A Causal Proximity Effect in Moral Judgment

Neele Engelmann (neele.engelmann@uni-goettingen.de)

Michael R. Waldmann (michael.waldmann@bio.uni-goettingen.de)
Department of Psychology
University of Göttingen, Germany

Abstract

In three experiments (total *N* = 1302) we investigated whether causal proximity affects moral judgments. We manipulated causal proximity by varying the length of chains mediating between actions and outcomes, and by varying the strengths of causal links. We demonstrate that moral judgments are affected by causal proximity with longer chains or weaker links leading to more lenient moral evaluations. Moreover, we identify outcome foreseeability as the crucial factor linking causal proximity and moral judgments. While effects of causal proximity on moral judgments were small when controlling for factors that were confounded in previous studies, knowledge about the presence of causal links substantially alters judgments of permissibility and responsibility. The experiments demonstrate a tight coupling between causal representations, inferences about mental states, and moral reasoning.

Keywords: Causal Reasoning; Moral Judgment; Causal Proximity; Causal Strength; Causal Chains

Suppose someone is contemplating an action that, as an unintended side effect, could cause serious harm to another person. For example, imagine a doctor in an emergency situation who has to decide which one of two live-saving drugs to administer to an unconscious patient. Both drugs will have the same stabilising effect on the patient, but both also have a risk of causing blood clots as a side effect. The exact probabilities are unknown, but the doctor remembers that when drug A causes blood clots, it happens via several intermediate steps. Drug A first needs to cause a number of intermediate events in the body before blood clots can develop. By contrast, when drug B causes blood clots, it does so directly. Which drug should the doctor choose?

If you prefer drug A, your preference may be an instance of a so-called causal proximity effect. Causal proximity refers to the position of a target cause (such as an action) relative to a target effect (such as a harmful outcome). A cause is traditionally called more proximate when fewer intermediate events connect it to a target outcome. Arguably, a cause may also be perceived as more proximate when its link to the effect is stronger, as spatio-temporal co-occurrence and causal strength tend to be correlated. We will explore both facets of proximity here.

It has been suggested that the length of a causal chain matters for our moral evaluations of agents and their actions. For example, Sloman, Fernbach, and Ewing (2009) note: "Actions that are connected to bad outcomes through fewer intermediate causes are more blameworthy" (p.11). In our example, administering the drug that can cause blood clots directly would thereby be predicted to be morally worse than administering the drug which can cause the same outcome via several intermediate steps. Similar effects can be expected when the strength of causal links is increased.

Are causal proximity effects rational?

Proximity effects have sometimes been described as biases (e.g., Johnson & Drobny, 1985). But this does not have to be the case. Proximity effects can naturally arise from the way in which moral reasoning about agents, actions, and outcomes is mediated by causal models (Waldmann, Wiegmann, & Nagel, 2017; Sloman et al., 2009).

In a causal model framework (see Waldmann, 2017; Sloman, 2005, for overviews), representing a chain of causally connected events generally means representing a number of events that are connected by probabilistic links. If each event in the chain actually occurs, there is a certain probability that the next event in the chain will occur as well. Say that in our example, there are five intermediate links between administering drug A and the development of blood clots, each of them with a probability of 0.15 conditional on its direct cause. Then $p(\text{blood clots}|\text{drug A}) = 0.15^5 = 0.00008$. For drug B, there is just one probabilistic link with a strength of 0.15, thus $p(\text{blood clots}|\text{drug B}) = 0.15$. In such a case, administering drug B would thus be much more likely to cause harm than administering drug A.

This calculation of course rests on the assumption that all single links, be it the direct relation or a component of the chain, are roughly equally strong[1] and that there are no alternative causes of the events in the chain. Nothing in our introductory example suggests that this needs to be the case. However, research has shown that a chain representation can indeed trigger the impression of a lower probabilistic dependency between a target cause and effect, and that this effect may be produced by people assigning roughly constant strength priors to verbally instructed probabilistic links (Stephan, Tentori, Pighin, & Waldmann, 2021; Bes, Sloman, Lucas, & Raufaste, 2012).

