and emphasizes the causal link between the action and its outcome. The counterfactual amplification should occur whether the actual outcome is good, Jill sustains minor bumps and bruises, and the counterfactual implies her injuries would have been worse in that she could have been seriously injured; or whether the actual outcome is bad, Jill sustains serious life-threatening injuries, and the counterfactual implies her injuries would have been worse in that she could have been killed. Importantly, judgments that a morally good action should have been taken will be *diminished* by a semi-factual, e.g., even if Ann had not rushed into the traffic, Jill’s injuries would have been the same. The semi-factual makes a comparison to a world in which the outcome is the same, it indicates that the outcome was over-determined and denies a causal link between the action and its outcome (e.g., McCloy & Byrne, 2002; Branscombe et al., 1996). Since judgments that the morally good action should have been taken can be expected to be high when no imagined alternative is asserted, the provision of a semi-factual should not encounter the floor effect conjectured to occur for judgments about morally bad actions. The semi-factual diminishment should occur whether the actual outcome is good, Jill sustains minor bumps and bruises, and the semi-factual implies her injuries would have been the same (still minor) or whether the actual outcome is bad, Jill sustains serious life-threatening injuries, and the semi-factual implies her injuries would have been the same (still serious).

The aversive-emotion hypothesis makes a different prediction: the counterfactual, if Ann had not rushed into the traffic, Jill’s injuries would have been worse, will amplify judgments that the action should have been taken compared to no imagined alternative because it shifts the focus to a worse outcome, Jill would have had worse injuries. Hence the negative emotion evoked by the imagined worse outcome and the desire to avoid it will result in the amplification of judgments that the action should have been taken. The same prediction applies whether the actual outcome was good, Jill sustained minor injuries, and so the negative emotion is evoked by the imagined bad outcome of serious injuries, or whether the actual outcome was bad, Jill sustained serious injuries, and the negative emotion is evoked by the imagined bad outcome of death. The semi-factual, even if Ann had not rushed into the traffic, Jill’s injuries would have been the same, will not affect judgments compared to no imagined alternative because it places the focus on the same actual outcome. The emotion an individual experiences when they hear the good outcome, Jill sustains minor injuries, will not be altered when they hear the semi-factual ‘even if Ann had not rushed into the traffic, Jill’s injuries would have been the same’ because they will think about the same good outcome, Jill’s minor injuries. Similarly the emotion an individual experiences when they hear the bad outcome, Jill sustains serious injuries, will not be altered when they hear the semi-factual ‘even if Ann had not rushed into the traffic, Jill’s injuries would have been the same’ because they will think about the same bad outcome, Jill’s serious injuries. We report five experiments to pit these alternative views against each other.

## 2. Experiment 1. Good outcomes

The aim of the experiment was to examine whether there is a moral hindsight effect of outcomes for moral judgments about good actions, that is, whether people tend to judge more often that a morally good action should have been taken when they know it succeeded and led to a good outcome, compared to when they do not know the outcome. The causal-inference hypothesis predicts that judgments will be increased when participants know that the good action led to a good outcome, compared to when they do not know the outcome, because the good outcome provides confirmation of its causal link to the action. The aversive-emotion hypothesis predicts that judgments will not be increased because its predictions of differences are based on negative emotional reactions only.

We also examine whether there are counterfactual amplification and semi-factual diminishment effects, for successful attempts to help that lead to a good outcome. We examine whether judgments that a morally good action should have been taken are increased when people hear a counterfactual that if the action had not been taken the outcome would have been worse; and whether they are decreased when people hear a semi-factual that even if the action had not been taken the outcome would have been same. The causal-inference hypothesis predicts that judgments will be amplified by a counterfactual because it emphasizes the causal link between the action and its outcome, and it predicts that judgments will be diminished by a semi-factual because it denies a causal link between the action and its outcome. The aversive-emotion hypothesis predicts that judgments will be amplified by the counterfactual because it shifts the focus to a worse outcome that evokes a negative emotion, and it predicts that judgments will not be affected by the semi-factual because it places the focus on the same actual outcome.

### 2.1. Method

#### 2.1.1. Participants and procedure

The participants in Experiments 1, 2, 3A and 3B were recruited on the online platform Prolific, [www.prolific.ac](http://www.prolific.ac), restricted to English speaking countries and paid 50 pence sterling for their participation. Participants were eliminated prior to any data analysis if they failed to answer correctly two trap questions commonly included in online studies (select the names of the people in the story from a set of 5 names, and select option '1' on a 1–7 scale to confirm that you are paying attention). For all studies, we report all our manipulations and measures, and each study's sample size was determined prior to data collection. We planned a sample size of 50 participants per condition to ensure sufficient power to observe a small to medium effect and set the pre-programmed recruitment stopping rule to recruit until 170 participants had been assigned by the program to the three conditions (on the assumption that some participants would fail the attention checks). The materials were presented using Surveygizmo software, [www.surveygizmo.com](http://www.surveygizmo.com). For all studies, informed consent was obtained from participants and approval was obtained from Trinity College Dublin's Psychology Research Ethics Committee.

The participants in Experiment 1 were 154 volunteers from the general public, 69 women and 85 men, with an age range of 18–72 years ( $M = 33.2$ ,  $SD = 11.8$ ), and a further 17 participants were eliminated prior to any data analysis because they failed to answer the trap questions correctly. Participants were assigned at random to the counterfactual ( $n = 53$ ), semi-factual ( $n = 42$ ), and facts ( $n = 59$ ) conditions.

#### 2.1.2. Materials and design

Each participant was presented with a story about a good action, and given a story ending that contained a good outcome, and a counterfactual, a semi-factual or a re-statement of the facts, as Table 2 shows. Participants completed a pre-outcome judgment that the action

should be taken and a post-outcome judgment that it should have been taken, on a 7 point scale with 1 labeled 'agree' and 7 'do not agree'. The design was mixed, with the between-participants factor of the story ending, and the within-participants factor of the judgments made pre- or post- outcome information.

