



Counterfactual and semi-factual thoughts in moral judgements about failed attempts to harm

Mary Parkinson & Ruth M. J. Byrne

To cite this article: Mary Parkinson & Ruth M. J. Byrne (2017) Counterfactual and semi-factual thoughts in moral judgements about failed attempts to harm, *Thinking & Reasoning*, 23:4, 409-448, DOI: [10.1080/13546783.2017.1345790](https://doi.org/10.1080/13546783.2017.1345790)

To link to this article: <http://dx.doi.org/10.1080/13546783.2017.1345790>



Published online: 06 Jul 2017.



[Submit your article to this journal](#)



Article views: 55



[View related articles](#)



[View Crossmark data](#)



Counterfactual and semi-factual thoughts in moral judgements about failed attempts to harm

Mary Parkinson and Ruth M. J. Byrne

School of Psychology and Institute of Neuroscience, Trinity College Dublin, University of Dublin, Dublin, Ireland

ABSTRACT

People judge that an individual who attempts to harm someone but fails should be blamed and punished more when they imagine how things could have turned out worse, compared to when they imagine how things could have turned out the same, or when they think only about what happened. This moral counterfactual amplification effect occurs when people believe the protagonist had no reason for the attempt to harm, and not when the protagonist had a reason, as Experiment 1 shows. It occurs for intentional failed attempts to harm and also for accidental near-misses, as Experiment 2 shows, but not for failed attempts in which the harm occurs anyway by another cause, for both general judgments about the event and specific judgments about the individual's actions, as Experiments 3 and 4 show. The implications for understanding the role of counterfactual thoughts in moral judgement are discussed.

ARTICLE HISTORY Received 15 April 2016; Accepted 16 June 2017

KEYWORDS Moral judgement; punishment; blame; counterfactual; semi-factual

How do people make moral judgements about failed attempts to harm? Suppose Jenny is taking a class in sculpture and she is assigned to work with a partner to weld together pieces of metal. She wants to burn her partner's hand and she believes that if she welds a piece of metal that her partner is holding the heat will travel down the metal to burn her. Jenny welds the metal, but her partner happens to let go and is not burned at all (Cushman, 2008). In four experiments, we examine people's judgements about how much Jenny should be punished, how much blame she deserves and how wrong her behaviour was. Our aim is to examine the effects of imagining what could have happened. We compare stories that contained a counterfactual about how things could have turned out worse, "if Jenny's partner had not let go of the metal, her hand would have been burned", or a semi-factual about how things could have turned out the same, "even if Jenny's partner had not let go of the metal, her hand would not have been burned". Our goal

CONTACT Mary Parkinson  parkinsm@tcd.ie

© 2017 Informa UK Limited, trading as Taylor & Francis Group

is to test whether people's judgements of punishment, blame and wrongness are amplified for failed attempts to harm when they imagine how things could have been worse.

Failed attempts to harm

A curious phenomenon about moral judgements is that even when an individual clearly intended to harm someone, the outcome matters (e.g., Nagel, 1993; Williams, 1981). An individual who attempts to harm and fails receives lighter consequences than one who attempts to harm and succeeds. For example, offences that are inchoate such as attempts to murder, conspiracy or incitement typically receive only half the legal sentence when an individual is convicted (e.g., Feinberg, 1995; Hart & Honore, 1959; Kadish, 1993). Consider two individuals who wanted to harm people driving underneath a bridge and threw bricks over a high wall on the bridge unable to see what was below. One individual's brick hit the pavement but the other individual's brick hit a car and killed someone. Participants blame the individual whose brick killed someone more than the individual whose brick hit the pavement, even though both individuals had the same intent and knowledge and carried out the same action (e.g., Lench, Domskey, Smallman, & Darbor, 2015). Hence, people make different judgements about failed and successful attempts to harm even when an individual failed to harm because of circumstances beyond their control rather than any decision to withdraw the attempt (e.g., Domskey, 2004; Enoch & Marmour, 2007; Nelkin, 2009; Nichols, Timmons, & Lopez, 2014; Young, Nichols, & Saxe, 2010). Young children's moral judgements also tend to be influenced by outcomes rather than intentions, whereas older children attempt to take intentions into account as well as outcomes (e.g., Piaget, 1965). Nonetheless, the influence of outcomes on moral judgements persists even for adults (e.g., Baron & Hershey, 1988; Cushman, 2008; Robinson, Goodwin, & Reisig, 2010; Walster, 1966). The failure of an attempt to harm is a matter of "moral luck" and so it can seem unfair and even irrational that people are more lenient in their moral judgements of an individual for a failed attempt to harm compared to a successful one (e.g., Young et al., 2010). However, people often rely on outcomes to provide additional information about a person's intentions, and the circumstances in which they acted (e.g., Baron & Hershey, 1988; Pizarro, Uhlmann, & Salovey, 2003). Their moral judgements are also affected by hindsight, that is, people with outcome knowledge tend to believe they would have predicted the outcome all along (e.g., Fleischhut, Meder, & Gigerenzer, 2017; Lench et al., 2015).

Imagined alternatives to reality

Our experiments examine whether people judge failed attempts to harm differently when they imagine how the outcome could have turned out. People

often create counterfactual alternatives to reality and imagine how an outcome could have turned out better or worse (see Byrne, 2016, for a review). These thoughts amplify emotional experiences such as guilt and regret, or relief after near-misses (e.g., Epstude & Roese, 2008; Kahneman & Miller, 1986; Larsen, McGraw, Mellers, & Cacioppo, 2004; Sweeny & Vohs, 2012). People think about what might have been to explain and defend past events (e.g., Ferrante, Giroto, Stragà, & Walsh, 2013; Markman, Mizoguchi, & McMullen, 2008; Walsh & Byrne, 2007). Their counterfactual thoughts help to identify causal relations (e.g., Byrne, 2005; Frosch & Byrne, 2012; Lagnado, Gerstenberg, & Zultan, 2013). They also help to form plans to prevent similar mistakes in the future (e.g., Epstude & Roese, 2008; Markman & McMullen, 2003).

Counterfactual thoughts affect judgements of blame for bad outcomes. Participants blame a victim more when they hear that the outcome of an attack would not have been as bad if the victim had acted differently, compared to when they hear that it would have been just as bad even if the victim had acted differently (e.g., Branscombe, Owen, Garstka, & Coleman, 1996). Similarly, participants blame a doctor more for prescribing medication that killed a patient when they believe he could have prescribed a different medication and the patient would have survived, compared to when they believe that even if he had prescribed a different medication the patient still would have died (e.g., Alicke, Buckingham, Zell, & Davis, 2008). And participants blame the individual who threw a brick from the overpass bridge that landed on the pavement just as much as the one whose brick landed on the car and killed someone, when they imagine how things could have turned out worse (e.g., Lench et al., 2015).

Failed attempts to harm and imagined alternatives

The aim of the four experiments we report was to examine how thoughts about the way an outcome could have turned out affect moral judgements about failed attempts to harm. In each experiment, participants were presented with stories in which an individual's attempt to harm fails, such as the example about Jenny welding metal (see Appendix for the full set of 24 stories). We created a counterfactual story ending in which participants were asked to consider how the outcome could have been worse, e.g., We assume that reading about how the outcome could have been worse will affect people's moral judgements about the failed attempt to harm because they will think about two possibilities, the facts – Jenny's partner let go of the metal and was not burned, and an imagined worse alternative – Jenny's partner did not let go and was burned (e.g., Byrne, 2005; Byrne & Tasso, 1999; Thompson & Byrne, 2002; Santamaria, Espino, & Byrne, 2005). People tend to focus on one or the other of these two possibilities (e.g., Markman & McMullen, 2003; McMullen & Markman, 2000). We propose that when people imagine how the

outcome could have been worse, they appreciate that the outcome very nearly turned out badly – Jenny’s partner almost got burned. Our view contrasts with a prominent suggestion that attempts to harm, such as attempted murder, are judged more leniently than actual harm, such as murder, because when people imagine how the outcome could have been worse, they appreciate that the actual outcome was in fact good – Jenny’s partner did not get burned (e.g., Kadish, 1993).

If Jenny’s partner had not let go of the metal, her hand would have been burned.

Our contrary proposal is based on the evidence that people’s judgements of a causal link between an action and its outcome are increased when they read a counterfactual (e.g., Branscombe et al., 1996; Gavanski & Wells, 1989; McCloy & Byrne, 2002; Spellman & Mandel, 1999). For example, when people hear that an athlete who took a painkiller experienced side effects of fatigue and nausea and lost a race, they judge that the painkiller caused the athlete to lose the race. They do so more often when they know about a counterfactual, other competitors took another painkiller with no side effects, compared to when they know about a semi-factual, other competitors took another painkiller with the same side effects (McCloy & Byrne, 2002). Hence, we suggest that a counterfactual such as “If Jenny’s partner had not let go, her hand would have been burned” emphasises that Jenny’s partner letting go was the cause of her hand not being burnt. Hence, it highlights that Jenny’s action would have caused harm but for her partner’s intervention (e.g., Teigen & Jensen, 2011). It underscores that Jenny almost caused harm, which was prevented by factors outside her control (see Appendix for further examples). As a result, we predict a moral counterfactual amplification effect: participants will judge that the protagonist should be punished and blamed more when they read about how things could have been worse compared to when they read only about the facts of what happened.

We compared stories about failed attempts to harm that contained a counterfactual to stories that contained no counterfactual, and also to stories that contained a semi-factual about how things could have turned out the same, e.g., A semi-factual changes one of the actions that led to the outcome, but the outcome remains the same (e.g., Boninger, Gleicher, & Strathman, 1994; McCloy & Byrne, 2002). People’s judgements of a causal link between an action and its outcome are decreased when they read a semi-factual (e.g., Branscombe et al., 1996; McCloy & Byrne, 2002). The semi-factual emphasises that Jenny’s partner letting go was not the cause of the outcome. Indeed an “even if...” thought often makes the outcome appear inevitable (e.g., Byrne, 2016). As a result, we predict that participants will judge that the protagonist should be punished and blamed more when they read a counterfactual about

how things could have been worse compared to when they read a semi-factual about how things could have turned out the same.

'Even if Jenny's partner had not let go, her hand would not have been burned.'

In the first experiment, participants read stories that contained a counterfactual, a semi-factual or just the facts. We also compared failed attempts to harm that came about when the protagonist had no reason, e.g., to those that came about when the protagonist had a reason, e.g., We expect that counterfactual thoughts about how things could have been worse will increase people's blame and punishment judgements of a protagonist who has no reason. People can construct a simple model in which a single bad action by the protagonist almost caused a bad outcome in the no-reason story. In this simple model, the counterfactual emphasises the causal role of the protagonist's actions in the nearly occurring bad outcome. To make moral judgements people need to be able to construct a model that explicitly connects an action to its outcome (e.g., Cushman, 2013; Miller, Hannikainen, & Cushman, 2014). In contrast, we expect counterfactuals may not increase people's blame and punishment judgements of a protagonist who has an enduring reason. People need to construct a more complex model to capture a context of a persistent feud with possibly many bad actions, possibly by the two protagonists, in which this current single action by the protagonist almost caused a bad outcome. In this complex model, the counterfactual may be less effective in emphasising the causal role of the protagonist's action in the nearly occurring bad outcome, against a backdrop of possibly many bad actions and bad outcomes.

For no reason at all, Jenny wants to burn her partner's hand.

'Because of a long-standing feud, Jenny wants to burn her partner's hand.'