If people perceive a lower conditional probability of harm given action A than given action B, it naturally follows that

[1] Or, at least, that the links in the chain are sufficiently weak to lower $p(\text{outcome}|\text{action})$ relative to the direct relation.

action A is morally preferable. This should hold prospectively (before acting, as in our introductory example), but it may also be true for retrospective moral evaluations. For instance, an agent may be deemed less morally responsible or blameworthy for harm when their action produced the outcome via a chain rather than directly (Sloman et al., 2009). We posit that such proximity effects on *moral* judgments are mediated by the agents' foreseeability of the harmful consequences (see Lagnado & Channon, 2008, Kirfel & Lagnado, 2020, for effects of foreseeability in other contexts). If an action causes harm via a longer chain, the harm is seen as less likely and thus less foreseeable than in a direct relation (assuming roughly equal strength of causal links), justifying a more lenient moral evaluation of action and agent. Thus, whenever the assumption of a lower probability of harm in a chain is justified (e.g., roughly constant link strengths), proximity effects in causal and moral judgments are not a bias. Direct causal relations can also vary in strength, which we view as a different way of manipulating proximity. Again, we predict a harsher moral evaluation of action and agent the stronger the causal link between their action and a harmful outcome is.

Proximity effects in causal and moral reasoning

Surprisingly, cases like our example have rarely been investigated in the context of moral judgment. Research on causal and moral judgments about chains involving human actions has largely focused on comparisons *within* chains. For example, a debate has revolved around the question whether the first or the last element in a causal chain is selected as "the" (main or most important) cause of a final outcome, and how such judgments are affected by features of causes (such as being intentional actions vs. physical events), or by how much they raise the probability of the outcome (Lagnado & Channon, 2008, Spellmann, 1997, Hilton, McClure, & Sutton, 2010).

Our focus, in contrast, are comparisons *between* two causal chains with the same start (an action) and end (a harmful outcome), but a different number of intermediate events (none vs. several). We found just one study that directly investigated such cases. Johnson and Drobny (1985) presented participants with a case in which a truck driver forgets to replace a safety pin in the steering column of his truck. In the "simple chain" condition, the steering fails and results in an accident. Subsequently, gasoline spills and ignites causing a house to burn down. In the "complex chain" condition, the gasoline first pours into a river, floats across it, ignites grass on the other side, then a field, and finally also burns down the house. In the condition with the longer chain, participants considered the truck driver to be less liable for the damage to the house, and they also indicated that he could foresee the outcome to a lesser extent. However, Johnson and Drobny were interested in legal rather than moral judgments, described negligent omissions instead of actions, and provided their participants with extensive jury instructions. Moreover, the two conditions vary in several confounded aspects. The complex condition presents a chain whose elements are both spatially and temporally more extended than the simple condition. Plus, background knowledge or assumptions about the described events may have had an influence. If participants for example assigned a very low probability to burning gasoline floating across a river, the obtained effects may have been produced by the perception of one very weak but necessary link, instead of being generated by the chain representation as such.

In the moral domain, so-called deviant causal chains have been shown to attenuate judgments of blame (Pizarro, Uhlmann, & Bloom, 2003). In these chains, an actor brings about an intended harmful outcome, but in an unexpected and unusual way. Coincidentally, the described deviant chains are often also longer than their "regular" counterparts. Moreover, in deviant chains foreseeability is often altered because the agent achieves the goal in an unforeseeable fashion. An example would be the case of an unpractised gunman who intends to shoot someone. His shot misses the victim, but startles a herd of pigs that trample the victim to death (cf. Davidson, 2001, p.72). Thus, while these studies investigate moral judgments, it is not clear whether their results are due to length, foreseeability, or deviancy of the chains.

Experiment 1: Varying Chain Length

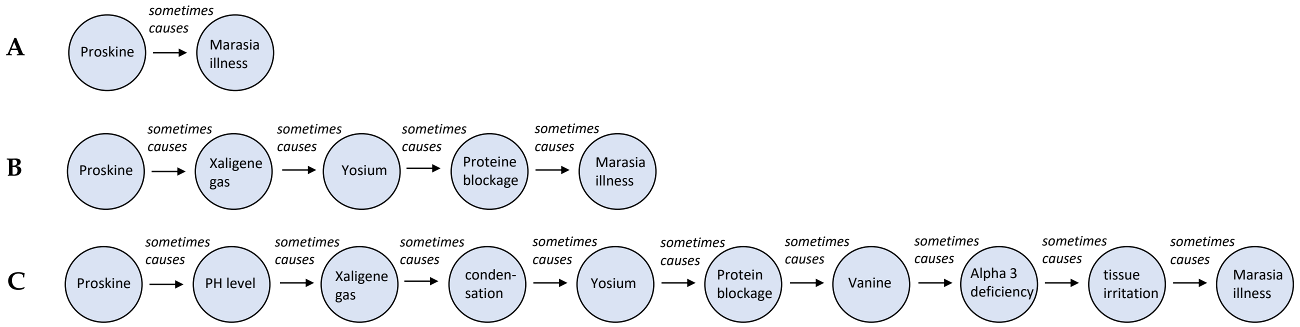
The aim of this experiment was to test for causal proximity effects in chains in a controlled setting, with little to no background knowledge about the events. If people assigned strength priors to links that are roughly constant, we should observe a lower estimated $p(\text{outcome}|\text{action})$ in chains compared to direct relations, along with a more lenient moral evaluation of action and agent in chains. We asked participants to morally evaluate action and agent both prospectively ("is it okay to act?") and retrospectively ("to what extent is the agent morally responsible for the harmful outcome?"). Unlike Johnson and Drobny, we used artificial materials that did not draw on prior knowledge about causal strength, or spatial and temporal relations. The strengths of the links were instructed using verbal labels suggesting equal link strengths (see Figure 1 for an illustration).