In each of the experiments, as a baseline materials check we also measured (a) moral elevation and helping behavior, that is, whether people felt uplifted by the story, and whether they wished to do something good, and (b) judgments that the action was morally good (immediately after the judgments that it should be taken). The manipulations did not affect these measures (for details of these tasks and their results see the [Supplemental Materials](#)).

### 2.2. Results and discussion

The data for each of the experiments are available in the [Dataset Supplemental Materials](#). A 3 (story ending: counterfactual, semi-factual, facts)  $\times$  2 (judgment: pre-outcome, post-outcome) ANOVA with repeated measures on the second factor on judgments that the action should have been taken showed a main effect of judgment,  $F(1, 151) = 15.597$ ,  $p < .001$ ,  $\eta_p^2 = .094$ , a main effect of story ending,  $F(2, 151) = 6.914$ ,  $p < .001$ ,  $\eta_p^2 = .084$ , and an interaction of the two,  $F(2, 151) = 18.64$ ,  $p < .001$ ,  $\eta_p^2 = .198$ , as Fig. 1 shows.

The decomposition of the interaction, with a Bonferroni adjusted alpha of  $p < .0056$  for 9 comparisons, showed that participants' post-outcome judgments were different from their pre-outcome ones in each of the three conditions. After they heard the outcome and a statement about the facts they were more inclined to agree that the action should have been taken, compared to their judgments before,  $t(58) = 4.743$ ,  $p < .001$ , 95% confidence interval of difference of means (CI) [.51, 1.25],  $d = 0.623$ , indicating a moral hindsight effect for good outcomes from good actions. Their post-outcome judgments were amplified by the counterfactual,  $t(52) = 5.198$ ,  $p < .001$ , 95% CI [.74, 1.67],  $d = 0.715$ , and diminished by the semi-factual,  $t(41) = 2.804$ ,  $p < .008$ , 95% CI [−1.06, −.17],  $d = 0.449$ . Hence a comparison of their post-outcome judgments showed that judgments were diminished for the semi-factual compared to the counterfactual,  $t(93) = 4.613$ ,  $p < .001$ , 95% CI [−2.63, −1.05],  $d = 0.957$ , and the facts,  $t(99) = 5.08$ ,  $p < .001$ , 95% CI [1.17, 2.68],  $d = 1.02$ , indicating a semi-factual diminishment effect; there was no difference between the counterfactual and facts,  $t < 1$  (which may reflect a ceiling effect for counterfactual amplification of judgments of good actions – a mirror-image of the floor effect for semi-factual diminishment of judgments of bad actions in previous studies). A baseline check comparison of their pre-outcome judgments showed that they were equivalent in the three conditions: counterfactual vs. semi-factual,  $t < 1$ , counterfactual vs. facts,  $t(110) = 1.265$ ,  $p = .209$  and semi-factual vs. facts,  $t(99) = 1.169$ ,  $p = .245$ .

The results are consistent with the causal-inference hypothesis, but we examine good actions that fail to help and lead to a bad outcome before we consider their implications for the competing hypotheses further.

## 3. Experiment 2. Bad outcomes

The aim of the experiment was to examine whether the moral hindsight, counterfactual amplification and semi-factual diminishment effects occur for moral judgments about failed attempts to help that lead to a bad outcome. The aim was to examine whether people tend to judge less often that a morally good action should have been taken when they know it failed and led to a bad outcome, compared to when they do not know the outcome. The aversive-emotion hypothesis predicts that judgments will be reduced when participants know that the good action led to a bad outcome, compared to when they do not know the outcome, because they will experience a negative reaction to the bad outcome. The causal-inference hypothesis predicts no effect on

**Table 2**  
Materials and judgment tasks in Experiment 1 (information in italics is provided for clarification here).

<i>Story:</i>	Ann is waiting for a bus with her son. Nearby a young child, Jill trips over an uneven pavement stone and stumbles into the road and falls down near the traffic lights at the busy intersection. Just then, a very large truck drives through the intersection. Ann rushes to help, she can see that the truck driver cannot see Jill struggling on the ground and will not be able to stop in time. She decides that the only way to help is to run into the road and hold Jill down so they both lie still as the truck passes over them. She knows that there is a terrible risk that the truck could crush them both.
<i>Pre-outcome judgment:</i>	Ann should run into the road to help Jill.
<i>Outcome:</i>	Ann ran into the road and held Jill down. When the truck drove on and passersby got to them, Jill had sustained only minor bumps and bruises from the truck.
<i>Story Endings:</i>	
<i>Facts:</i>	The police examined the videos of the incident later and observed Ann rushing to help and Jill sustaining injuries from the truck.
<i>Counterfactual:</i>	The police examined the videos of the incident later and observed that if Ann had not rushed to help, Jill’s injuries would have been worse.
<i>Semi-factual:</i>	The police examined the videos of the incident later and observed that even if Ann had not rushed to help, Jill’s injuries would have been the same.
<i>Post-outcome judgment:</i>	Knowing what I know now, I believe Ann should have run into the road to help Jill.

judgments, because the bad outcome does not provide disconfirmation of the causal link of the good action to a good outcome.

We also examine whether there is a counterfactual amplification effect for failed attempts to help that lead to a bad outcome. We test whether people tend to judge more often that a morally good action that led to a bad outcome should have been taken when they hear a counterfactual that if the action had not been taken the outcome would have been worse. Note that a counterfactual such as, if Ann had not rushed into the traffic, Jill’s injuries would have been worse, can be uttered sensibly not only when Ann’s action succeeded and Jill sustained only bumps and bruises, but also even when Ann’s action did not succeed and Jill sustained serious injuries. Similarly we examined whether there is a semi-factual diminishment effect for failed attempts to help that lead to a bad outcome. We examined whether people tend to judge less often that a morally good action that led to a bad outcome should have been taken when they hear a semi-factual that if the action had not been taken the outcome would have been the same. A semi-factual such as, even if Ann had not rushed into the traffic, Jill’s injuries would have been the same, can also be uttered sensibly not only when Ann’s action succeeded and Jill sustained only bumps and bruises, but also even when Ann’s action did not succeed and Jill sustained serious injuries. The predictions of the aversive-emotion and causal-inference hypotheses for the effects of counterfactuals and semi-factuals remain the same regardless of whether the outcome is good or bad (see Table 1). The causal-inference hypothesis predicts that judgments will be amplified by a counterfactual because it emphasizes the causal link between the action and its outcome, and it predicts that judgments will be diminished by a semi-factual because it denies a causal link between

the action and its outcome. The aversive-emotion hypothesis predicts that judgments will be amplified by the counterfactual because it shifts the focus to a worse outcome that evokes a negative emotion, and it predicts that judgments will not be affected by the semi-factual because it places the focus on the same actual outcome.