The second and third experiments examine the effects of counterfactuals and semi-factuals on people's moral judgements about failed attempts to harm compared to two different sorts of control comparisons. In Experiment 2, the comparison was to accidental near-misses of harm, and in Experiment 3, the comparison was to failed attempts to harm in which harm occurs by another cause. In the first comparison, participants were told that the protagonist wanted to bring about the outcome for the failed attempt to harm: whereas they were told the protagonist did not want to bring about the outcome for the accidental near-miss: We predicted that a counterfactual amplification effect would occur in both cases. People often spontaneously create counterfactuals after accidental near-misses and we anticipate that thoughts about how things could have turned out worse will amplify moral judgements in such situations, just as in failed attempts to harm.

'Jenny wants to burn her partner's hand. Jenny believes that if she welds a piece of metal that her partner is holding the heat will travel down the metal to burn her partner's hand.'

'Jenny does *not* want to burn her partner's hand. Jenny believes that if she welds a piece of metal that her partner is holding the heat will *not* travel down the metal to burn her partner's hand.'

In the second comparison, participants were told: We predicted that a counterfactual amplification effect would not occur in this case. Thoughts about how things could have been worse amplify moral judgements because they highlight the causal role of the protagonist, and in this case, the protagonist did not cause the outcome. The fourth experiment compared the effects of general judgements, e.g., "how much should Jenny be punished?" to specific judgements, e.g., "how much should Jenny be punished for her attempt to harm her partner?" The conditions of the four experiments are sketched in Table 1.

'Jenny welds the metal, but her partner happens to let go and is not burned at all. However, Jenny's partner picks up a different piece of hot metal and is burned.'

Experiment 1

The aim of the experiment was to establish whether there is a moral counterfactual amplification effect in people's judgements of punishment, blame and wrongness for failed attempts to harm. Participants read stories that contained a counterfactual in which the outcome was worse, e.g., or a semi-factual in which the outcome was the same, e.g., or no imagined alternative, e.g., We predict that a counterfactual amplification effect will be observed, in that participants will make harsher moral judgements of punishment, blame and wrongness when they read a counterfactual, compared to when they read a semi-factual or no alternative.

'If Jenny's partner had not let go of the metal that Jenny was welding, her hand would have been burned.'

'Even if Jenny's partner had not let go of the metal that Jenny was welding, her hand would not have been burned.'

'Jenny's partner continues with her work, and everyone else carries on with their work.'

In addition, we compared failed attempts to harm that came about when the protagonist had no reason, e.g., to those that came about when the protagonist had a reason, e.g., People's moral judgements are influenced by their assessment of a protagonist's intentions (e.g., Knobe, 2010; Newman, De Freitas, & Knobe, 2015; Phillips et al., 2015). They consider the individual's reasons and motivation for carrying out an immoral act, for example, whether it was accidental or coerced (e.g., Malle, Guglielmo, & Monroe, 2014; Martin & Cushman, 2016; Monroe, Dillon & Malle, 2014; Young & Tsoi, 2013). We predict that the counterfactual amplification effect will occur when the protagonist is

**Table 1.** Examples of the materials in the conditions of the four experiments.

	Example of story	Counterfactual	Semi-factual
Experiment 1 Failed attempt – no reason	Jenny wants to burn her partner's hand. She believes if she welds metal her partner is holding the heat will travel down to burn her partner's hand. Jenny welds the metal but her partner happens to let go and is not burned at all.	<i>Worse outcome:</i> If Jenny's partner had not let go of the metal that Jenny was welding, her hand would have been burned.	<i>Same outcome:</i> Even if Jenny's partner had not let go of the metal that Jenny was welding, her hand would not have been burned.
Failed attempt – Reason	For no reason at all, Jenny wants to burn her partner's hand.	<i>As Example</i>	<i>As Example</i>
Experiment 2 Failed attempt	Because of a long-standing feud, Jenny wants to burn her partner's hand.	<i>As Example</i>	<i>As Example</i>
Near-miss	<i>As Example</i> Jenny does <i>not</i> want to burn her partner's hand. She believes if she welds metal her partner is holding the heat will <i>not</i> travel down to burn her partner's hand. Her partner happens to let go and is not burned at all.	The metal Jenny welded transmitted heat quickly. <i>As Example</i> <i>As above</i>	The metal Jenny welded did not transmit heat quickly. <i>As Example</i> <i>As above</i>
Experiment 3 Failed attempt Other-cause	<i>As Example</i> <i>As Example</i> However, Jenny's partner picks up a different piece of hot metal and is burned.	<i>As Example</i> <i>Better outcome:</i> If Jenny's partner had not let go of the metal that Jenny was welding, her hand would not have been burned.	<i>As Example</i> <i>Same outcome:</i> Even if Jenny's partner had not let go of the metal that Jenny was welding, her hand would have been burned.
Experiment 4 Failed attempt	<i>As Example</i>	General judgements How much should Jenny be punished? How much blame does Jenny deserve? How wrong was Jenny's behaviour?	Specific judgements How much should Jenny be punished for her attempt to harm her partner? How much blame does Jenny deserve for her attempt to harm her partner? How wrong was Jenny's attempt to harm her partner?
Other-cause	<i>As Example</i> However, Jenny's partner picks up a different piece of hot metal and is burned.	<i>As above</i>	<i>As above</i>

described as having no reason, because the counterfactual emphasises the causal role of the individual's action in the nearly occurring bad outcome, but not when the protagonist is described as having a reason.

'For no reason at all, Jenny wants to burn her partner's hand.'

'Because of a long-standing feud, Jenny wants to burn her partner's hand.'

Method

Participants and procedure

The participants were 312 volunteers, 156 men and 156 women, aged 17–55 years, with a mean age of 35 years, recruited from the online platform, Crowdfunder (see [Crowdfunder.com](http://www.crowdfunder.com)). The 312 participants were those who remained when 33 participants were eliminated prior to any data analysis on the basis of several criteria used to screen participants: participants were asked to select a name of a story protagonist from a list of five names and participants who chose an incorrect name were eliminated, participants were asked to participate only if they had not taken part in something similar before; they were asked at the end of the experiment whether they had done something similar before and if so, what its subject was, and participants who had taken part in something similar before were eliminated. Participants were assigned at random to one of six conditions: reason counterfactual ($n = 58$), no-reason counterfactual ($n = 45$), reason semi-factual ($n = 59$), no-reason semi-factual ($n = 52$), reason control ($n = 50$), no-reason control ($n = 48$). The materials were presented via SurveyGizmo (see <http://www.surveymgizmo.com/>). Participants were directed to it from Crowdfunder. Participants were compensated \$0.15 for their time. They were asked to read each story carefully and to answer the questions in the order in which they were asked. They were told there were no right or wrong answers. The experiment took approximately 5–10 minutes to complete.

Materials and design

Participants were assigned at random to the six conditions in a 3 (imagined alternative: counterfactual vs. semi-factual vs. control) \times 2 (motivation: reason vs. no-reason) fully between-participants design. They received two scenarios with different contents in a different randomized order for each participant (see Appendix). The participants completed three moral judgements: (1) Punishment, e.g., "How much should Jenny be punished?" (2) Blame, e.g., "How much blame does Jenny deserve?" and (3) Wrongness, e.g., "How wrong was Jenny's behaviour?". They provided their judgements on seven-point scales with 1 anchored as "not at all", 4 as "somewhat" and 7 as "very much".

Results and discussion

The data for all of the experiments are available at <https://reasoningandimagination.wordpress.com/data-archive/>

We carried out three ANOVAs of a 3 (imagined alternative: counterfactual vs. semi-factual vs. no-alternative) \times 2 (motivation: reason vs. no-reason) between participants design, on the punishment, blame and wrongness judgements (see Figure 1).

Punishment judgements. There was a main effect of imagined alternative, $F(2, 306) = 3.79, p = .024, \eta p^2 = .024$, as the protagonist was judged to be more deserving of punishment when participants read a counterfactual than no-alternative, $F(1, 306) = 7.57, p = .006, \eta p^2 = .024$; there were no differences between a counterfactual and a semi-factual, $F(1, 306) = 1.65, p = .2$, or between a semi-factual and no-alternative, $F(1, 306) = 2.36, p = .13$, as Figure 1 shows. There was no main effect of motivation, $F(1, 306) = 1.02, p = .31$, and the two variables did not interact, $F(2, 306) = 1.84, p = .16$.

Blame judgements. There were no main effects of imagined alternative or motivation, $F < 1$ in each case, but the two variables interacted, $F(2, 306) = 4.57, p = .011, \eta p^2 = .029$. Planned contrasts showed that in the no-reason condition, participants blamed the protagonist more when they read a counterfactual than no-alternative, $F(2, 306) = 6.33, p = .012, \eta p^2 = .02$, or a semi-

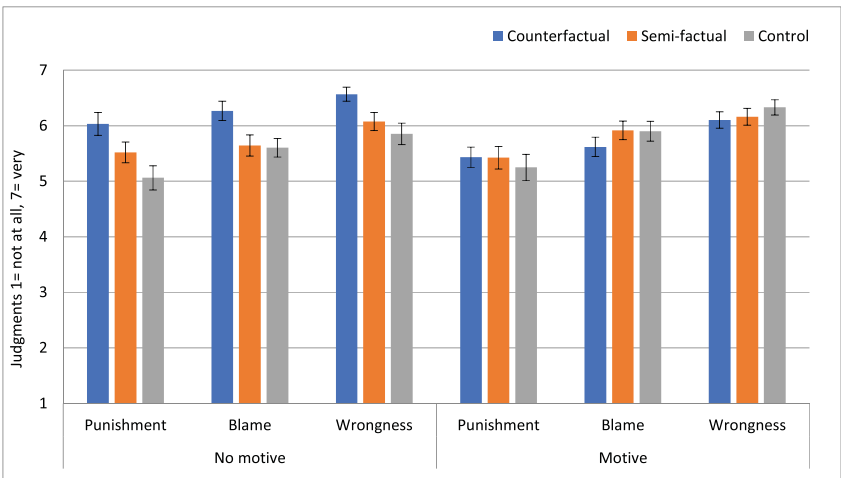


Figure 1. Moral judgements in response to "How much should Jenny be punished?", "How much blame does Jenny deserve?" and "How wrong was Jenny's behaviour?" in Experiment 1. Error bars are standard error of the mean.

factual, $F(2, 306) = 5.81$, $p = .017$, $\eta p^2 = .019$, and there was no difference between the semi-factual and no-alternative conditions, $F < 1$. In contrast, in the reason condition, there were no differences between the counterfactual vs. no-alternative, $F(2, 306) = 1.3$, $p = .26$, counterfactual vs. semi-factual, $F(1, 306) = 1.58$, $p = .21$, or semi-factual vs. no-alternative, $F < 1$, as Figure 1 shows. The reason and no-reason conditions differed for the counterfactual $F(1, 306) = 6.57$, $p = .011$, $\eta p^2 = .021$, but not the semi-factual, $F(1, 306) = 1.26$, $p = .26$, or no-alternative conditions, $F(1, 306) = 1.33$, $p = .25$.

Wrongness judgements. There were no main effects of imagined alternative, $F(2, 306) = 1.44$, $p = .24$ or motivation, $F < 1$, but the two variables interacted, $F(1, 306) = 4.44$, $p = .013$, $\eta p^2 = .028$. Planned contrasts showed that in the no-reason condition, participants judged the protagonist's behaviour to be wrong more often when they read a counterfactual than no-alternative, $F(1, 306) = 9.41$, $p = .002$, $\eta p^2 = .03$, or a semi-factual, $F(1, 306) = 4.62$, $p = .032$, $\eta p^2 = .015$, there was no difference between the semi-factual and no-alternative, $F < 1$. In contrast, in the reason condition, there were no differences, $F < 1$ in each case, as Figure 1 shows. The reason and no-reason conditions differed for the counterfactual, $F(1, 306) = 4.34$, $p = .038$, $\eta p^2 = .014$ and no-alternative, $F(1, 306) = 4.43$, $p = .03$, $\eta p^2 = .014$; there were no differences for the semi-factual condition $F < 1$.