Methods

Design, Material and Procedure We created three cover stories about agents causing undesired harm to another person.[2] In each cover story, there was a chain version (three intermediate events and four probabilistic links between action and outcome) and a direct version (no intermediate events and just one probabilistic link). Each participant saw the direct version of one cover story, and the chain version of a different story (in random order). In total, there were six possible Latin square combinations of cover story and structure. This

[2] see https://osf.io/85s23/ for material, data, and code of analyses and figures for all studies reported in this paper. All analyses were conducted and all figures created using R (R Core Team, 2020) and RStudio (RStudio Team, 2020), as well as the following packages: *effsize* (Torchiano, 2020), *ez* (Lawrence, 2016), *ggpubr* (Kassambara, 2020), *MBESS* (Kelley, 2020), *reshape2* (Wickham, 2007), and the *tidyverse* (Wickham et al., 2019).

Figure 1: Example illustrations for direct relations (A), chains in Experiment 1 (B), and chains in Experiment 3 (C).



and all following experiments were implemented online using Unipark Questback.

In the beginning of each story, generic information about the relationship between action and outcome was presented. Here's an example: "A group of scientists is investigating the effects of exposure of to a certain chemical called Proskine. In their lab studies, they found that when Proskine is produced and stored, the following mechanism can unfold: Exposure to Proskine sometimes causes Marasia illness, a new and severe respiratory condition." In the chain condition, the second part of this story read: "Proskine sometimes causes Xaligene gas to develop in its environment. When Xaligene gas develops, it sometimes reacts and causes another chemical, Yosium, to form as well. When Yosium is present, it sometimes causes certain proteins in the body to be blocked upon exposure. When these proteins are blocked, this sometimes causes Marasia illness, a new and severe respiratory condition." On the same page, an illustration of the causal structure was provided, depicting the cited events as nodes, with arrows between them representing the causal links. The arrows were labelled with "sometimes causes" (see Figure 1).

After the generic information, we presented participants with the case of an agent who plans to carry out the action in question (in the example: creating and storing Proskine). The agent was always described as aware of the information from the previous page, but not desiring the negative outcome (in the example, the agent is a chemist who needs to create and store Proskine for research). The stories also mentioned the presence of potential victims of the harmful action. In the chemical scenario, we stated: "The lab is shared with several colleagues". Before giving any information about the occurrence of the harmful outcome, we asked participants to answer the following *prospective* moral question: "From a moral point of view, is it okay for [agent] to [perform the target action]?" Ratings were given on a scale ranging from 1 ("not at all") to 10 ("fully"). On a subsequent page, the actual occurrence of the negative outcome was described (in the example: a colleague in the same lab develops Marasia illness), and participants were asked to indicate the extent to which the agent was morally responsible for this outcome (i.e., a retrospective moral evaluation) on an identical scale. After moral judgments for both cases were recorded, the cases were presented anew and participants were asked to estimate the probability of the harmful outcome given the action. Answers were given on a slider ranging from 0 to 100%.

We predicted the following pattern of results: 1) the action should be seen as more allowed ("okay") in the chain condition compared to the direct condition, 2) the agent should be held less morally responsible for the outcome in the chain compared to the direct condition, and 3) $p(\text{outcome}|\text{action})$ should be estimated as lower in the chain compared to the direct condition.