### 3.1. Method

#### 3.1.1. Participants and procedure

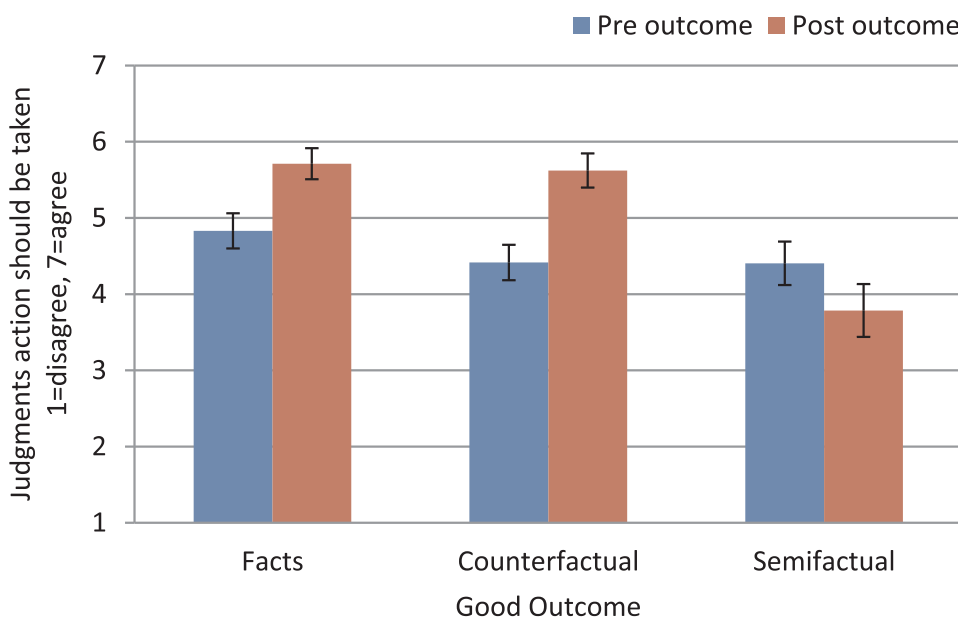
The participants were a new set of 145 volunteers recruited on Prolific, 71 women and 74 men, with an age range of 18–67 years ( $M = 31.32$ ,  $SD = 11.04$ ). A further 17 participants were eliminated prior to any data analysis because they failed the trap questions. Participants were assigned at random to the counterfactual ( $n = 66$ ), semi-factual ( $n = 36$ ), and facts ( $n = 43$ ) conditions and the procedure was the same as the previous experiment.

#### 3.1.2. Materials and design

The same materials were used as the previous experiment except that the outcome was bad, ‘When the truck drove on and passersby got to them, Jill had sustained serious life threatening injuries from the truck’. Participants made the same judgments as the previous experiment and the design of the experiment was the same.

### 3.2. Results and discussion

An ANOVA of the same design as the previous experiment on judgments that the action should have been taken showed no main



**Fig. 1.** Judgments that the action should have been taken, made before and after information about the outcome and a statement of the facts, counterfactual, or semi-factual in Experiment 1 for good outcomes. Responses are reverse scored so that higher bars indicate higher agreement that the action should have been taken. Error bars are standard error of the mean.

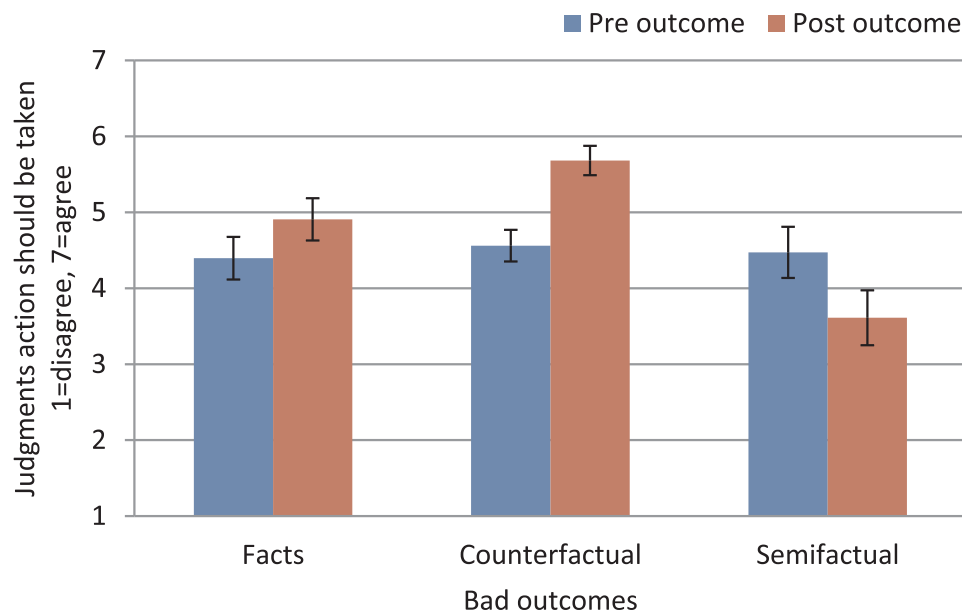


Fig. 2. Judgments that the action should have been taken made before and after information about the outcome and a statement of the facts, counterfactual, or semi-factual in Experiment 2 for bad outcomes. Responses are reverse scored. Error bars are standard error of the mean.

effect of judgment,  $F(1, 142) = 3.131, p = .079$ , a main effect of story ending,  $F(2, 142) = 5.313, p < .006, \eta_p^2 = .07$ , and an interaction of the two,  $F(2, 145) = 15.969, p < .001, \eta_p^2 = .184$ , as Fig. 2 shows.