The experiment shows that participants judged individuals more harshly when they read a story that contained a counterfactual about how the outcome could have been worse compared to when they read no alternative. This amplification was observed for judgements of punishment, blame and wrongness. The results indicate that imagining how things could have been worse intensifies moral judgements about failed attempts to harm, compared to not imagining any alternative. Participants also judged individuals more harshly when they read a story that contained a counterfactual compared to one that contained a semi-factual about how the outcome could have turned out the same. There were no differences in the harshness of judgements when participants read a semi-factual or no alternative. The comparison clarifies that it is not merely imagining how things could have turned out that leads to the amplification effect, but imagining how they could have turned out worse. The result corroborates the suggestion that the counterfactual underscores that Jenny's action would have caused harm but for her partner's intervention.

The experiment also shows that the counterfactual amplification occurs when participants are told explicitly that the protagonist's attempt to harm occurred for no reason, but it does not occur when the attempt occurred for an enduring reason, such as a long-standing feud. The result is consistent with the suggestion that the counterfactual emphasises the causal role of the

protagonist's action in nearly bringing about a bad outcome in the simple model of the no-reason scenario, but not in the more complex model that includes the context of an enduring feud with possibly many other causal links. People may explain the protagonist's behaviour in terms of the feud in general, rather than by causal and counterfactual reasoning about the specific attempt. Of course participants may consider that protagonists who are described as having no reason for their attempt to harm are somehow suspicious. They may consider them to be morally worse than those described as having an enduring feud as a reason, given that reasons can act as a justification for behaviour. However, this explanation predicts a main effect of motivation – people should judge protagonists more harshly when they have no reason. The results showed no main effect of motivation for any of the moral judgements. In the remaining experiments, we convey the protagonists' lack of reason implicitly rather than explicitly. The next experiment examines intentions further by comparing failed attempts to harm to unintentional near-misses.

Experiment 2

The aim was to examine people's moral judgements for failed attempts to harm compared to accidental near-misses, e.g., People readily create counterfactuals for near-misses, for example, they tend to judge that a person who misses a flight by five minutes will feel worse than one who misses a flight by half an hour (e.g., Kahneman & Tversky, 1982; McMullen & Markman, 2000; McMullen & Markman, 2002). Of course, we predict that people will judge an individual more harshly for a failed attempt to harm than for an accidental near-miss. But we also predict that participants will judge an individual more harshly when they read about how the outcome could have been worse rather than about how it could have turned out the same, for an accidental near-miss just as for a failed attempt to harm.

'Jenny does *not* want to burn her partner's hand. She believes that if she welds a piece of metal that her partner is holding the heat will *not* travel down the metal to burn her partner's hand. Jenny welds the metal, but her partner happens to let go and is not burned at all.'

In this comparison, the counterfactuals all referred to a worse outcome, e.g., "If Jenny's partner had not let go of the metal that Jenny was welding, her hand would have been burned", and the semi-factuals to the same outcome, "Even if Jenny's partner had not let go of the metal that Jenny was welding, her hand would not have been burned". We examine judgements of punishment, blame and wrongness, and we also test whether participants are primed by counterfactuals more than semi-factuals to read a punishment statement quickly, e.g., "Jenny deserves to be punished". The time

participants take to read the punishment statement can provide some information about the extent to which the statement concurs with their expectations, and hence whether their expectations of punishment increase when they are primed by a counterfactual rather than a semi-factual.

Method

Participants and procedure

The participants were 28 university students at Trinity College Dublin who participated for course credit. There were 15 women and 13 men, aged between 17 and 51, with a mean age of 22 years. Participants were tested individually or in small groups of two or three individuals. They completed the experiment on a Macintosh computer running Superlab software to present the materials, and to record latencies and responses. The instructions were the same as the previous experiment. They performed a practice trial at the outset. The experiment took approximately 10 minutes to complete.

Materials and design

Participants acted as their own controls and received 24 scenarios in a 2 (intent: accidental vs. intentional) x 2 (imagined alternative: counterfactual vs. semi-factual) design with repeated measures on both factors. They received 24 different contents, 6 instances in each of the 4 conditions (see Appendix). We assigned the contents randomly to create four sets of different combinations of contents; participants were assigned at random to receive one set so that each participant received six instances in each condition, but not the same content twice. Each participant received the stories in a different randomised order.

Latencies were recorded for the final sentence in each story which was an assertion "<the individual> deserves to be punished". Participants pressed the space bar to see each sentence, and the latency to read the final punishment statement, e.g., "Jenny deserves to be punished" was recorded from when participants pressed the space bar and the punishment statement appeared on screen, to when they pressed the space bar to read the next sentence (which was the first moral judgement). The tasks and measures were the same as the previous experiment.

Results and discussion

We carried out four within-participant ANOVAs with a design of 2 (intent: failed attempt vs. accidental near-miss) x 2 (imagined alternative: counterfactual vs. semi-factual) on the four measures (see [Figure 2](#)).

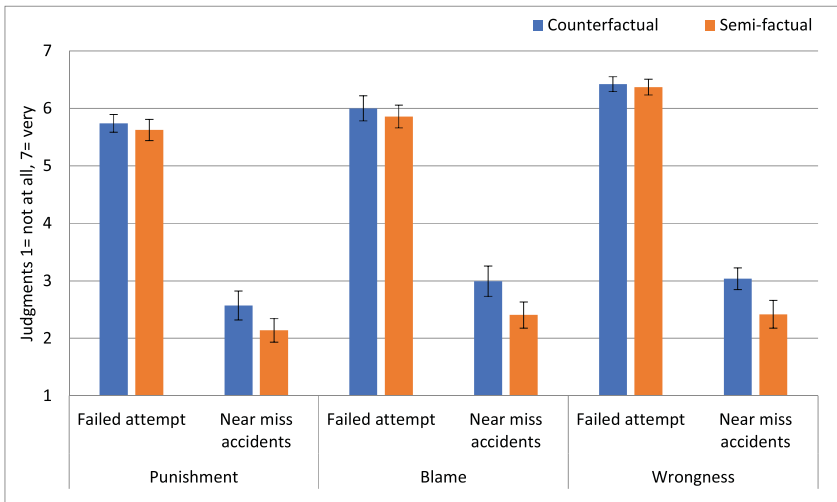


Figure 2. Moral judgements in response to “How much should Jenny be punished?”, “How much blame does Jenny deserve?” and “How wrong was Jenny’s behaviour?” in Experiment 2. Error bars are standard error of the mean.

Punishment latencies and judgements. There was no main effect of imagined alternatives, $F < 1$, or intent, $F(1, 27) = 1.17, p = .29$, and the two variables did not interact, $F < 1$ in the ANOVA on the latencies to read the punishment assertion, transformed to Log 10. Prior to any data analysis, outliers were removed, defined as latencies more than three times the interquartile range of each condition, for each participant, resulting in 3% of latencies eliminated. Participants rapidly read the punishment assertion for a failed attempt with a counterfactual ($M = 1069.7$ msec, $SD = 374.8$), or a semi-factual ($M = 1171.07$ msec, $SD = 589.27$), and for near-misses with a counterfactual ($M = 1176.68$ msec, $SD = 448.01$) or a semi-factual ($M = 1167.3$ msec, $SD = 416.57$). Nonetheless, participants subsequently judged the protagonist’s behaviour to be more deserving of punishment when they read a counterfactual compared to a semi-factual, $F(1, 27) = 5.16, p = .031, \eta p^2 = .16$, and they judged the protagonist’s behaviour to be more deserving of punishment for failed attempts than accidental near-misses $F(1, 27) = 280.51, p = .000, \eta p^2 = .91$, and the two variables did not interact, $F(1, 27) = 1.05, p = .32$, as Figure 2 shows.

Blame judgements. The protagonist’s behaviour was judged to be more deserving of blame when participants read a counterfactual about an imagined worse outcome compared to when they read a semi-factual about the same outcome, $F(1, 27) = 6.85, p = .014, \eta p^2 = .2$, and more deserving of blame for failed attempts than accidental near-misses, $F(1, 27) = 197.6, p = .000, \eta p^2 = .88$, and the two variables did not interact $F < 1$, as Figure 2 shows.

Wrongness judgements. Similarly, the protagonist's behaviour was judged to be more wrong when participants read a counterfactual than a semi-factual, $F(1, 27) = 7.35$, $p = .011$, $\eta p^2 = .21$, and more wrong for failed attempts than accidental near-misses $F(1, 27) = 278.62$, $p = .000$, $\eta p^2 = .91$, and the two variables did not interact, $F(1, 27) = 3.37$, $p = .078$, $\eta p^2 = .11$, as [Figure 2](#) shows.

The experiment shows, as expected, that people's judgements of punishment, blame and wrongness were harsher for an individual who attempted to harm than one who caused an accidental near-miss. The results also show that, in both cases, people's judgements were harsher when they read a counterfactual about how the outcome could have been worse, than when they read a semi-factual. The results indicate that imagining how things could have been worse increases moral opprobrium for failed attempts to harm and for accidental near-misses, perhaps because of a judgement that an individual did not take every precaution to prevent harm (e.g., Malle et al., 2014; Young et al., 2010).

A potential alternative explanation for the effect of counterfactuals on moral judgements of near-miss accidents is that in the counterfactual condition, the protagonist has false beliefs (Jenny believes that heating the metal will not burn her partner but it would have if her partner had not moved her hand), whereas in the semi-factual condition, the protagonist has true beliefs (Jenny believes that heating the metal will not burn her partner and in fact it would not have even if her partner had not moved her hand). Blame for accidents and near-miss accidents is affected by a protagonist's false beliefs (e.g., Young et al., 2010). However, this explanation can be ruled out by the observation that in the failed attempts condition, protagonists were judged more harshly when they had true beliefs (this time in the counterfactual condition, Jenny believes that heating the metal will burn her partner and in fact if her partner had not moved her hand it would have) compared to false beliefs (this time in the semi-factual condition: Jenny believes that heating the metal will burn her partner but even if her partner had not moved her hand she would not have been burned).

Although there was no interaction between imagined alternative and intention, an inspection of [Figure 2](#) shows differences between counterfactuals and semi-factuals for near-misses but relatively small differences between them for failed attempts. We suspect the difference between counterfactuals and semi-factuals may have been minimised by the inclusion of the explanatory sentence about the facts, e.g., "The metal Jenny welded transmitted heat quickly", immediately preceding the counterfactual, "if Jenny's partner had not let go of the metal that Jenny was welding, her hand would have been burned" and "The metal Jenny welded did not transmit heat quickly", immediately preceding the semi-factual, "if Jenny's partner had not let go of the metal that Jenny was welding, her hand would not have been burned". The factual statements may have prompted participants to focus on

the facts of what occurred, e.g., Jenny's partner was not burned, rather than on the counterfactual alternative, e.g., Jenny's partner would have been burned. We examine the effects of counterfactuals and semi-factuals on failed attempts to harm when this potential direction to the facts is removed in the next experiment, in which we compare failed attempts to harm, to situations in which the attempt failed, but the harm occurred by other means.

Experiment 3

The aim of the experiment was to compare failed attempts to harm to "other-cause" situations in which the attempt to harm failed, but harm occurred anyway, not caused by the protagonist (e.g., Cushman, 2008): Importantly, the counterfactual for the other-cause story: refers to an outcome that is *better* than the actual one (unlike those in the previous experiments). The semi-factual alternative refers to an outcome that is the same as the actual one: As in the previous experiments, we expect that participants' moral judgements about failed attempts to harm will be amplified by counterfactuals more than semi-factuals. In contrast, we expect no such effects for the other-cause stories because thinking about what could have happened does not highlight the protagonist's causal role in the outcome, as it was caused by other means.