Participants To achieve 90% power for observing all three effects at a minimum effect size of $d = .20$ each, we planned for a power of 97% for each of three one-sided t-tests ($0.97^3 \approx 0.91$ power to detect all three effects). This resulted in a required sample size of 300 participants. We recruited 304 participants on the platform *prolific.co*. Inclusion criteria (identical for all further experiments) were being a native English speaker, not having participated in previous studies using similar material, an acceptance rate of at least 90% of previous tasks on the platform, and not completing the survey via smartphone. Participants received a compensation of £0.45 for an estimated four minutes of their time. Five participants were excluded due to failing a simple attention check[3], leaving data of 299 participants for the analyses ($M_{age} = 34$, $SD_{age} = 12.1$, 63% women, 37% men, <1% no answer).

Results and Discussion

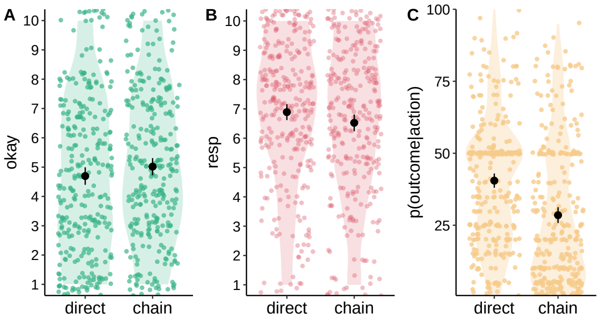
Figure 2 shows the results for all three measures. For the two moral questions, we observed significant, albeit small effects in the predicted directions (*okay*: $M_{direct} = 4.70$, $SD_{direct} = 2.62$, $M_{chain} = 5.02$, $SD_{chain} = 2.55$, $t_{298} = 1.96$, $p = .025$, $d = 0.12$ [0; 0.25]; *moral responsibility*: $M_{direct} = 6.89$, $SD_{direct} =$

[3] "If Peter is taller than Alex, and Alex is taller than Max, who is the shortest among them?" This attention check was used in all studies reported here.

2.39, M_{chain} = 6.53, SD_{chain} = 2.51, t_{298} = 2.41, p = .008, d = 0.15 [0.03; 0.27]). For the causal measure on the other hand, we observed a medium-sized effect in the predicted direction, M_{direct} = 40.51, SD_{direct} = 21.7, M_{chain} = 28.51, SD_{chain} = 24.19, t_{298} = 10.61, $p < .001$, d = 0.52 [0.42; 0.62]. No corrections of p-values were applied (although, given our conjunctive hypothesis, we could have increased the alpha-level per test).

Thus, while both kinds of moral judgment were indeed more lenient when a longer chain was described between action and outcome, the effects were relatively small. However, participants clearly perceived a difference in the strengths of the causal relations connecting action and outcome between the direct and the chain conditions. In chains, outcomes were estimated as less likely to occur than in direct relations. A possible explanation is that while $p(\text{outcome}|\text{action})$ matters for moral judgments, a relatively large difference is required. We test this hypothesis in the next experiment by directly manipulating causal strength.

Figure 2: Mean ratings for whether it is okay to act (A), agents' moral responsibility (B), and $p(\text{outcome}|\text{action})$ (C) per structure condition in Experiment 1. Error bars are 95% CIs.



Experiment 2: Varying Strength

In Experiment 1, we compared a direct probabilistic causal relation with a chain that contained several probabilistic links of equal strength, which entails lower overall strength in the chain than in the direct condition given equal link strengths. Thus, both the causal strength of the relation between action and outcome and chain length was varied. In Experiment 2, we focused only on direct causal relations while manipulating their strength. Moreover, strength is conveyed more saliently here by presenting numeric values.

Methods

Design, Material and Procedure We varied $p(\text{outcome}|\text{action})$ in three levels: .30, .60, and .90. The manipulation was delivered within subject. The cover stories were otherwise identical to the ones in Experiment 1. Each participant saw each link strength in the context of a different cover story (in random order), in one of three possible Latin square combinations.

Instead of learning about direct relations versus chains, participants in this experiment were presented with generic information about $p(\text{outcome}|\text{action})$ for each case. To convey strength information, we presented relative frequencies. For the "chemical" example, the instruction read: "A group of scientists is investigating Marasia illness, a new and severe respiratory condition. They suspected that it may be related to exposure with Proskine, a newly developed chemical that is sometimes used in pharmaceutical labs. The scientists therefore reviewed the health records of 1000 employees of pharmaceutical companies who have been in contact with Proskine. For comparison, they also reviewed the records of 1.000 employees who do the same job, but have not been in contact with this specific chemical. These are their results: Of the 1000 people who have been in contact with Proskine, [300/600/900] contracted Marasia illness. Of the 1000 people who have not been in contact with Proskine, no one contracted Marasia illness." On the subsequent pages, the task proceeded exactly as in Experiment 1, with identical measures.