The decomposition of the interaction, with a Bonferroni alpha of  $p < .0056$ , showed no differences in participants' judgments before and after they heard the bad outcome and a statement about the facts,  $t(42) = 1.759, p = .086$ , indicating no moral hindsight effect for bad outcomes from good actions. Judgments that the action should be taken were amplified by the counterfactual,  $t(65) = 5.179, p < .001, 95\% \text{ CI } [.69, 1.55], d = 0.639$ , and diminished by the semi-factual,  $t(35) = 4.148, p < .001, 95\% \text{ CI } [-1.28, -.44], d = 0.695$ . Hence, a comparison of post-outcome judgments showed they were diminished for the semi-factual compared to the counterfactual,  $t(100) = 5.545, p < .001, 95\% \text{ CI } [-2.81, -1.33], d = 1.11$ , and the facts,  $t(77) = 2.886, p < .005, 95\% \text{ CI } [.40, 2.19], d = 0.658$ ; there was no significant difference between the counterfactual and facts on the adjusted alpha,  $t(107) = 2.361, p < .02$ , consistent with Experiment 1. A baseline check comparison of pre-outcome judgments showed they were equivalent in the three conditions,  $t < 1$  in every case.

Given that participants in the two experiments were drawn from the same Prolific population, we compared post-outcome judgments in the facts conditions of Experiment 1 and 2: when participants knew the outcome, they judged that the action should have been taken more often for good outcomes than bad ones,  $M = 2.2881$  vs.  $M = 3.093, t(100) = 2.392, p < .02, 95\% \text{ CI } [-1.47, -.137], d = 0.478$ .

The discovery of an asymmetric moral hindsight effect, for good outcomes and not for bad ones, is consistent with the causal-inference hypothesis. It predicts a moral hindsight effect for morally good actions that succeed and result in a good outcome, because the good outcome provides confirmation of its causal link to the action. People decide, with hindsight, that Ann could have known that her action would save Jill. It predicts no effect of outcome knowledge for morally good actions that fail and result in a bad outcome because the bad outcome does not necessarily disconfirm the causal link between the action and the expected good outcome, it may instead indicate that other factors intervened to prevent the outcome. Hence, people will not decide, with hindsight, that Ann could have known that her action would not save Jill.

The results go against the aversive-emotion hypothesis. It predicts that people will judge that the morally good action should have been

taken less often when they know it failed and resulted in a bad outcome, e.g., Jill sustains life-threatening injuries, compared to when they do not know the outcome, because they experience a negative reaction to the bad outcome. But the results show no such reduction in judgments. It also predicts that people will judge that the morally good action should have been taken as often when they know it succeeded, compared to when they do not know the outcome, because the hypothesis focuses only on negative emotions. But the results show that judgments are increased in this case.

The discovery of a semi-factual diminishment effect, as well as a counterfactual amplification effect, is also consistent with the causal-inference hypothesis. The semi-factual denies a causal link between the action and its outcome, and so judgments are reduced, whereas the counterfactual emphasizes the causal link between the action and its outcome, and so judgments are increased. The discovery of a semi-factual diminishment effect goes against the aversive-emotion hypothesis: although it predicts that counterfactuals will amplify judgments because they shift the focus to a worse outcome, it also predicts that semi-factuals will not diminish judgments compared to no imagined alternative because they place the focus on the same actual outcome.

The results also rule out another alternative explanation based on costs and benefits. Participants could understand the modal auxiliary 'should' to indicate they must make a judgment based on the idea that the protagonist should incur the cost of a given action only when that action is the best one available to the protagonist that results in a benefit to the other person. This account is distinct from the causal-inference hypothesis in that it does not attempt to consider the protagonist's perspective and what they could have anticipated, but instead proposes that the participant makes a judgment based on what is, after the fact, the best course of action that the protagonist could have taken, from a 'god's eye view'. The counterfactual information may change whether or not participants consider that the action the protagonist carried out was actually the best option that they had, and the semi-factual information may lead them to consider that the protagonist incurred a cost and yet the outcome turned out the same as it would have without the cost. However, on this account, judgments that the action should have been taken should be decreased when the action fails and the outcome is bad, since the bad outcome indicates that there was no beneficial outcome even though the protagonist incurred a cost. The finding that judgments that the action should have been taken are

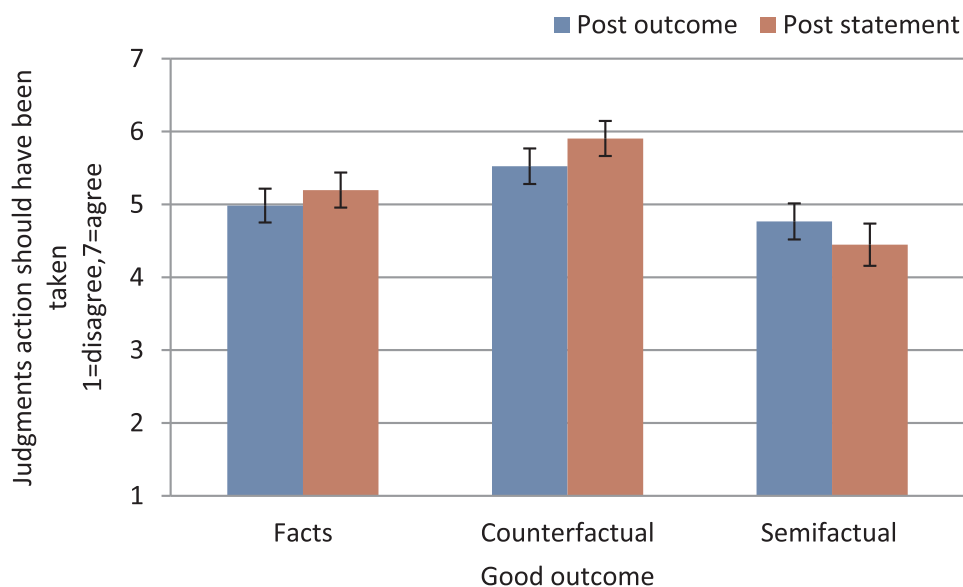


Fig. 3. : Judgments that the action should have been taken made after information about the outcome and after a story ending statement of the facts, counterfactual, or semi-factual in Experiment 3A for good outcomes. Responses are reverse scored. Error bars are standard error of the mean.

not decreased when the outcome is bad go against this costs-and-benefits view.