'Jenny welds the metal, but her partner happens to let go and is not burned at all. However, Jenny's partner picks up a different piece of hot metal and is burned.'

'If Jenny's partner had not let go of the metal that Jenny was welding, her hand would not have been burned.'

'Even if Jenny's partner had not let go of the metal that Jenny was welding, her hand would have been burned.'

The other-cause stories also allow us to make two important comparisons. First, we can compare two key conditions that control for whether the outcome would have been worse or not: the counterfactual failed-attempts condition and the semi-factual other-cause one. We suggest that people's moral judgements of failed attempts to harm are amplified when they imagine that the protagonist's intended harm could have occurred. But an alternative explanation is that their moral judgements are amplified when they imagine a worse outcome. In the failed-attempts stories, the real outcome is that no harm occurred and the counterfactual imagines an alternative in which the protagonist's intended harm could have occurred (her partner could have been burned) and so the outcome would have been worse (the harm would have occurred). The other-cause stories allow us to separate these two factors. In the other-cause stories, the real outcome is that harm occurred (by other means), and the semi-factual imagines an alternative in which the protagonist's intended harm could have occurred (her partner could have been

burned) and the outcome would have been the same (the harm occurred anyway by other means). The comparison of these two conditions allows us to examine the effects of imagining that the protagonist's intended harm occurred (her partner was burned), controlled for whether the outcome would have been worse or not. We suggest that it is the former that amplifies moral judgements and so we predict no differences between these two conditions.

Second, we can compare two key conditions that control for "pivotality": the semi-factual failed-attempts condition and the semi-factual other-cause one. One alternative explanation for the difference between the counterfactual and semi-factual conditions in the previous experiments is the protagonist's "pivotality": protagonists are judged to be more causally responsible for an outcome if it could not have occurred without their input (e.g., Lagnado et al., 2013). The semi-factual in the failed-attempts story conveys that the protagonist was not pivotal and she could not have changed the outcome, and moreover, she could never have caused harm: it was inevitable that Jenny's partner's hand would *not* be burned. But the semi-factual in the other-cause condition conveys that the protagonist was not pivotal and she could not have changed the outcome (in this instance, the harm was caused by Jenny's partner lifting a different piece of hot metal), but she *could* have caused harm. These conditions allow us to disentangle the role of the protagonist's potential to cause harm from the role of the protagonist's causal pivotality. We suggest that it is the former that amplifies moral judgements and so we predict that protagonists will be judged more harshly in the semi-factual other-cause harm condition than the semi-factual failed attempts condition.

We examined people's moral judgements about failed attempts to harm, this time omitting the assertion that focused on the facts, e.g., "The metal Jenny welded transmitted heat quickly".

Method

Participants and procedure

The participants were 26 university students at Trinity College Dublin who participated for course credit. There were 14 women and 12 men, aged between 18 and 36 years with a mean age of 23 years. The procedure was the same as the previous experiment.

Design and materials

Participants acted as their own controls in a 2 (outcome: failed attempt vs. other cause) x 2 (alternative outcome: counterfactual vs. semi-factual) design. Participants received 16 scenarios with different contents, 4 instances in each of the 4 conditions. Each story contained a counterfactual alternative, or a

semi-factual alternative. Participants also received eight filler stories in which no counterfactual or semi-factual alternative was mentioned. The contents were assigned to conditions using a Latin square design, and the stories were presented in a different random order to each participant. The measures were the same as the previous experiment.

Results and discussion

We carried out four within-participant ANOVAs with the design of 2 (imagined alternative: counterfactual vs. semi-factual) \times 2 (outcome: failed attempt vs. other-cause). We report planned contrasts on the interactions to test the hypothesis of a difference between counterfactuals and semi-factuals for failed attempts, and not for other-causes (see Figure 3).

Punishment latencies and judgements. Analyses of log-transformed response times revealed that, in failed attempt scenarios, participants read the punishment statement “Jenny deserves to be punished” more quickly for a failed attempt when it was primed by a counterfactual ($M = 1367.67$ msec, $SD = 495.52$) than a semi-factual ($M = 1660.53$ msec, $SD = 731.66$), $F(1, 25) = 6.38$, $p = .018$, $\eta p^2 = .2$. There were no differences for other-cause stories between the counterfactual ($M = 1551.56$ msec, $SD = 554.36$) and semi-factual conditions ($M = 1527.06$ msec, $SD = 973.06$), $F < 1$, as shown by contrasts on the interaction, $F(1, 25) = 4.89$, $p = .036$, $\eta p^2 = .16$. There were no main effects of either variable, $F < 1$. Participants also judged the protagonist’s behaviour to be deserving of punishment more when they read a counterfactual than a semi-factual for a failed attempt, $F(1, 25) = 5.62$, $p = .026$, $\eta p^2 = .18$, and there was no effect for other-cause stories, $F < 1$, as shown by planned contrasts on the interaction, $F(1, 25) = 4.78$, $p = .038$, $\eta p^2 = .16$, in the ANOVA conducted on the punishment judgements. There were no main effects of outcome, $F < 1$, or imagined alternative, $F(1, 25) = 1.33$, $p = .26$, as Figure 3 shows.

Blame judgements. There was a main effect of outcome, as individuals were blamed more for failed attempts than other-cause stories, $F(1, 25) = 4.58$, $p = .042$, $\eta p^2 = .16$, there was no main effect of imagined alternative $F < 1$, and the variables did not interact, $F(1, 25) = 3.8$, $p = .063$, as Figure 3 shows.

Wrongness judgements. There were no differences in judgements of how wrong the protagonist’s behaviour was between counterfactuals and semi-factuals for failed attempts $F(1, 25) = 3.72$, $p = .065$, or other-cause scenarios $F < 1$, as shown by planned contrasts on the interaction, $F(1, 25) = 4.4$, $p = .046$, $\eta p^2 = .15$. The interaction arises because participants judge the protagonist’s behaviour to be more wrong when they consider a semi-factual in the other-cause condition than the failed-attempts condition, $F(1, 25) = 6.63$, $p = .016$, ηp^2

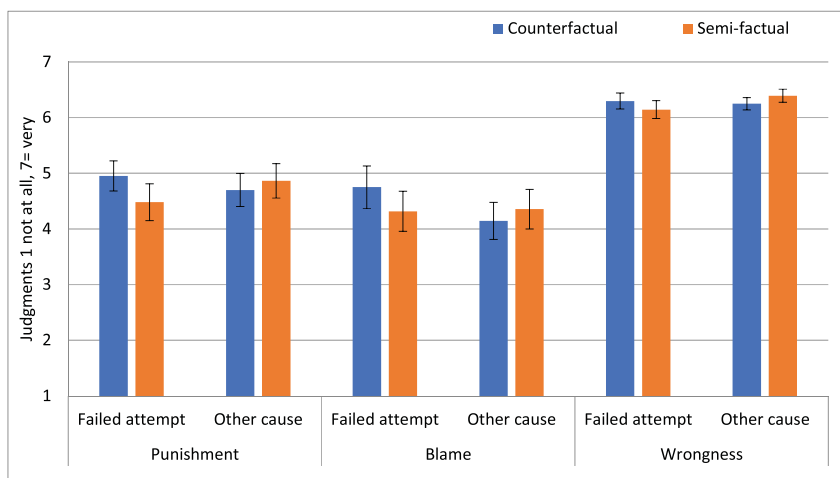


Figure 3. Moral judgements in response to “How much should Jenny be punished?”, “How much blame does Jenny deserve?” and “How wrong was Jenny’s behaviour?” in Experiment 3. Error bars are standard error of the mean.

= .21. There was no main effect of outcome $F(1, 25) = 1.48, p = .24$ or imagined alternative $F < 1$, as Figure 3 shows.

To rule out the possibility that individuals were merely judged more harshly when the imagined outcome was worse than reality, rather than when the imagined outcome corresponded to the harm that the individual had intended, we carried out additional planned contrasts to compare two key conditions: the counterfactual failed attempts condition and the semi-factual other-cause condition. As we predicted, there was no difference between these two key conditions for latencies of punishment $F < 1$, judgements of punishment, $F < 1$, blame $F(1, 25) = 2.92, p = .1$, or wrongness $F(1, 25) = 1.6, p = .22$.

To rule out the possibility that individuals were judged more harshly because of their causal “pivotality” rather than their potential to harm, we carried out additional planned contrasts to compare two further conditions: the semi-factual failed attempts condition, and the semi-factual other-cause condition. As we predicted, protagonists were judged more severely in the semi-factual other-cause condition than the semi-factual failed attempts condition for judgements of punishment, $F(1, 25) = 6.85, p = .015, \eta^2 = .22$, and wrongness $F(1, 25) = 6.63, p = .016, \eta^2 = .21$, although not for punishment latencies, or blame, $F < 1$ in each case.

The results show that for failed attempts to harm, participants judge a protagonist more harshly when they are primed by a counterfactual than a semi-factual, and no such effects occur for failed attempts to harm when the harm occurs by other causes. Importantly, participants judged a protagonist more

harshly when the intended harm could have occurred regardless of whether it was worse than reality. Also, they judged a protagonist more harshly when the protagonist could have caused harm even when they were not causally pivotal. Overall, the results support the hypothesis that people make harsher moral judgements about a protagonist's failed attempt to harm when a counterfactual about how it could have been worse highlights the role of factors outside the protagonist's control.

Notably, the counterfactuals affected the time participants took to read punishment assertions and their judgements of punishment, but they did not affect their judgements of blame or wrongness in this experiment. One possibility is that participants interpret the scope of the question differently for the different measures. Accordingly, we examine the scope of judgements in Experiment 4.

Experiment 4

The aim of the experiment was to compare general judgements, e.g., "How much blame does Jenny deserve?" to specific judgements, e.g., "How much blame does Jenny deserve *for her attempt to harm her partner?*" (see [Table 1](#) for examples). Participants blame a protagonist more for a failed attempt that results in no harm compared to a failed attempt in which harm occurs by other means, even though they judge the protagonist's behaviour to be wrong in both situations (e.g., Cushman, 2008). The experiment tests two alternative possible explanations. One possibility is that the effect arises because participants consider that the individual deserves blame for the attempt, but not for the causally unrelated harmful outcome. The general judgement, e.g., "How much blame does Jenny deserve?" is interpreted to refer to blame for the *attempt* to harm when the attempt fails and results in no harm, but it may be interpreted to refer to the *outcome* of harm when the attempt to harm fails but harm occurs through other means (e.g., Malle et al., 2014). In contrast, the judgement "How wrong was Jenny's behaviour?" unambiguously refers to the act, rather than to the outcome, and therefore judgements are similar whether no harm occurs, or a harmful outcome occurs via another means. This explanation predicts that it is only for general judgements that participants will blame a protagonist more for a failed attempt that results in no harm compared to a failed attempt in which harm occurs by other means, and yet judge the protagonist's behaviour to be wrong in both situations; the difference will not occur for specific judgements, e.g., "How much blame does Jenny deserve *for her attempt to harm her partner?*". An alternative possibility is that judgements that an action deserves blame and punishment depend on reasoning about the cause of the outcome prompted by the bad outcome, whereas judgements that an action is wrong do not require thinking about the outcome (Cushman, 2008). This explanation

predicts that for both general and specific judgements, participants will blame a protagonist more for a failed attempt that results in no harm compared to a failed attempt in which harm occurs by other means, and yet judge the protagonist's behaviour to be wrong in both situations.