We predicted the following pattern of results: 1) The agent's action should be assessed as more allowed ("okay") the less likely its negative effect (.30 > .60 > .90), and 2) the agent should be held less morally responsible for the negative outcome the less likely it was to result from their action (.30 < .60 < .90). We also expected participants to accurately infer $p(\text{outcome}|\text{action})$ from the presented numbers, which can be seen as a manipulation check in this case.

Participants We aimed for a sample size of 292 valid participants in this experiment. In three one-way, repeated-measures ANOVAs, this will yield a power of 91% to detect a small effect (η_p^2 = .01) on all measures. We invited 300 participants to take part in the experiment. Participants received a compensation of £0.60 for an estimated six minutes of their time. Nine participants were excluded due to failing a simple attention check, leaving data of 291 participants for the analyses (M_{age} = 31.59, SD_{age} = 11.47, 59% women, 39% men, 2% another identity or no answer).

Results and Discussion

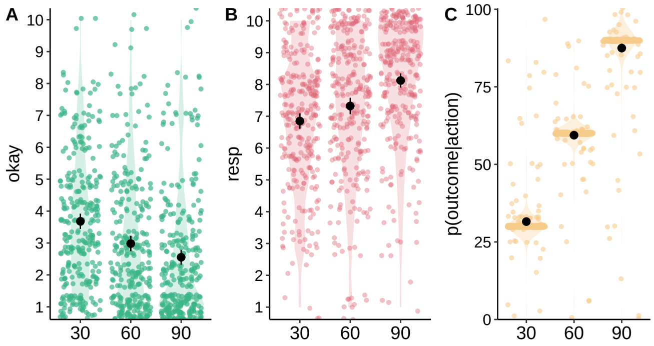
See Figure 3 for results. Prospectively, actions were regarded as more permissible ("okay") the lower the probability of harm was ($M_{30\%}$ = 3.68, $SD_{30\%}$ = 2.09, $M_{60\%}$ = 2.98, $SD_{60\%}$ = 2.07, $M_{90\%}$ = 2.55, $SD_{90\%}$ = 2.02, $F_{2,580}$ = 29.39, $p <$ .001, η_p^2 = .09 [0.06; 0.13][4], which is confirmed by a negative linear trend in group means (t_{870} = -6.60, $p < .001$), and no detectable quadratic trend (t_{870} = 0.91, p = .365). Retrospectively, agents were held less morally responsible for harm the weaker the probabilistic relation between their action and the outcome was ($M_{30\%}$ = 6.85, $SD_{30\%}$ = 2.11, $M_{60\%}$ = 7.32, $SD_{60\%}$ = 2.24, $M_{90\%}$ = 8.12, $SD_{90\%}$ = 1.95, $F_{2,580}$ = 43.34, p

[4] We report 90% confidence intervals for all η_p^2, see Steiger (2004))

$< .001$, $\eta_p^2 = 0.13$ [0.09; 0.17]). There was a positive linear trend in group means $t_{870} = 7.32$, $p < .001$), and no detectable quadratic trend ($t_{870} = 1.08$, $p = .28$). Finally, responses to the query about p(outcome|action) confirmed that the strengths of probabilistic relations between action and outcome were accurately inferred ($M_{30\%} = 31.56$, $SD_{30\%} = 9.31$, $M_{60\%} = 59.41$, $SD_{60\%} = 7.68$, $M_{90\%} = 87.49$, $SD_{90\%} = 11.96$,, $F_{2,580} = 2919.31$, $p < .001$, $\eta_p^2 = 0.91$ [0.90; 0.92]), with a positive linear trend in group means ($t_{870} = 68.76$, $p < .001$), and no detectable quadratic trend ($t_{867} = 0.17$, $p = 0.87$). No corrections of p-values were applied (although, given our conjunctive hypothesis, we could have increased the alpha-level per test for the moral questions).