#### 4. Experiment 3A and 3B: Imagined alternatives only

In the previous experiments the outcome and imagined alternatives were presented together. For example, in Experiment 1, in which the good action led to a good outcome, participants were told in the counterfactual condition,

‘Ann ran into the road and held Jill down. When the truck drove on and passersby got to them, Jill had sustained only minor bumps and bruises from the truck. The police examined the videos of the incident later and observed that if Ann had not rushed to help, Jill’s injuries would have been worse.’

A more stringent test of the causal inference hypothesis is to examine whether participants’ judgments are amplified by counterfactuals and diminished by semi-factuals, even when participants have already committed to a judgment based on information about the outcome. In the next experiments, participants were told the outcome, for example in the good outcome condition, they were told,

‘Ann ran into the road and held Jill down. When the truck drove on and passersby got to them, Jill had sustained only minor bumps and bruises from the truck.’

And they were asked to make a post-outcome judgment,

‘Ann should have run into the road to help Jill.’

Then they were given the imagined alternative separately, for example, in the counterfactual condition they were told,

‘The police examined the videos of the incident later and observed that if Ann had not rushed to help, Jill’s injuries would have been worse.’

And they were asked to make a post-story ending judgment,

‘Knowing what I know now, I believe Ann should have run into the road to help Jill.’

We test good outcomes in Experiment 3A and bad ones in Experiment 3B. We expect to replicate the results of Experiment 1 and Experiment 2, even when the outcome and imagined alternatives are

presented separately and participants have committed to a judgment based on information about the outcome before they have heard the imagined alternative.

#### 4.1. Method

##### 4.1.1. Participants and procedure

In Experiment 3A the participants were a new set of 150 volunteers from Prolific, 76 women and 74 men, with an age range of 18–69 years ( $M = 33.51$ ,  $SD = 11.39$ ). A further 21 participants were eliminated prior to any data analysis because they failed the trap questions. In Experiment 3B the participants were a new set of 151 volunteers, 94 women, 56 men and 1 non-binary person, with an age range of 18–67 years ( $M = 36.78$ ,  $SD = 12.05$ ). Once again a further 21 participants were eliminated because they failed the trap questions. The procedure was the same as the previous experiments.

##### 4.1.2. Materials and design

The materials were the same as the previous experiments except that participants were given the outcome before they made their first judgment (and so all of their judgments were post-outcome judgments), they were then given the counterfactual, semi-factual, or facts story ending and they made their second judgment (see Table S1 in the Supplemental Materials). In Experiment 3A the outcome was good and in Experiment 3B it was bad. The design was the same as the previous experiments.

#### 4.2. Results and discussion

In Experiment 3A, an ANOVA of the same design as the previous experiments on judgments that the action should have been taken showed no main effect of judgment,  $F < 1$ , a main effect of story ending,  $F(2, 147) = 4.953$ ,  $p < .008$ ,  $\eta_p^2 = .063$ , and an interaction of the two,  $F(2, 147) = 3.94$ ,  $p < .022$ ,  $\eta_p^2 = .051$ . The decomposition of the interaction, with a Bonferroni alpha of  $p < .0056$ , showed only one significant difference on the adjusted alpha, that post story-ending judgments were amplified for the counterfactual compared to the semi-factual,  $t(87) = 3.819$ ,  $p < .001$ , 95% CI  $[-2.22, -0.699]$ ,  $d = 0.818$ , as Fig. 3 shows.

In Experiment 3B an ANOVA of the same design on judgments that the action should have been taken showed no main effect of judgment,

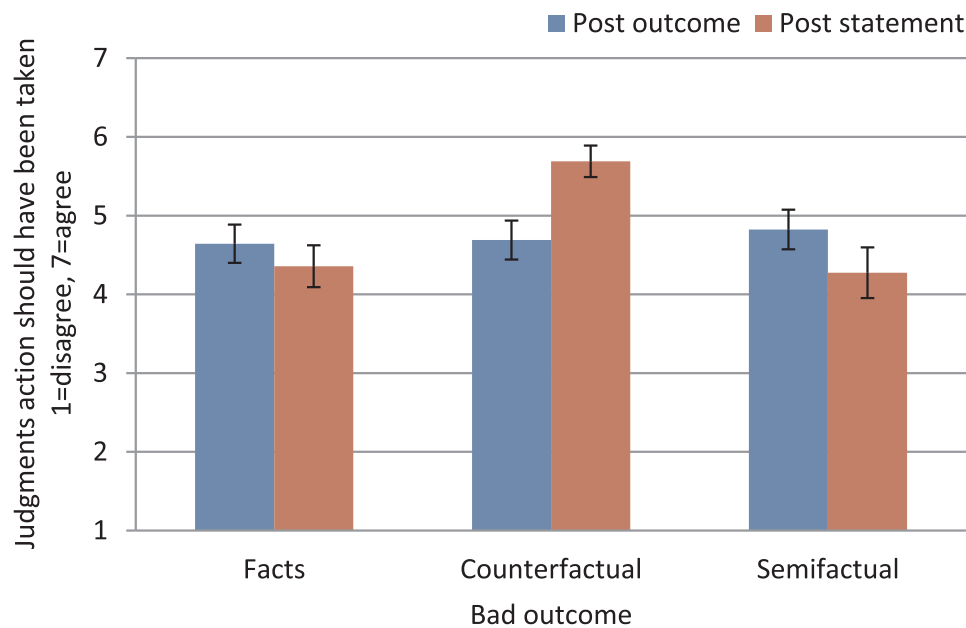


Fig. 4. Judgments that the action should have been taken made after information about the outcome and after a story ending statement of the facts, counterfactual, or semi-factual in Experiment 3B for bad outcomes. Responses are reverse scored. Error bars are standard error of the mean.