Method

Participants and procedure

The participants were 68 volunteers, 51 men and 17 women, aged 17–55 years, with a mean age of 32 years, recruited from the online platform, Crowdfunder (see Crowdfunder.com). The 68 participants were those who remained when 33 participants were eliminated prior to any data analysis on the basis of the same criteria used to screen participants as outlined in Experiment 1, and in addition, participants were given a prompt to generate counterfactuals relating to a story provided at the end of the experiment, and those who failed to provide an appropriate response were eliminated. The materials were presented via SurveyGizmo (see <http://www.surveymzmo.com/>), which participants were directed to from Crowdfunder. Participants were compensated \$.30 for their time. They were asked to read each story carefully and to answer the questions in the order in which they were asked. They were told there were no right or wrong answers. They performed a practice trial at the outset. The experiment took approximately 10 minutes to complete.

Materials and design

Participants were assigned at random to the general judgement group or the specific judgement group. They received eight stories in a 2 (outcome: failed attempt vs. other-cause) x 2 (judgement focus: general vs. specific) design with repeated measures on the first factor. We assigned the contents randomly, in eight combinations of contents; participants were assigned at random to receive one set so that each participant received either the failed attempt or other-cause version of each story, and no participant received the same content for both. Each participant received the eight stories in a different randomised order.

Participants in the general judgement group completed the same three target judgements as the previous experiments. Participants in the specific judgement group completed three judgements as follows: (1) "How much should Jenny be punished for her attempt to harm her partner?", (2) "How much blame does Jenny deserve for attempting to harm her partner?", (3) "How wrong was Jenny's attempt to harm her partner?". They provided their judgements on seven-point scales with 1 anchored as "not at all", 4 as "somewhat" and 7 as "very much". Participants also generated counterfactuals after they had completed their judgements but there were an insufficient number to be analysed and so we do not report them here.

Results and discussion

We carried out three ANOVAs with the design of a 2 (outcome: failed attempt vs. other-cause) x 2 (judgement focus: general vs. specific) with repeated measures on the first factor (see Figure 4).

Punishment judgements. There was a main effect of judgement focus, as the protagonist was judged to be more deserving of punishment for specific judgements rather than general ones, $F(1, 66) = 16.3, p < .001, \eta p^2 = .2$. There was a main effect of outcome, as protagonists were judged more harshly for failed attempts than for other-caused harm, $F(1, 66) = 4.93, p < .03, \eta p^2 = .07$, as Figure 4 shows. The two variables did not interact $F < 1$.

Blame judgements. There was a main effect of judgement focus, as the protagonist was judged to be more deserving of blame for specific judgements rather than general ones, $F(1, 66) = 11.93, p < .001, \eta p^2 = .15$. There was a main effect of outcome, as protagonists were judged more harshly for failed attempts than for other-caused harm, $F(1, 66) = 14.36, p = .008, \eta p^2 = .18$ as Figure 4 shows. The two variables did not interact $F < 1$.

Wrongness judgements. There was no main effect of judgement focus $F < 1$, or outcome, $F(1, 66) = 1.98, p = .16$, and the two variables did not interact $F < 1$, as Figure 4 shows.

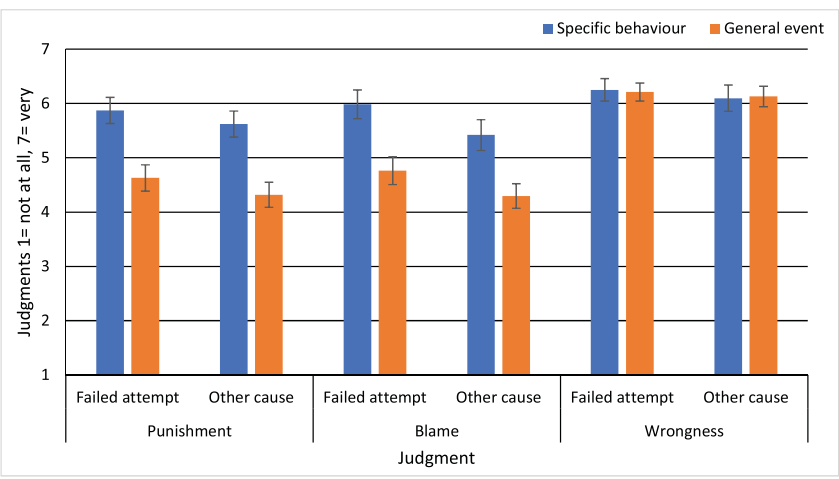


Figure 4. Moral judgements in response to “How much should Jenny be punished/for her attempt to harm her partner?”, “How much blame does Jenny deserve/ for attempting to harm her partner?” and “How wrong was Jenny’s behaviour/attempt to harm her partner?” in Experiment 4. Error bars are standard error of the mean.

The results show that participants judged that protagonists should be punished and blamed more for a failed attempt than other-caused harm, not only for general judgements but also for specific judgements; however, they judged protagonists' actions to be wrong equally for a failed attempt and other-caused harm. The experiment rules out the explanation that the tendency to blame a protagonist more for a failed attempt that results in no harm compared to a failed attempt in which harm occurs by other means occurs only because of a confound in the focus of the judgement. Instead, it is consistent with the explanation that people make judgements of blame and punishment by reasoning about the cause of the outcome, whereas when they make judgements that an action is wrong they may not think as much about the outcome.

General discussion

People make harsher moral judgements of punishment, blame and wrongness about a protagonist who engaged in a failed attempt to harm another person when they think about how things could have been worse, e.g., "if Jenny's partner had not let go of the metal, her hand would have been burned" compared to when they think about how things could have turned out the same, e.g., "even if Jenny's partner had not let go of the metal, her hand would not have been burned", and compared to when they think only about the facts, as the first experiment showed. The moral counterfactual amplification effect occurred when the protagonist was described as having no reason, but not when the protagonist was described as having an enduring reason. People's reasons for their actions play a key role in moral judgements (e.g., Malle et al., 2014; Reeder, Monroe, & Pryor, 2008). In many situations, people know or infer the reasons why an actor carried out an action (e.g., Baron & Hershey, 1988). The observation that the moral counterfactual amplification effect occurs only when the actor's reasons are unknown suggests important limitations to the effect. People rely on counterfactuals to help them to identify causes and intentions (e.g., Alicke, Mandel, Hilton, Gerstenberg, & Lagnado, 2015; McEleney & Byrne, 2006; Roesse & Epstude, *in press*; Walsh & Byrne, 2007). Hence counterfactuals may amplify moral judgements when the actor's reason is unknown because they help to identify the actor's causal and intentional role. When the actor's reasons are known, the role of counterfactuals may be less prominent. One fruitful avenue for future research is to examine situations in which an actor's motivations are ambiguous.

Counterfactuals amplify moral judgements for accidental near-misses as well as failed attempts to harm, as the second experiment showed. Of course people judge a protagonist more harshly for a failed attempt to harm than an accidental near-miss, but in both cases, thoughts about how things could

have been worse led to harsher punishment, blame and wrongness judgements, compared to thoughts about how things could have turned out the same. Counterfactuals amplify moral judgements for failed attempts to harm, but not when the harm occurs anyway by another cause, as the third experiment showed. Participants judge a protagonist more harshly when the intended harm could have occurred regardless of whether it is worse than reality, and even when the protagonist is not causally pivotal. The tendency to judge failed attempts that result in no harm more harshly than failed attempts that result in harm caused by other means reflects reasoning about the cause of the outcome, as Experiment 4 shows.

When an individual tries to harm someone and fails, it may be reasonable to rely on knowledge about the outcome to make inferences with hindsight about the likelihood of the outcome or the person's reasons for their actions (e.g., Baron & Hershey, 1988; Fleischhut et al., 2017; Lench et al., 2015). But it can also seem unfair that an individual who tries to harm someone is punished and blamed less if their attempt fails than if it succeeds, for reasons outside their control (e.g., Young et al., 2010). Counterfactual thoughts about how things could have been worse can "de-bias" judgements of various sorts (e.g., Hirt, Kardes, & Markman, 2004; Kray & Galinsky, 2003). The experiments show that when people think about how things could have been worse they judge that an individual should be punished and blamed for an attempt to harm that fails, more than when they think only about what happened or when they think about how things could have turned out the same.

The results support the explanation that people make harsher moral judgements about failed attempts to harm when thoughts about how things could have been worse highlight the factors outside the protagonist's control which prevented the harm from occurring. Counterfactual thoughts about how things could have been worse focus people's attention on the harm that almost happened, and so people make harsher moral judgements. The result rules out the proposal that thoughts about how things could have been worse focus people's attention on the good outcome that actually happened, which leads them to make lenient moral judgements about failed attempts to harm (pace; Kadish, 1993). Of course, counterfactual thoughts about how things could have been worse often help people to feel relief and solace, e.g., the thought that a person who is injured in a car crash could have died may provide some comfort (e.g., Larsen et al., 2004; Markman, Gavanski, Sherman, & McMullen, 1993; Medvec, Madey, & Gilovich, 1995; Sweeny & Vohs, 2012) and in some circumstances such thoughts can lead to diminished blame (e.g., Nario-Redmond & Branscombe, 1996). But moral judgements about failed attempts to harm stand in stark contrast to this backdrop: counterfactual thoughts about how things could have been worse lead to increased blame and punishment.

Acknowledgments

The research was supported by a PhD scholarship from the Irish Research Council (Humanities and Social Sciences) to the first author. We thank Maedhbh Etain Kelly, Jessica Dully, Jennifer Barry and Michael Desmond Collins for help with coding the data. Some of the results were presented at the European Society for Cognitive Psychology annual conference in Cyprus in 2015 and at the Interdisciplinary Workshop on Counterfactual Thinking in Toronto in 2016.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Irish Research Council for the Humanities and Social Sciences.

References

- 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(10), 1371–1381.
- Alicke, M. D., Mandel, D. R., Hilton, D. J., Gerstenberg, T., & Lagnado, D. A. (2015). Causal conceptions in social explanation and moral evaluation: A historical tour. *Perspectives on Psychological Science*, 10(6), 790–812.
- Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, 54(4), 569.
- Boninger, D. S., Gleicher, F., & Strathman, A. (1994). Counterfactual thinking: From what might have been to what may be. *Journal of Personality and Social Psychology*, 67(2), 297–307.
- Branscombe, N. R., Owen, S., Garstka, T. A., & Coleman, J. (1996). Rape and accident counterfactuals: Who might have done otherwise and would it have changed the outcome? *Journal of Applied Social Psychology*, 26(12), 1042–1067.
- Byrne, R. M. (2016). Counterfactual thought. *Annual Review of Psychology*, 67, 135–157.
- Byrne, R. M., & Tasso, A. (1999). Deductive Reasoning with factual, possible and counterfactual conditionals. *Memory and Cognition*, 27(4), 726–740.
- Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge: MIT Press.
- Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, 108(2), 353–380.
- Cushman, F. (2013). Action, outcome, and value a dual-system framework for morality. *Personality and Social Psychology Review*, 17(3), 273–292.
- Domskey, D. (2004). There is no door. *The Journal of Philosophy*, 101(9), 445–464.
- Enoch, D., & Marmour, A. (2007). The case against moral luck. *Law and Philosophy*, 26(4), 405–436.
- Epstude, K., & Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review*, 12(2), 168–192.
- Feinberg, J. (1995). Equal punishment for failed attempts: Some bad but instructive arguments against it. *Arizona Law Review*, 37, 117–133.