In sum, we demonstrated that the probabilistic strength of the relationship between action and outcome clearly influenced moral evaluations. However, a very large effect on the causal measure, p(outcome|action) only led to medium-sized effects on the moral measures. Thus, the chain manipulation in Experiment 1 may not have decreased the perceived p(outcome|action) enough to produce a large effect on moral judgments. In Experiment 3 we went back to comparing direct causal relations with chains but increased the length of the chain hoping for a stronger effect. Moreover, we tested the hypothesis that foreseeability mediates the effect.

Figure 3: Mean ratings for whether it is okay to act (A), agents' moral responsibility (B), and p(outcome|action) (C) per strength condition in Experiment 2. Error bars are 95% CIs.



Experiment 3: The Role of Foreseeability

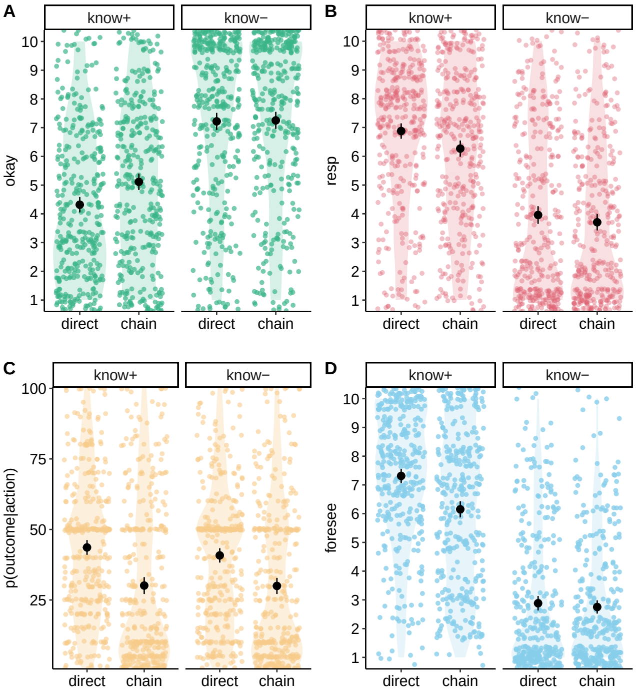
In this experiment, we used the same task as presented in Experiment 1, but with a stronger chain manipulation (nine probabilistic links instead of four, see Figure 1). In addition, we added new conditions in which agents were unaware of the possible harm that may result from their action. A longer causal chain does not only entail lower causal strength, but should normally also decrease the foreseeability of the negative outcome compared to a direct relation. However, this difference in outcome foreseeability between chains and direct relations depends on agents' awareness of the relation. When someone is unaware of any relation between their action and a harmful outcome, it seems hardly morally relevant whether action and outcome are related directly or by a longer chain. If proximity effects on moral judgments are mediated by outcome foreseeability, they should be eliminated without proximity knowledge.

Methods

Design, Materials, and Procedure We implemented a 2 (structure: direct vs. chain, within-subjects) x 2 (proximity knowledge: yes vs. no, between-subjects) mixed design. In the chain conditions, nine probabilistic links were instructed instead of four (see OSF for the full material). The direct conditions were identical to the ones in Experiment 1. In the"proximity knowledge" conditions (*know+*), agents were aware of the relation between their action and the possible harmful outcome (as in Experiments 1 and 2). In the "no proximity knowledge" conditions (*know-*), they were not. In the "chemical" cover story, for instance, we stated in *know-*: "Since the scientists studying Proskine have not published their results so far, Mary cannot be aware of them. To the best of her knowledge, there are no special risks associated with producing and storing Proskine." The cover stories were combined with the levels of the structure manipulation in a Latin square as described in Experiment 1. Procedure and measures also were the same as in Experiment 1, with the addition of a question about foreseeability at the end of the experiment, presented along with the question about p(outcome|action). The new question read: "To what extent could [agent] foresee that someone would be harmed by [her/his] action?", and responses were provided on a scale ranging from 1 ("not at all") to 10 ("fully"). This question was intended primarily as a manipulation check.

Participants We aimed for a sample size of 700 valid participants in this experiment. The sample size was determined by a simulation (see OSF for code), focusing on the two moral questions (*okay* and *resp*). Based on pilot studies, we planned for proximity effects of $d = 0.22$ for *okay* and of $d = 0.28$ for *resp*, in one-sided, paired t-tests in *know+*. We predicted null effects in two-sided paired t-tests for both measures in *know-*. If these patterns obtained, we should thus also observe a significant structure x proximity knowledge interaction in mixed ANOVAs for both *okay* and *resp*. With 700 participants, we achieve a power of >90% to detect the full set of the specified effects. We invited 720 participants to take part in the experiment. Participants received a compensation of £0.65 for an estimated six minutes of their time. Sixteen participants were excluded due to failing a simple attention check or due to completing the survey from a smartphone against instructions, leaving data of 704 participants for the analyses ($M_{age} = 34.81$, $SD_{age} = 13.14$, 50% women, 49% men, 1% non-binary or no answer).