$F < 1$ , or story ending,  $F(2, 148) = 2.781, p = .065$ , but a significant interaction of the two,  $F(2, 148) = 18.972, p < .001, \eta_p^2 = .204$ . The decomposition of the interaction, with a Bonferroni alpha of  $p < .0056$ , again showed that post story-ending judgments were amplified for the counterfactual compared to the semi-factual,  $t(107) = 3.829, p < .001, 95\% \text{ CI} [-2.15, -.68], d = 0.74$ ; and they were amplified for the counterfactual compared to the facts,  $t(98) = 4.085, p < .001, 95\% \text{ CI} [-1.98, -.685], d = 0.825$ ; post-story ending judgments were also amplified compared to post-outcome ones for the counterfactual condition,  $t(57) = 5.385, p < .001, 95\% \text{ CI} [.628, 1.37], d = 0.727$ . There were no other significant differences on the adjusted alpha, as Fig. 4 shows.<sup>1</sup>

The results confirm that counterfactuals amplify and semi-factuals diminish judgments that an action should have been taken even when they are presented separately from information about the outcome, for both successful and failed attempts to help. The semi-factual diminishment effect corroborates the predictions of the causal-inference hypothesis and goes against those of the aversive-emotion one.

## 5. Experiment 4

The aim was to replicate the asymmetric moral hindsight effect in a direct comparison of good and bad outcomes, in a laboratory based study, with a between-participants design for outcome information, extended to other materials. We used four stories, two transport accidents similar to the previous experiments, and two isomorphs based on less dramatic sports incidences, in which either a good or bad outcome could be expected or viewed as likely.

### 5.1. Method

#### 5.1.1. Participants and procedure

The participants were 45 volunteers who were psychology students from Trinity College Dublin, 34 women and 10 men, and one participant who chose not to provide information on age or gender, with an age range of 18 to 42 years ( $M = 22.8, SD = 5.27$ ). One further

<sup>1</sup> Given the gender imbalance in this experiment we also checked that gender had no influence on post-outcome or post-story ending judgments,  $t < 1$  in both cases.

participant was eliminated because of failure to complete the items. They were tested in small groups and the materials were presented in a booklet with each story and each judgment on a separate page.

#### 5.1.2. Materials and design

Participants read four stories based on transports and sports contents (see the Supplemental materials). They received two good outcome and two bad outcome versions of each sort of content, presented in a different randomized order to each participant. We created two sets that assigned the contents to the good and bad outcome versions of the stories in two different ways to control for content. The pre-outcome group ( $n = 24$ ) read the story and made a judgment before they heard the outcome and another one after they heard the outcome. The post-outcome group ( $n = 21$ ) read the story and its outcome and then made a single judgment. The design was mixed, with a within-participants variable of good versus bad outcome and a between-participants variable of pre- versus post-outcome judgments.

### 5.2. Results and discussion

A 2 (outcome: good vs bad)  $\times$  2 (group: pre vs post outcome group) ANOVA with repeated measures on the first factor on judgments that the action should have been taken, made by participants after they knew the outcome, showed a main effect of outcome,  $F(1, 43) = 27.29, p < .001, \eta_p^2 = .388$ , as participants judged the action should have been taken more when the outcome was good than when it was bad. There was no main effect of group,  $F < 1$ , and no interaction,  $F(1, 43) = 1.388, p = .245$ , as Fig. 5 shows.

A second analysis on judgments made by the group who made judgments before and after knowing the outcome, which was a 2 (outcome: good vs bad)  $\times$  2 (judgment: pre vs post outcome judgment) repeated measures ANOVA, showed no main effects of outcome,  $F(1, 23) = 3.381, p = .079$ , or judgment,  $F(1, 23) = 4.057, p = .056$ , but an interaction of the two,  $F(1, 23) = 8.972, p < .006, \eta_p^2 = .28$ . The decomposition of the interaction with four comparisons using a Bonferroni adjusted alpha of  $p < .0125$  showed that participants judged more often that the action should have been taken for good outcomes than bad ones when they knew the outcome,  $t(23) = 2.73, p < .012, 95\% \text{ CI} [-1.67, -.23], d = 0.674$ , there was no difference when they did not know the outcome,  $t < 1$ ; and more often when they



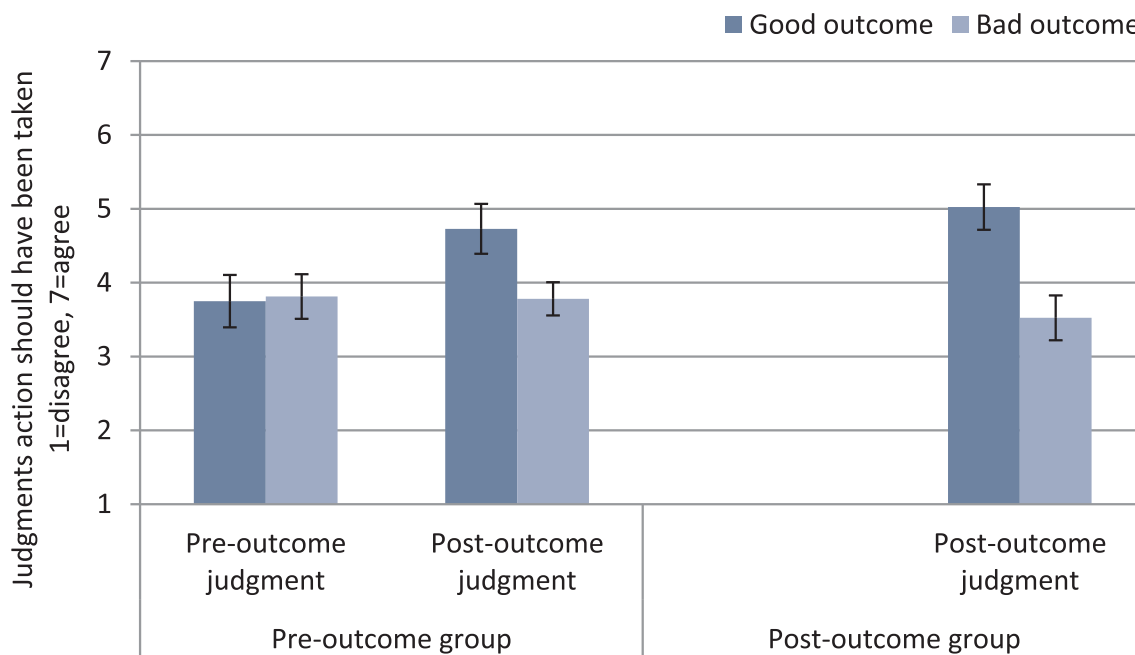


Fig. 5. Judgments that the action should have been taken before and after outcome information in the pre-outcome condition and after outcome information in the post-outcome condition for good or bad outcomes in Experiment 4. Responses are reverse scored. Error bars are standard error of the mean.

knew the outcome than when they did not for good outcomes,  $t(23) = 3.75$ ,  $p < .001$ , 95% CI [0.44, 1.52],  $d = 0.578$ , there was no difference for bad outcomes,  $t < 1$ .