- Ferrante, D., Giroto, V., Stragà, M., & Walsh, C. (2013). Improving the past and the future: A temporal asymmetry in hypothetical thinking. *Journal of Experimental Psychology: General*, 142(1), 23–27.
- Fleischhut, N., Meder, B., & Gigerenzer, G. (2017). Moral hindsight. *Experimental Psychology*, 64, 110–123.
- Frosch, C. A., & Byrne, R. M. (2012). Causal conditionals and counterfactuals. *Acta Psychologica*, 141(1), 54–66.
- Gavanski, I., & Wells, W. L. (1989). Counterfactual processing of normal and exceptional events. *Journal of Experimental Social Psychology*, 25(4), 314–325.
- Hart, H. L. A., & Honore, A. M. (1959). *Causation in the law*. Oxford: Clarendon Press.
- Hirt, E. R., Kardes, F. R., & Markman, K. D. (2004). Activating a mental simulation mind-set through generation of alternatives: Implications for debiasing in related and unrelated domains. *Journal of Experimental Social Psychology*, 40(3), 374–383.
- Kadish, S. H. (1993). The criminal law and the luck of the draw. *Journal of Criminal Law Criminology and Police Science*, 84, 679–702.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93(2), 136.
- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York, NY: Cambridge University Press.
- Knobe, J. (2010). Person as scientist person as moralist. *Behavioral and Brain Sciences*, 33(4), 315–329.
- Kray, L. J., & Galinsky, A. D. (2003). The debiasing effect of counterfactual mind-sets: Increasing the search for disconfirmatory information in group decisions. *Organizational Behavior and Human Decision Processes*, 91(1), 69–81.
- Lagnado, D. A., Gerstenberg, T., & Zultan, R. I. (2013). Causal responsibility and counterfactuals. *Cognitive Science*, 37(6), 1036–1073.
- Larsen, J. T., McGraw, A. P., Mellers, B. A., & Cacioppo, J. T. (2004). The agony of victory and thrill of defeat mixed emotional reactions to disappointing wins and relieving losses. *Psychological Science*, 15(5), 325–330.
- Lench, H. C., Domsy, 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., & McMullen, M. N. (2003). A reflection and evaluation model of comparative thinking. *Personality and Social Psychology Review*, 7(3), 244–267.
- Markman, K. D., Gavanski, I., Sherman, S. J., & McMullen, M. N. (1993). The mental simulation of better and worse possible worlds. *Journal of Experimental Social Psychology*, 29(1), 87–109.
- Markman, K. D., Mizoguchi, N., & McMullen, M. N. (2008). “It would have been worse under Saddam:” Implications of counterfactual thinking for beliefs regarding the ethical treatment of prisoners of war. *Journal of Experimental Social Psychology*, 44(3), 650–654.
- Martin, J. W., & Cushman, F. (2016). Why we forgive what can’t be controlled. *Cognition*, 147, 133–143.
- McCloy, R., & Byrne, R. M. (2002). Semifactual “even if” thinking. *Thinking & Reasoning*, 8(1), 41–67.
- McEleney, A., & Byrne, R. M. (2006). Spontaneous counterfactual thoughts and causal explanations. *Thinking and Reasoning*, 12(2), 235–225.
- McMullen, M. N., & Markman, K. D. (2000). Downward counterfactuals and motivation: The wake-up call and the Pangloss effect. *Personality and Social Psychology Bulletin*, 26(5), 575–584.

- McMullen, M. N., & Markman, K. D. (2002). Affective impact of close counterfactuals: Implications of possible futures for possible pasts. *Journal of Experimental Social Psychology*, 38(1), 64–70.
- Medvec, V. H., Madey, S. F., & Gilovich, T. (1995). When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality and Social Psychology*, 69(4), 603.
- Miller, R. M., Hannikainen, I. A., & Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion*, 14(3), 573.
- Monroe, A. E., Dillon, K. D., & Malle, B. F. (2014). Bringing free will down to earth: People's psychological concept of free will and its role in moral judgment. *Consciousness and Cognition*, 27, 100–108.
- Nagel, T. (1993). Moral luck. In D. Statman (Ed.), *Moral luck* (pp. 57–71). Albany: State University of New York Press.
- Nario-Redmond, M. R., & Branscombe, N. R. (1996). It could have been better or it might have been worse: Implications for blame assignment in rape cases. *Basic and Applied Social Psychology*, 18(3), 347–366.
- Nelkin, D. K. (2004). Moral luck. In E. M. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Stanford, CA: Stanford University.
- Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science*, 39(1), 96–125.
- Nichols, S., Timmons, M., & Lopez, T. (2014). Using experiments in ethics–ethical conservatism and the psychology of moral luck. In *Empirically informed ethics: Morality between facts and norms* (pp. 159–176). Switzerland: Springer International Publishing.
- Phillips, J., Luguri, J. B., & Knobe, J. (2015). Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition*, 145, 30–42.
- Piaget, J. (1965). *The moral judgment of the child*. New York, NY: Free Press. (Originally published in 1932).
- Pizarro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise the role of perceived metadesires. *Psychological Science*, 14(3), 267–272.
- Reeder, G. D., Monroe, A. E., & Pryor, J. B. (2008). Impressions of Milgram's obedient teachers: Situational cues inform inferences about motives and traits. *Journal of Personality and Social Psychology*, 95(1), 1–17.
- Robinson, P. H., Goodwin, G. P., & Reisig, M. D. (2010). Disutility of injustice. *The New York University Law Review*, 85, 1940–2032.
- Roese, N., & Epstude, K. (in press). The functional theory of counterfactual thinking: New evidence, new challenges, new insights. *Advances in Experimental Social Psychology*. Advance online publication. doi:10.1016/bs.aesp.2017.02.001
- Santamaria, C., Espino, O., & Byrne, R. M. (2005). Counterfactual and Semifactual conditionals prime alternative possibilities. *Journal of Experimental Psychology: Learning Memory and Cognition*, 31(5), 1149–1154.
- Spellman, B. A., & Mandel, D. R. (1999). When possibility informs reality: Counterfactual thinking as a cue to causality. *Current directions in Psychological Science*, 8(4), 120–123.
- Sweeny, K., & Vohs, K. D. (2012). On near misses and completed tasks the nature of relief. *Psychological Science*, 23(5), 464–468.
- Teigen, K. H., & Jensen, T. K. (2011). Unlucky victims or lucky survivors? Spontaneous counterfactual thinking by families exposed to the tsunami disaster. *European Psychologist*, 16(1), 48–57.
- Thompson, V. A., & Byrne, R. M. (2002). Reasoning counterfactually: Making inferences about things that didn't happen. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 28(6), 1154–1170.

- Walsh, C. R., & Byrne, R. M. (2007). How people think “if only...” about reasons for actions. *Thinking & Reasoning*, 13(4), 461–483.
- Walster, E. (1966). Assignment of responsibility for an accident. *Journal of Personality and Social Psychology*, 3(1), 73–79.
- Williams, B. (1981). *Moral luck: Philosophical papers 1973-1980*. Cambridge: Cambridge University Press.
- Young, L., & Tsoi, L. (2013). When mental states matter, when they don’t, and what that means for morality. *Social and Personality Psychology Compass*, 7(8), 585–604.
- Young, L., Nichols, S., & Saxe, R. (2010). Investigating the neural and cognitive basis of moral luck: It’s not what you do but what you know. *Review of Philosophy and Psychology*, 1(3), 333–349.

Appendix

The 24 stories for the failed attempts to harm used in Experiments 1–4 are provided. The explanatory sentence about the facts in *italics* was used in Experiment 2 only. The words used to convey an accidental near-miss in Experiment 2 are given in parentheses. The sentences for the other-causes used in Experiment 3 are also provided. The first two stories were used in Experiment 1 and the reasons appeared in that experiment only.

(1) Burn (from Cushman)

Jenny is taking a class in sculpture. She is assigned to work with a partner to weld together pieces of metal.

REASON: Because of a long-standing feud,

NO REASON: For no reason at all,

Jenny wants (does not want) to burn her partner’s hand. Jenny believes that if she welds a piece of metal that her partner is holding the heat will (not) travel down the metal to burn her partner’s hand.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Jenny welds the metal, but her partner happens to let go and is not burned at all.

COUNTERFACTUAL: *The metal Jenny welded transmitted heat quickly.*

If Jenny’s partner had not let go of the metal that Jenny was welding, her hand would have been burned.

SEMI-FACTUAL: *The metal Jenny welded did not transmit heat quickly.*

Even if Jenny’s partner had not let go of the metal that Jenny was welding, her hand would not have been burned.

CONTROL: Jenny’s partner continues with her work, and everyone else carries on with their work.

OTHER CAUSE: Jenny welds the metal, but her partner happens to let go and is not burned by Jenny. However, Jenny’s partner picks up a different piece of hot metal and is burned.

COUNTERFACTUAL: If Jenny’s partner had not let go of the metal that Jenny was welding, her hand would not have been burned.

SEMI-FACTUAL: Even if Jenny’s partner had not let go of the metal that Jenny was welding, her hand would still have been burned.



(2) Weight (from Cushman)

Tom works at a construction site. He is holding a 50-pound steel beam that belongs on the ground below. Tom's supervisor is on the ground below, taking a break.

REASON: Because of a mutual dislike between Tom and his supervisor,

NO REASON: Without any clear motive,

Tom wants (does not want) to drop the beam on his supervisor and break his legs. Tom believes that his supervisor is not about to walk underneath the beam, and if he drops the beam it will (not) hit his supervisor or break his legs.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Tom drops the beam, the supervisor walks beneath the beam, and the beam happens to miss Tom's supervisor, who is just fine.

COUNTERFACTUAL: Tom is quite far above his supervisor. If the beam Tom dropped had not missed Tom's supervisor, his legs would have been broken.

SEMI-FACTUAL: Tom is not far above his supervisor. Even if the beam Tom dropped had not missed Tom's supervisor, his legs would not have been broken.

CONTROL: Tom's supervisor continues with his work, and everyone else carries on with their work.

OTHER CAUSE: Tom drops the beam, the supervisor walks beneath the beam, but the beam happens to miss Tom's supervisor. However, just then another beam happens to break loose from a pile to the side and rolls over the supervisor, breaking his legs.

COUNTERFACTUAL: If the beam Tom dropped had not missed Tom's supervisor, his legs would not have been broken.

SEMI-FACTUAL: Even if the beam Tom dropped had not missed Tom's supervisor, his legs would still have been broken.

(3) Burn, Chemical

Oran is taking a class in chemistry. He is assigned to work with a partner to combine solutions. Oran wants (does not want) to burn his partner's hand. Oran believes that if he pours some solution into the jar that his partner is holding some (none) of the solution will spill down the side of the jar to burn his partner's leg.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Oran pours the solution, but his partner happens to hold the jar away from his body and is not burned at all.

COUNTERFACTUAL: *The acid in the jar Oran held was very strong.* If Oran's partner had not held the jar away from his body, his hand would have been burned.

SEMI-FACTUAL: *The acid in the jar Oran held was very weak.* Even if Oran's partner had not held the jar away from his body, his hand would not have been burned.

OTHER CAUSE: Oran pours the solution, but his partner happens to hold the jar away from his body and is not burned by Oran. However Oran's partner spills solution from a different jar on his leg and is burned.

COUNTERFACTUAL: If Oran's partner had not held the jar away from his body, his hand would not have been burned.

SEMI-FACTUAL: Even if Oran's partner had not held the jar away from his body, his hand would still have been burned

(4) Burn, Cooking

Ann is taking a class in cooking. She is assigned to work with a partner to sauté some vegetables. Ann wants (does not want) to burn her partner's hand. Ann believes that if she moves the pan that her partner is holding the gas flames will (not) reach up around the pan to burn her partner's finger.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Ann moves the pan, but her partner happens to let go and is not burned at all.

COUNTERFACTUAL: *The gas flame can reach very high.* If Ann's partner had not let go of the pan Ann moved, her hand would have been burned.

SEMI-FACTUAL: *The gas flame cannot reach very high.* Even if Ann's partner had not let go of the pan Ann moved, her hand would not have been burned.

OTHER CAUSE: Ann moves the pan, but her partner happens to let go and is not burned by Ann. However her partner moves a different pan and is burned.