Figure 4: Mean ratings for whether it is okay to act (A), agents' moral responsibility (B), $p(\text{outcome}|\text{action})$ (C), and agents' outcome foreseeability (D) per structure and proximity knowledge condition in Experiment 3. Error bars are 95% CIs.



Results and Discussion

Figure 4 shows the results. In the *know+* conditions, we found the predicted proximity effects on the moral questions, with larger effect sizes than in Experiment 1 (*okay*: $M_{direct} = 4.32$, $SD = 2.56$, $M_{chain} = 5.11$, $SD = 2.75$, $t_{351} = 5.08$, $p < .001$, $d = 0.30$ [0.18; 0.42], *resp*: $M_{direct} = 6.88$, $SD = 2.55$, $M_{chain} = 6.27$, $SD = 2.69$, $t_{351} = 3.8$, $p < .001$, $d = 0.23$ [0.11; 0.35]). As predicted, the effects largely disappeared in the *know-* conditions, although a very small significant effect remained for attributions of moral responsibility (*okay*: $M_{direct} = 7.22$, $SD = 2.83$, $M_{chain} = 7.25$, $SD = 2.83$, $t_{351} = 0.26$, $p = 0.793$, *resp*: $M_{direct} = 3.96$, $SD = 2.91$, $M_{chain} = 3.70$, $SD = 2.71$, $t_{351} = 2.03$, $p = 0.043$, $d = 0.09$ [0; 0.18]). The predicted interaction between structure and foreseeability was thus found for the *okay* question ($F_{1,702} = 14.01$, $p < .001$, $\eta_p^2 = .02$ [0.01; 0.04]), but not for moral responsibility ($F_{1,702} = 3.11$, $p = 0.078$). Independent of proximity knowledge, participants thought that the final outcomes were less likely to occur in the long chains than in the direct causal relation ($t_{703} = 14.22$, $p < .001$, $d = 0.46$ [0.39; 0.53], no interaction). However, the effect size was similar to the one we found in Experiment 1 ($d = 0.52$). As expected, participants only ascribed less outcome foreseeability with increased chain length to agents who were aware of the relation between action and outcome but not to agents without knowledge about the causal relation (interaction: $F_{1,702} = 36.89$, $p < .001$, $\eta_p^2 = .05$ [0.03; 0.08], see OSF for the full analysis and all descriptive statistics). No corrections of p-values were applied (although, given our conjunctive hypothesis, we could have increased the alpha-level per test for the moral judgment questions).

General Discussion

We set out to test the hypothesis that (1) instructing a chain of probabilistically linked events between an action and a harmful outcome would lead to a more lenient moral evaluation of the agent and the action, compared to instructing a direct relation. We expected this pattern because (2) participants should perceive the harmful outcome as less likely to actually occur in a chain than in a direct relation. Moreover, we predicted that (3) the effect will be mediated by participants' attributions of outcome foreseeability to agents.

We found evidence for (1) in two experiments, but with surprisingly small effect sizes. The actions were only seen as slightly more permissible, and agents only judged as slightly less responsible in the chain compared to the direct conditions. Our data in all experiments are *consistent* with (2). The probability of the final outcome given action was indeed perceived as lower in chains than in direct relations. In all experiments, a lower $p(\text{outcome}|\text{action})$ was also associated with a more lenient moral evaluation. However, it is unclear whether these effects were caused by the difference in the structure of the causal models (direct vs. chain), or by the lowered strenghts of the relation between action and final outcome. It is possible to experimentally dissociate these two factors. For a more rigorous test, we would need to keep chain length constant and vary $p(\text{outcome}|\text{action})$ independently (see Stephan et al., 2021). We have conducted such a study in the meantime and found that the effect of chain length on moral judgments is at least substantially mediated by $p(\text{outcome}|\text{action})$ (Engelmann & Waldmann, *manuscript in preparation*). Further experiments are ongoing. Finally, Experiment 3 provides support for (3), the mediating role of outcome foreseeability. Chain length largely ceased to affect moral judgments when agents were unaware of the presence of the chain or of the direct relation. A very small effect persisted for moral responsibility, reminiscent of the moral luck literature (Young, Nichols, & Saxe, 2010). We will explore this puzzling effect further in future research.