The results confirm that an asymmetric moral hindsight effect occurs in various domains. It shows that judgments that a morally good action should have been taken are increased when people know the good action led to a good outcome, compared to when they do not know the outcome; but judgments that the good action should have been taken are not decreased when people know it led to a bad outcome. Moreover, it shows that judgments that a morally good action should have been taken are increased when people know the good action led to a good outcome, compared to when they know it led to a bad outcome. It replicates the findings in a laboratory based study with student participants tested in small groups, and extends the findings to a set of materials that concern not only self-sacrificial actions in life-and-death situations but also those in less dramatic encounters in sporting situations.

## 6. General discussion

A moral hindsight effect occurs for good actions with good outcomes, but not for good actions with bad outcomes. It may seem unfair to esteem successful attempts to help more than failed attempts to help, just as it seems unfair to condemn successful attempts to harm more than failed attempts to harm. But the tendency may reflect a rational reassessment of the putative causal relations between the action and its outcome. The discovery of an asymmetric moral hindsight effect corroborates a causal-inference hypothesis: the effect arises because people infer that the good outcome provides confirmation of its causal link to the good action, and they decide, with hindsight, that the protagonist could have known that her good action would lead to a good outcome. The bad outcome does not necessarily provide disconfirmation of the causal link between the good action and the anticipated good outcome because other factors may have intervened to prevent the outcome, and so people do not decide, with hindsight, that the protagonist could have known that her good action would not lead to a good outcome.

The discovery of an asymmetric moral hindsight effect goes against an aversive-emotion hypothesis. It predicts that judgments that the morally good action should have been taken will be reduced when it

results in a bad outcome because people will experience a negative emotional reaction to the bad outcome; and it predicts that judgments that the morally good action should have been taken will be unaffected when it results in a good outcome. But the results go against both of these predictions. The discovery indicates that moral hindsight does not arise from an aversive emotional reaction to an outcome. The results also rule out an alternative explanation that participants make a ‘god’s eye view’ judgment that the protagonist should incur the cost of the self-sacrificial good action only when it is the best action to result in the benefits of a good outcome. On this account, judgments that the action should have been taken should be decreased when the action fails and the outcome is bad, since the bad outcome indicates that there was no beneficial outcome even though the protagonist incurred a cost. The finding that judgments that the action should be taken are not decreased go against the costs-and-benefits view.

Imagined alternatives modify moral hindsight: a counterfactual amplifies judgments that an action should have been taken whereas a semi-factual diminishes them, for morally good actions that lead to good outcomes or bad outcomes. Imagined alternatives to reality have widespread effects on moral judgment (e.g., Byrne, 2005; Malle, Guglielmo, & Monroe, 2014; McCloy & Byrne, 2000; Monroe & Malle, 2017; Phillips, Luguri, & Knobe, 2015). The discovery rules out the aversive-emotion idea that imagined alternatives affect judgments because they shift the focus from the actual outcome, since semi-factuals place the focus on the same actual outcome. Instead it suggests that imagined alternatives affect judgments because they enrich the construction of a causal inferential model that explicitly represents the relations between the action and the outcome (e.g., Byrne, 2017; Sloman & Lagnado, 2015). Both discoveries are consistent with the idea that moral judgments implicate inferential processes (e.g., Bonnefon & Trémolière, 2017; see also Parkinson & Byrne, 2017b; Parkinson & Byrne, 2018).

## Acknowledgements

We thank Mary Parkinson for helpful discussions, Tiago Almeida for contributions to related preliminary studies and to Experiment 4, Rory Vignoles, Evie Alkin and Meg Ryan for help with data collection. We are grateful for comments on the results at talks at the 2015 APS

Convention in New York, the 2016 International Thinking Conference in Providence, and the 2017 Experimental Psychology Conference in Australia.

### Funding

This research was supported by the John Templeton Foundation, grant number 48054 awarded to Ruth Byrne. We confirm the funders had no role in the study design, collection, analysis or interpretation of data, writing of the report or decision to submit for publication.

### Conflict of interest

None.

### Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cognition.2018.05.010>.