COUNTERFACTUAL: If Ann's partner had not let go of the pan Ann moved, her hand would not have been burned.

SEMI-FACTUAL: Even if Ann's partner had not let go of the pan Ann moved, her hand would still have been burned.

(5) Weight (from Cushman)

Jerry is studying in a library. He is leaning on a heavy bookshelf that is rocking precariously. Jerry's classmate is on the other side of the shelf, looking for a book. Jerry wants (does not want) to push the bookshelf onto his classmate and break his ribs. Jerry believes that his classmate is not about to walk behind the bookshelf, and that if he pushes the shelf it will (not) fall on his classmate and break his ribs.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Jerry pushes the bookshelf, his classmate walks behind the bookshelf, and the bookshelf happens to miss Jerry's classmate, who is just fine.



COUNTERFACTUAL: *The bookshelf Jerry pushed is very heavy.* If the bookshelf Jerry pushed had not missed Jerry's classmate, his ribs would have been broken.

SEMI-FACTUAL: *The bookshelf Jerry pushed is not very heavy.* Even if the bookshelf Jerry pushed had not missed Jerry's classmate, his ribs would not have been broken.

OTHER CAUSE: Jerry pushes the bookshelf, his classmate walks behind the bookshelf, and the bookshelf happens to miss Jerry's classmate. However, just then another bookshelf happens to topple from the other side and falls on Jerry's classmate, breaking his ribs.

COUNTERFACTUAL: If the bookshelf Jerry pushed had not missed Jerry's classmate, his ribs would not have been broken.

SEMI-FACTUAL: Even if the bookshelf Jerry pushed had not missed Jerry's classmate, his ribs would still have been broken.

(6) Weight, Keg

Alex works in a pub. He is holding a large steel keg that belongs in the cellar below. Another bartender is in the cellar below, taking a break. Alex wants (does not want) to throw the keg onto the other bartender and break his ribs. Alex believes that the other bartender is not about to walk underneath the cellar door, and that if he drops the keg it will (not) hit the other bartender and break his ribs.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Alex drops the keg, the other bartender walks beneath the keg, and the keg happens to miss the other bartender, who is just fine.

COUNTERFACTUAL: *The keg Alex dropped is full.* If the keg Alex dropped had not missed the other bartender, his ribs would have been broken.

SEMI-FACTUAL: *The keg Alex dropped is empty.* Even if the keg Alex dropped had not missed the other bartender, his ribs would not have been broken.

OTHER CAUSE: Alex drops the keg, the other bartender walks beneath the keg, and the keg happens to miss the other bartender, who is just fine. However, just then another keg happens to break loose from a pile to the side and fall onto the other bartender, breaking his ribs.

COUNTERFACTUAL: If the keg Alex dropped had not missed the other bartender, his ribs would not have been broken.

SEMI-FACTUAL: Even if the keg Alex dropped had not missed the other bartender, his ribs would still have been broken.

(7) Game (from Cushman)

Kevin is eating at a diner when a man challenges him to a game of darts. The man throws his darts well and gets a very high score. Kevin wants (does not want) to hit the man's hand with a dart and pierce it. Kevin believes that the man is (not) about to reach toward the dart

board to collect his darts, and that if he throws his dart he will not hit the man's hand or pierce it.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Kevin throws his dart, the man reaches out, but the dart happens to miss the man and hit the board harmlessly.

COUNTERFACTUAL: *The dart Kevin threw has a sharp steel tip.* If the dart Kevin threw had not missed the man, his hand would have been pierced.

SEMI-FACTUAL: *The dart Kevin threw has a dull plastic tip.* Even if the dart Kevin threw had not missed the man, his hand would not have been pierced.

OTHER CAUSE: Kevin throws his dart, the man reaches out, but Kevin's dart happens to miss the man and hit the board harmlessly. However, another player throws a dart and it pierces the man's hand.

COUNTERFACTUAL: If the dart Kevin threw had not missed the man, his hand would not have been pierced.

SEMI-FACTUAL: Even if the dart Kevin threw had not missed the man, his hand would still have been pierced.

(8) Game, Snooker

Barry is having lunch in a pub, when a man challenges him to a game of snooker. The man aims his shots well and begins winning the game. Barry wants (does not want) to hit the man's hand with a snooker ball and bruise it. Barry believes that the man is (not) about to reach toward the snooker table to collect his chalk, and that if he takes his shot he will not hit the man's hand and bruise it.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Barry takes his shot, the man reaches out, but the ball happens to miss the man and hit the side of the table harmlessly.

COUNTERFACTUAL: *Barry's shot travelled slowly.* If the shot Barry took had not missed the man, his hand would have been bruised.

SEMI-FACTUAL: *Barry's shot travelled fast.* Even if the shot Barry took had not missed the man, his hand would not have been bruised.

OTHER CAUSE: Barry takes his shot, the man reaches out, but the ball happens to miss the man and hit the side of the table harmlessly. However, another player takes a shot on their table and it bruises the man's hand.

COUNTERFACTUAL: If the shot Barry took had not missed the man his hand would not have been bruised.

SEMI-FACTUAL: Even if the shot Barry took had not missed the man, his hand would still have been bruised.

(9) Game, Soccer

Adam is playing on a soccer pitch when a man challenges him to a game of penalty shots. The man plays very well and gets a very high

number of goals. Adam wants (does not want) to hit the back of the man's head with the soccer ball and bruise it. Adam believes that the man is about to walk between the goalposts to collect his ball and that if he takes his shot he will (not) hit the man's head and bruise it.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Adam takes his shot, as the man reaches the goalposts, but the ball happens to miss the man and hit the net harmlessly.

COUNTERFACTUAL: *Adam's shot had strong force.* If the shot Adam took had not missed the man, his head would have been bruised.

SEMI-FACTUAL: *Adam's shot had weak force.* Even if the shot Adam took had not missed the man, his head would not have been bruised.

OTHER CAUSE: Adam takes his shot, as the man reaches the goalposts, but the ball happens to miss the man and hit the net harmlessly. However, another player takes a shot and it hits the man's head.

COUNTERFACTUAL: If the shot Adam took had not missed the man, his head would not have been bruised.

SEMI-FACTUAL: Even if the shot Adam took had not missed the man, his head would still have been bruised.

(10) Medical (from Cushman)

Bruce is a dentist filling in the cavity of his patient. He must drill into the patient's tooth just above a major nerve. Bruce wants (does not want) to hit the patient's nerve, (nor) to cause the patient excruciating pain. Bruce believes that if he switches the drill to a higher speed he will (not) hit the nerve or cause the pain.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Bruce switches the drill to a higher speed and misses the nerve. The patient undergoes no pain at all.

COUNTERFACTUAL: *The man's nerve is sensitive.* If Bruce's drill had not missed the nerve, the man would have been in pain.

SEMI-FACTUAL: *The man's nerve is not sensitive.* Even if Bruce's drill had not missed the nerve, the man would not have been in pain.

OTHER CAUSE: Bruce switches the drill to a higher speed and misses the nerve. The patient undergoes no pain at all as a result of Bruce, but just then a nurse hits a nerve on the other side of the patient's mouth and causes the patient pain.

COUNTERFACTUAL: If Bruce's drill had not missed the nerve, the man would have not have been in pain.

SEMI-FACTUAL: Even if Bruce's drill had not missed the nerve, the man would still have been in pain.

(11) Medical, Physiotherapist

Teresa is a physiotherapist massaging the back of her patient. She must massage the patient's back close to where the patient has a damaged

muscle. Teresa wants (does not want) to put pressure on the patient's damaged muscle, and (nor) to cause the patient excruciating pain. Teresa believes that if she rubs the patient's spine above the damaged muscle she will (not) hit the damaged muscle and cause the pain.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Teresa rubs the patient's spine just above the damaged muscle and misses the damaged muscle. The patient undergoes no pain at all.

COUNTERFACTUAL: *The damaged muscle is tender.* If Teresa had not missed the damaged muscle, the patient would have been in pain.

SEMI-FACTUAL: *The damaged muscle is numb.* Even if Teresa had not missed the damaged muscle, the patient would not have been in pain.

OTHER CAUSE: Teresa rubs the patient's spine just above the damaged muscle and misses the damaged muscle. The patient undergoes no pain at all as a result of Teresa, but just then a trainee physiotherapist puts pressure on the muscle from the lower back and causes the patient pain.

COUNTERFACTUAL: If Teresa had not missed the damaged muscle, the patient would not have been in pain.

SEMI-FACTUAL: Even if Teresa had not missed the damaged muscle, the patient would still have been in pain.

(12) Medical, acupuncturist

Robert is an acupuncturist inserting needles into the foot of his patient. He must insert the needle into the patient's foot just above a joint. Robert wants (does not want) to hit the patient's joint, (nor) to cause the patient excruciating pain. Robert believes that if he inserts the needle at an oblique angle he will (not) hit the joint or cause the pain.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Robert inserts the needle at an oblique angle and misses the joint. The patient undergoes no pain at all.

COUNTERFACTUAL: *The joint is highly sensitive.* If Robert's needle had not missed the joint, the man would have been in pain.

SEMI-FACTUAL: *The joint is insensitive.* Even if Robert's needle had not missed the joint, the man would not have been in pain.

OTHER CAUSE: Robert inserts the needle at an oblique angle and misses the joint. The patient undergoes no pain at all as a result of Robert, but just then an assistant hits a joint on the patient's other foot and causes the patient pain.

COUNTERFACTUAL: Robert's needle had not missed the joint, the man would have not been in pain.

SEMI-FACTUAL: Even if Robert's needle had not missed the joint, the man would still have been in pain.

(13) Cut (from Cushman)

Maria is a hair dresser cutting a customer's hair. The haircut is almost finished. There is only one more piece of hair to trim, and it is right beside the customer's ear. Maria wants (does not want) to cut off a piece of the customer's ear. Maria believes that if she trims the hair at a sharp angle she will (not) cut off a piece of the customer's ear.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Maria trims the hair at a sharp angle, but just then the customer sneezes and Maria misses the ear. The haircut is finished and the customer is perfectly fine.

COUNTERFACTUAL: *The scissors are very sharp.* If the customer had not sneezed, a piece of the customer's ear would have been cut off.

SEMI-FACTUAL: *The scissors are not very sharp.* Even if the customer had not sneezed, a piece of the customer's ear would not have been cut off.

OTHER CAUSE: Maria trims the hair at a sharp angle, but just then the customer sneezes and Maria misses the ear. While sneezing the customer knocks into the straight razor from a shaving kit, and a piece of the customer's ear is cut off.

COUNTERFACTUAL: If the customer had not sneezed, a piece of the customer's ear would not have been cut off.

SEMI-FACTUAL: Even if the customer had not sneezed, a piece of the customer's ear would still have been cut off.

(14) Cut, Fabric

Sarah is an assistant at a fabric shop cutting a customer's fabric. The fabric is almost finished. There is only one more piece of fabric to trim, and it is right beside the customer's hand. Sarah wants (does not want) to cut off a piece of the customer's finger. Sarah believes that if she trims the fabric at a sharp angle she will (not) cut off a piece of the customer's finger.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Sarah trims the fabric at a sharp angle, but just then the customer waves and Maria misses the finger. The fabric is finished and the customer is perfectly fine.

COUNTERFACTUAL: *The blade has no safety guard.* If the customer had not waved, a piece of the customer's finger would have been cut off.

SEMI-FACTUAL: *The blade has a safety guard.* Even if the customer had not waved, a piece of the customer's finger would not have been cut off.

OTHER CAUSE: Sarah trims the fabric at a sharp angle, but just then the customer waves and Sarah misses the finger. The fabric is finished and

the customer is perfectly fine. While waving the customer knocks into the Stanley knife from a tailoring kit, and a piece of the customer's finger is cut off.