An unexpected and interesting observation in all experiments was that medium (Experiment 1, Experiment 3) or large (Experiment 2) differences in $p(\text{outcome}|\text{action})$ only translated into small (Experiment 1, Experiment 3) or medium (Experiment 2) effects on the moral judgment measures. The only manipulation that pushed moral judgments across the scale midpoints (from permissible to impermissible and from responsible to not responsible) was the knowledge manipulation in Experiment 3. In the causal reasoning literature, it is sometimes claimed that people care more about causal structure than about causal strength (e.g., Bes et al., 2012). Possibly, a similar effect obtains in moral reasoning, where causal reasoning is combined with inferences about others' mental states: Once agents know about the mere existence of a causal link between an action and a harmful outcome (as is the case in all our scenarios except the *know*-conditions of Experiment 3), we may be reluctant to judge their actions as permissible or blameless, even when harm becomes increasingly unlikely. While causal strength is clearly not irrelevant, a negative impression based on the causal link between an action and harm may prevail. A current example is the reluctance of some people to get vaccinated against Covid-19 because of extremely unlikely side-effects of some of the available vaccines, despite those risks being dramatically outweighed by the benefits.

Given that our chain manipulation here did not dissociate causal strength from causal structure (as explained above), it follows that the knowledge manipulation in Experiment 3 also did not dissociate *knowledge about causal strength* from *knowledge about causal structure*. It is clearly possible for agents to be aware of the presence of a causal link without knowing its strength. Likewise, agents might know about a statistical association between events without knowing if and how they are causally connected. We are presently conducting further experiments that aim to illuminate how these components combine to inform moral judgments.

References

Bes, B., Sloman, S., Lucas, C. G., & Raufaste, (2012). Nonbayesian inference: Causal structure trumps correlation. *Cognitive Science*, *36*(7), 1178–1203.

Davidson, D. (2001). *Essays on actions and events: Philosophical essays* (Vol. 1). Oxford University Press.

Engelmann, N., & Waldmann, M. R. (*manuscript in preparation*). Causal structure, causal strength, and moral judgment.

Hilton, D. J., McClure, J., & Sutton, R. M. (2010). Selecting explanations from causal chains: Do statistical principles explain preferences for voluntary causes? *European Journal of Social Psychology*, *40*(3), 383–400.

Johnson, J. T., & Drobny, J. (1985). Proximity biases in the attribution of civil liability. *Journal of Personality and Social Psychology*, *48*(2), 283–296.

Kassambara, A. (2020). ggpubr: 'ggplot2' based publication ready plots [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=ggpubr (R package version 0.4.0)

Kelley, K. (2020). Mbess: The mbess r package [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=MBESS (R package version 4.8.0)

Kirfel, L., & Lagnado, D. A. (2021). Causal judgments about atypical actions are influenced by agents' epistemic states. *Cognition*, *212*, 104721.

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*(3), 754–770.

Lawrence, M. A. (2016). ez: Easy analysis and visualization of factorial experiments [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=ez (R package version 4.4-0)

Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, *39*(6), 653–660.

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

RStudio Team. (2020). Rstudio: Integrated development environment for r [Computer software manual]. Boston, MA.

Sloman, S. (2005). *Causal models: How people think about the world and its alternatives*. Oxford University Press.

Sloman, S., Fernbach, P. M., & Ewing, S. (2009). Causal models: The representational infrastructure for moral judgment. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 50, pp. 1–26). Academic Press.

Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General*, *126*(4), 323.

Steiger, J. H. (2004). Beyond the f test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological methods*, *9*(2), 164.

Stephan, S., Tentori, K., Pighin, S., & Waldmann, M. R. (2021). Interpolating causal mechanisms: The paradox of knowing more. *Journal of Experimental Psychology: General*.

Torchiano, M. (2020). effsize: Efficient effect size computation [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=effsize (R package version 0.8.1)

Waldmann, M. R. (2017). *The Oxford handbook of causal reasoning*. Oxford University Press.

Waldmann, M. R., Wiegmann, A., & Nagel, J. (2017). Causal models mediate moral inferences. In J. Bonnefon & B. Tremolière (Eds.), *Moral inferences* (pp. 37–55). London: Routledge/Taylor Francis Group.

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12), 1–20.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

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.