### References

- Algoe, S. B., & Haidt, J. (2009). Witnessing excellence in action. *The Journal of Positive Psychology, 4*(2), 105–127.
- Alicke, M. D., Buckingham, J., Zell, E., & Davis, T. (2008). Culpable control and counterfactual reasoning in the psychology of blame. *Personality and Social Psychology Bulletin, 34*, 1371–1381.
- Aquino, K., McFerran, B., & Laven, M. (2011). Moral identity and the experience of moral elevation in response to acts of uncommon goodness. *Journal of Personality and Social Psychology, 100*(4), 703.
- Baron, J., & Hershey, J. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*(4), 569–579.
- Bonnefon, J. F., & Trémolière, B. (2017). *Moral inferences*. Oxford: Routledge.
- Branscombe, N. R., Owen, S., Garstka, T. A., & Coleman, J. (1996). Rape and accident counterfactuals. *Journal of Applied Social Psychology, 26*(12), 1042–1067.
- Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31*, 61–83.
- Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge, MA: MIT press.
- Byrne, R. M. J. (2016). Counterfactual thought. *Annual Review of Psychology, 67*, 153–157.
- Byrne, R. M. J. (2017). Counterfactual thinking: From logic to morality. *Current Directions in Psychological Science, 26*(4), 314–322.
- Cherubini, P., Castelvécchio, E., & Cherubini, A. M. (2005). Generation of hypotheses in Wason's 2–4–6 task: An information theory approach. *The Quarterly Journal of Experimental Psychology Section A, 58*(2), 309–332.
- Cox, K. S. (2010). Elevation predicts domain-specific volunteerism 3 months later. *The Journal of Positive Psychology, 5*(5), 333–341.
- Cushman, F. (2008). Crime and punishment. *Cognition, 108*(2), 353–380.
- Cushman, F. (2013). Action, outcome, and value. *Personality and Social Psychology Review, 17*(3), 273–292.
- Diessner, R., Iyer, R., Smith, M. M., & Haidt, J. (2013). Who engages with moral beauty? *Journal of Moral Education, 42*(2), 139–163.
- Fischhoff, B. (1975). Hindsight does not equal foresight. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–299.
- Fleischhut, N., Meder, B., & Gigerenzer, G. (2017). Moral hindsight. *Experimental Psychology, 64*, 110–123.
- Freeman, D., Aquino, K., & McFerran, B. (2009). Overcoming beneficiary race as an impediment to charitable donations. *Personality and Social Psychology Bulletin, 35*(1), 72–84.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105–2108.
- Gubbins, E., & Byrne, R. M. (2014). Dual processes of emotion and reason in judgments about moral dilemmas. *Thinking & Reasoning, 20*(2), 245–268.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Strauss & Giroux.
- Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D. A., & Knobe, J. (2015). Causal superseding. *Cognition, 137*, 196–209.
- Lagnado, D. A., Gerstenberg, T., & Zultan, R. I. (2013). Causal responsibility and counterfactuals. *Cognitive Science, 37*(6), 1036–1073.
- Lench, H. C., Domskey, D., Smallman, R., & Darbor, K. E. (2015). Beliefs in moral luck: When and why blame hinges on luck. *British Journal of Psychology, 106*(2), 272–287.
- Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry, 25*(2), 147–186.
- Markman, K. D., Mizoguchi, N., & McMullen, M. N. (2008). It would have been worse under Saddam. *Journal of Experimental Social Psychology, 44*(3), 650–654.
- Martin, J. W., & Cushman, F. (2016a). Why we forgive what can't be controlled. *Cognition, 147*, 133–143.
- Martin, J. W., & Cushman, F. A. (2016b). The adaptive logic of moral luck. In J. Sytma, & W. Buckwalter (Eds.). *The Blackwell companion to experimental philosophy* (pp. 190–202). Oxford: Wiley Blackwell.
- McCloy, R., & Byrne, R. M. (2000). Counterfactual thinking about controllable events. *Memory & Cognition, 28*(6), 1071–1078.
- McCloy, R., & Byrne, R. M. (2002). Semifactual “even if” thinking. *Thinking & Reasoning, 8*(1), 41–67.
- McEleney, A., & Byrne, R. M. J. (2006). Spontaneous causal and counterfactual thoughts. *Thinking and Reasoning, 12*, 235–255.
- Mikhail, J. (2007). Universal moral grammar. *Trends in Cognitive Sciences, 11*(4), 143–152.
- Monroe, A. E., & Malle, B. F. (2017). Two paths to blame. *Journal of Experimental Psychology: General, 146*(1), 123.
- Oaksford, M., & Chater, N. (2017). Causal models and conditional reasoning. In M. Waldmann (Ed.). *Oxford handbook of causal reasoning* (pp. 327–346). Oxford: Oxford University Press.
- Oeberst, A., & Goeckenjan, I. (2016). When being wise after the event results in injustice: Evidence for hindsight bias in judges' negligence assessments. *Psychology, Public Policy, and Law, 22*(3), 271.
- Parkinson, M., & Byrne, R. M. J. (2017a). Counterfactual and semifactual thoughts in moral judgments about failed attempts to harm. *Thinking and Reasoning, 23*(4), 409–448.
- Parkinson, M., & Byrne, R. M. J. (2017b). Moral judgments of risky choices: A moral echoing effect. *Judgment & Decision Making, 12*(3), 236–252.
- Parkinson, M., & Byrne, R. M. J. (2018). Judgments of moral responsibility and wrongness for intentional and accidental harm and purity violations. *Quarterly Journal of Experimental Psychology, 71*(3), 779–789.
- Phillips, J., Luguri, J. B., & Knobe, J. (2015). Unifying morality's influence on non-moral judgments. *Cognition, 145*, 30–42.
- Rasga, C., Quelhas, A. C., & Byrne, R. M. (2016). Children's reasoning about other's intentions. *Cognitive Development, 40*, 46–59.
- Roese, N., & Epstude, K. (2017). The functional theory of counterfactual thinking. *Advances in Experimental Social Psychology, 56*, 1–79.
- Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science, 7*(5), 411–426.
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*(4), 296–320.
- Sarin, A., Lagnado, D. A., & Burgess, P. W. (2017). The intention-outcome asymmetry effect. *Experimental Psychology, 64*(2), 124–141.
- Schnall, S., & Roper, J. (2012). Elevation puts moral values into action. *Social Psychological and Personality Science, 3*(3), 373–378.
- Schnall, S., Roper, J., & Fessler, D. M. (2010). Elevation leads to altruistic behavior. *Psychological Science, 21*(3), 315–320.
- Sloman, S. A., & Lagnado, D. (2015). Causality in thought. *Annual Review of Psychology, 66*, 223–247.
- Timmons, S., & Byrne, R. M. J. (2018). Moral fatigue: the effects of cognitive fatigue on moral reasoning. *Quarterly Journal of Experimental Psychology*. <http://dx.doi.org/10.1177/1747021818772045> in press.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458.
- Walsh, C. R., & Byrne, R. M. J. (2007). How people think “if only...” about reasons for actions. *Thinking & Reasoning, 13*(4), 461–483.
- Young, L., Nichols, S., & Saxe, R. (2010). Investigating the neural and cognitive basis of moral luck. *Review of Philosophy and Psychology, 1*(3), 333–349.