COUNTERFACTUAL: If the customer had not waved, a piece of the customer's finger would not have been cut off.

SEMI-FACTUAL: Even if the customer had not waved, a piece of the customer's finger would still have been cut off.

(15) Cut, Gardening

Shane is a gardener mowing his employer's lawn. The lawn is almost finished. There is only one more piece of lawn to trim, and it is right beside the customer's foot. Shane wants (does not want) to cut off a piece of his employer's foot. Shane believes that if he jerks the lawnmower at a sharp angle he will (not) cut off a piece of his employer's foot.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Shane jerks the lawnmower at a sharp angle, but just then his employer steps away and Shane misses the foot. The lawn is finished and his employer is perfectly fine.

COUNTERFACTUAL: *The employer is wearing canvas shoes.* If the employer had not stepped away, a piece of the employer's foot would have been cut off.

SEMI-FACTUAL: *The employer is wearing sturdy boots.* Even if the employer had not stepped away, a piece of the customer's foot would not have been cut off.

OTHER CAUSE: Shane jerks the lawnmower at a sharp angle, but just then his employer steps away and Shane misses the foot. While stepping away his employer knocks into the saw from a toolkit and a piece of his employer's foot is cut off.

COUNTERFACTUAL: If the employer had not stepped away, a piece of the employer's foot would not have been cut off.

SEMI-FACTUAL: Even if the employer had not stepped away, a piece of the customer's foot would still have been cut off.

(16) Poison (from Cushman)

Steve, Ken and Pat are roommates. There is a rat in their apartment, and Ken made some cookies with rat poison to kill the rat before leaving for the weekend. Steve sees the cookies on the counter. Steve wants (does not want) Pat to eat the cookie and become very ill. Steve believes that Pat knows (not) to eat the cookie and will leave it untouched.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Steve hands a cookie to Pat, but Pat is distracted and leaves the cookie untouched. Pat is not ill at all, continues on with his day as usual and leaves for the weekend.

COUNTERFACTUAL: *There was a large amount of poison in the cookie Pat handed Steve.* If Pat had not left the cookie Steve gave him untouched, he would have been ill.

SEMI-FACTUAL: *There was only a small amount of poison in the cookie Pat handed Steve.* Even if Pat had not left the cookie Steve gave him untouched, he would not have been ill.

OTHER CAUSE: Steve hands a cookie to Pat, but Pat is distracted and leaves the cookie untouched. However, entirely on his own, Pat discovers the other cookies Ken baked and eats one, becoming very ill.

COUNTERFACTUAL: If Pat had not left the cookie Steve gave him untouched, he would not have been ill.

SEMI-FACTUAL: Even if Pat had not left the cookie Steve gave him untouched, he would still have been ill.

(17) Poison, detergent

Grace, Emma and James are colleagues. They share a canteen, and Grace leaves some dishwasher detergent in a jar next to the coffee, before leaving for the weekend. Emma sees the jar on the counter. Emma wants (does not want) poison James and make him very ill. Emma believes that the white powder in the jar is sugar, and that it will (not) make him ill.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Emma hands a cup of poisoned coffee to James, but James is distracted and leaves the coffee untouched.

COUNTERFACTUAL: *Emma didn't put in a lot of detergent.* If James had not left the coffee Emma gave him untouched, he would have been ill.

SEMI-FACTUAL: *Emma put in a lot of detergent.* Even if James had not left the coffee Emma gave him untouched he would not have been ill.

OTHER CAUSE: Emma hands a cup of poisoned coffee to James, but James is distracted and leaves the coffee untouched. However, entirely on his own, James discovers the jar and puts some powder in his coffee, becoming very ill.

COUNTERFACTUAL: If James had not left the coffee Emma gave him untouched, he would not have been ill.

SEMI-FACTUAL: Even if James had not left the coffee Emma gave him untouched he would still have been ill.

(18) Poison, rubbing alcohol

Darragh, Laura and Cormac are housemates. Darragh uses an alcohol rub for muscle aches, which he leaves on the kitchen counter in an unmarked bottle, before leaving for the weekend. Laura sees the bottle on the counter. Laura wants (does not want) Cormac to drink the shot

and become very ill. Laura thinks that the bottle contains vodka, and will (not) make Cormac ill.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Laura pours a shot of the rubbing alcohol for Cormac, but Cormac is distracted and leaves the shot untouched.

COUNTERFACTUAL: *Laura fully filled the shot glass.* If Cormac had not left the shot Laura gave him untouched, he would have been ill.

SEMI-FACTUAL: *Laura half-filled the shot glass.* Even if Cormac had not left the shot Laura gave him untouched he would not have been ill.

OTHER CAUSE: Laura pours a shot of the rubbing alcohol for Cormac, but Cormac is distracted and leaves the glass untouched. However, entirely on his own, Cormac discovers the bottle later and drinks a shot of the rubbing alcohol, becoming very ill.

COUNTERFACTUAL: If Cormac had not left the shot Laura gave him untouched, he would not have been ill.

SEMI-FACTUAL: Even if Cormac had not left the shot Laura gave him untouched he would still have been ill.

(19) Swing (from Cushman)

John is an adult man walking through a carnival. He comes to a stand where you can punch a mechanical target to win a prize. The owner of the stand happens to be squatting beneath the target. John wants (does not want) to hit the owner and break his nose. John believes that the owner is (not) about to stand up, and that if he punches towards the target he will not hit the owner or break his nose.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: John punches towards the target, the owner stands up, and John happens to miss the owner, who is fine.

COUNTERFACTUAL: *John's arms are strong.* If John had not missed the owner, the owner's nose would have been broken.

SEMI-FACTUAL: *John's arms are weak.* Even if John had not missed the owner, the owner's nose would not have been broken.

OTHER CAUSE: John punches towards the target, the owner stands up, and John happens to miss the owner. However, another customer hits the owner and breaks his nose.

COUNTERFACTUAL: If John had not missed the owner, the owner's nose would not have been broken.

SEMI-FACTUAL: Even if John had not missed the owner, the owner's nose would still have been broken.

(20) Swing: golf

Eoin is an adult man playing golf at a driving range. He comes to a spot where you can take practise swings. An employee of the range happens to be standing close behind the spot.



Eoin wants (does not want) to hit the employee and break his shin. Eoin believes that the employee is (not) about to step closer, and that if he takes a practise swing he will not hit the employee and break his shin.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Eoin takes his practise swing, the employee steps closer, and Eoin happens to miss the employee, who is fine.

COUNTERFACTUAL: *Eoin is very strong.* If Eoin had not missed the employee, the employee's shin would have been broken.

SEMI-FACTUAL: *Eoin is not very strong.* Even if Eoin had not missed the employee, the employee's shin would not have been broken.

OTHER CAUSE: Eoin takes his practise swing, the employee steps closer, and Eoin happens to miss the employee, who is fine. However, another customer hits the employee and breaks his shin.

COUNTERFACTUAL: If Eoin had not missed the employee, the employee's shin would not have been broken.

SEMI-FACTUAL: Even if Eoin had not missed the employee, the employee's shin would still have been broken.

(21) Swing, fair

Dave is an adult man walking through a town fair. He comes to a stand where you can swing a rubber mallet to hit a target on the ground to win a prize. The owner of the stand happens to have his foot next to the target. Dave wants (does not want) to hit the owner and break his toes. Dave believes that the owner is (not) about to move his foot onto the target, and that if he punches towards the target he will (not) hit the owner and break his toes.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Dave swings the mallet towards the target, the owner moves his foot, and Dave happens to miss the owner, who is fine.

COUNTERFACTUAL: *The mallet is coated in steel.* If Dave had not missed the owner, the owner's toes would have been broken.

SEMI-FACTUAL: *The mallet is coated in rubber.* Even if Dave had not missed the owner, the owner's toes would not have been broken.

OTHER CAUSE: Dave swings the mallet towards the target, the owner moves his foot, and Dave happens to miss the owner. However, another customer hits the owner and breaks his toes.

COUNTERFACTUAL: If Dave had not missed the owner, the owner's toes would not have been broken.

SEMI-FACTUAL: Even if Dave had not missed the owner, the owner's toes would still have been broken.

(22) Trip (from Cushman)

Amy is sitting in a crowded train station. She is about to put her feet up on the seat across from her to relax. Just then she notices a passenger

rushing to catch his train. Amy wants (does not want) the passenger to trip and twist his ankle. Amy believes that the passenger is (not) going to run past her, and that if she puts her feet up on the seat across from her the passenger will (not) trip over her or twist his ankle.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Amy puts her feet up and the passenger happens to run by without tripping at all.

COUNTERFACTUAL: *The floors of the station are marble.* If the passenger had not run by Amy without tripping, the passenger's ankle would have been twisted.

SEMI-FACTUAL: *The floors of the station are carpeted.* Even if the passenger had not run by Amy without tripping, the passenger's ankle would not have been twisted.

OTHER CAUSE: Amy puts her feet up and the passenger happens to run by without tripping at all. However, for different reasons, the passenger slips on something else further along and twists his ankle.

COUNTERFACTUAL: If the passenger had not run by Amy without tripping, the passenger's ankle would not have been twisted.

SEMI-FACTUAL: Even if the passenger had not run by Amy without tripping, the passenger's ankle would still have been twisted.

(23) Trip, umbrella

Fiona is sitting in a crowded coffee shop. She is about to lean her umbrella on the seat across from her. Just then she notices a waitress rushing to get to the kitchen. Fiona wants (does not want) the waitress to trip and twist her ankle. Fiona believes that the waitress is (not) going to run past her, and that if she leans her umbrella against the seat across from her the waitress will (not) trip over her umbrella and twist her ankle.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Fiona leans her umbrella on the seat opposite and the waitress happens to run by without tripping at all.

COUNTERFACTUAL: *There was nothing to block her fall.* If the waitress had not run by Fiona without tripping, the waitress' ankle would have been twisted.

SEMI-FACTUAL: *There was a booth to block her fall.* Even if the waitress had not run by Fiona without tripping, the waitress' ankle would not have been twisted.

OTHER CAUSE: Fiona leans her umbrella on the seat opposite and the waitress happens to run by without tripping at all. However, for different reasons, the waitress slips on something else further along and twists her ankle.

COUNTERFACTUAL: If the waitress had not run by Fiona without tripping, the waitress' ankle would not have been twisted.

SEMI-FACTUAL: Even if the waitress had not run by Fiona without tripping, the waitress' ankle would still have been twisted.

(24) Trip, Lead

Cian is sitting in a crowded office. He is about to plug in an extension lead to the set of sockets across from him. Just then he notices a co-worker rushing to leave the office. Cian wants (does not want) the passenger to trip and twist his ankle. Cian believes that his co-worker is (not) going to run past him, and that if plugs in the extension lead into a socket across from him, his co-worker will (not) trip over the lead and break his nose.

FAILED ATTEMPT/NEAR-MISS ACCIDENT: Cian plugs in the extension lead and the co-worker happens to hurry by without tripping at all.

COUNTERFACTUAL: *There was a hard desk in front of him.* If the co-worker had not run by Cian without tripping, the co-worker's ankle would have been twisted.

(25) SEMI-FACTUAL: *There was a soft couch in front of him.* Even if the co-worker had not run by Cian without tripping, the co-worker's ankle would have not been twisted.

OTHER CAUSE: Cian plugs in the extension lead and the co-worker happens to hurry by without tripping at all. However, for different reasons, the co-worker trips on something else further along and breaks his nose.

COUNTERFACTUAL: If the co-worker had not run by Cian without tripping, the co-worker's ankle would not have been twisted.

SEMI-FACTUAL: Even if the co-worker had not run by Cian without tripping, the co-worker's ankle would still have been twisted.