Cognition 156 (2016) 164–176

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

How prescriptive norms influence causal inferences

Jana Samland *, Michael R. Waldmann

Department of Psychology, University of Göttingen, Germany

ARTICLE INFO

Article history:
Received 10 June 2015
Revised 13 July 2016
Accepted 14 July 2016
Available online 31 August 2016

Keywords:
Causal reasoning
Moral judgment
Causal selection
Conversational pragmatics
Norms

ABSTRACT

Recent experimental findings suggest that prescriptive norms influence causal inferences. The cognitive mechanism underlying this finding is still under debate. We compare three competing theories: The culpable control model of blame argues that reasoners tend to exaggerate the causal influence of norm-violating agents, which should lead to relatively higher causal strength estimates for these agents. By contrast, the counterfactual reasoning account of causal selection assumes that norms do not alter the representation of the causal model, but rather later causal selection stages. According to this view, reasoners tend to preferentially consider counterfactual states of abnormal rather than normal factors, which leads to the choice of the abnormal factor in a causal selection task. A third view, the accountability hypothesis, claims that the effects of prescriptive norms are generated by the ambiguity of the causal test question. Asking whether an agent is a cause can be understood as a request to assess her causal contribution but also her moral accountability. According to this theory norm effects on causal selection are mediated by accountability judgments that are not only sensitive to the abnormality of behavior but also to mitigating factors, such as intentionality and knowledge of norms. Five experiments are presented that favor the accountability account over the two alternative theories.

1. Introduction

Most theories of moral judgments assume that moral evaluations presuppose causal facts: an agent is only held (morally) accountable for an outcome if she has actually causally contributed to its occurrence (see Shaver, 1985; Sloman, Fernbach, & Ewing, 2009; Weiner, 1995). However, the traditional claim that moral judgments are secondary to causal ones has been challenged by recent findings suggesting that the inverse relation also holds: causal judgments are also influenced by moral evaluations (Alicke, 1992; Alicke, Rose, & Bloom, 2011; Hitchcock & Knobe, 2009; Kominsky, Phillips, Gerstenberg, Lagnado, & Knobe, 2015). One example for this influence is the pen vignette (Knobe & Fraser, 2008) which describes a scenario in which faculty members and administrative assistants working in a philosophy department frequently take pens although only administrative assistants are allowed to do so. One day a faculty member and an administrative assistant both take a pen simultaneously, which leads to a problem. There are no pens left. Participants of experiments who were asked who *caused* the problem tend to choose the faculty member who violated the prescriptive norm over the administrative assistant who is allowed to take pens. Thus, normative evaluations seem to influence causal judgments. However, the cognitive processes underlying this so-called *norm effect* are still under dispute.

1.1. Possible influences of prescriptive norms on causal inferences

Although it is a well-established finding that norms affect causal judgments, it is less clear how these judgments are affected by norms. The literature suggests different possibilities: One possibility is that norms alter *causal model representations*, that is, they lead to changes of the causal structure or the estimated causal strengths (see Waldmann & Hagmayer, 2013, for an overview of causal model theories). An influence on causal strength, for example, is suggested by Liu and Ditto (2013): "[t]he more participants believed that the action was immoral even if it had beneficial consequences, the less they believed it would actually produce those consequences (...)" (p. 318). Consistent with the claim that the consideration of norms alters causal representations, the *culpable control model* of blame, proposed by Alicke (2000), states that participants tend to exaggerate the causal role of the norm-violating agent because they have a desire to blame her for the negative outcome. Thus, the first possibility is that prescriptive norms influence causal inferences by changing the size of the causal model's strength parameters.

* Corresponding author at: Department of Psychology, University of Göttingen, Gosslerstr. 14, 37073 Göttingen, Germany.
E-mail address: jana.samland@psych.uni-goettingen.de (J. Samland).

http://dx.doi.org/10.1016/j.cognition.2016.07.007
0010-0277/

A second possibility how norms could influence causality is that normative evaluations influence *causal selection* judgments without affecting intuitions about the underlying causal model. Causal selection refers to the fact that in situations in which several factors contribute to an outcome, people often select one over the other factors and name it 'the cause' (see Cheng & Novick, 1991). For example, although subjects may know that a forest fire depends on both a lightning bolt and oxygen, they typically select the first factor as the cause.

Hitchcock and Knobe (2009; and similarly Halpern & Hitchcock, 2014) have proposed a theory, the *counterfactual reasoning account of causal selection*, that traces causal selection back to counterfactual reasoning about abnormal factors. Abnormality in this account may refer to all types of norm violations including statistical, moral, or proper functioning norms. According to Hitchcock and Knobe's theory, abnormal factors stimulate reasoning about a possible world in which the abnormal factor had instead been normal, whereas thinking about an alternative behavior of a normal factor is less likely (see also Hesslow, 1988; Hilton & Slugoski, 1986; Kahneman & Miller, 1986). The greater salience (or *relevance*; see Phillips, Luguri, & Knobe, 2015) of the counterfactual contrast of the abnormal factor leads to its choice as the cause. On this account counterfactual reasoning can be regarded as the mediator between the violation of a norm and causal selection. Abnormality is only one means leading to an increase in salience of a counterfactual alternative; there are many other ways (see also Kominsky et al., 2015; Phillips et al., 2015).

1.2. The ambiguity of causal queries: The accountability hypothesis

The theories we have discussed so far claim that prescriptive norms either influence parameters of causal models or guide causal selection through counterfactual reasoning about abnormal factors. However, there is an alternative to the view that prescriptive norms affect causal judgments. One general problem of studies investigating norm effects is the notorious ambiguity of the term "cause." Especially in the context of human actions, it can both refer to the question of whether a mechanism underlying a causal relation is present and to the question of whether an agent can be held *accountable* for an outcome. As Deigh (2008) points out, Hart and Honoré (1959) have already argued "(...) that the statement that someone has caused harm either means no more than that the harm would not have happened without ('but for') his action or (...) it is a disguised way of asserting the 'normative judgment' that he is responsible in the first sense, i.e., that it is proper or just to blame or punish him or make him pay" (pp. 61) (see also Alicke, Mandel, Hilton, Gerstenberg, & Lagnado, 2015; Lagnado & Channon, 2008; Suganami, 2011; Sytsma, Livengood, & Rose, 2012, for related views). The ambiguity of queries about the cause in scenarios demonstrating norm effects is grounded in the presupposition relation between accountability and causation. Agents are only held accountable for outcomes they have caused.[1] Thus, causal test questions may either narrowly refer to the causal process linking the agent's behavior to the morally charged outcome, or they could refer to the more comprehensive set of factors determining accountability.

Based on the idea of conversational or experimental pragmatics (see e.g., Noveck & Reboul, 2008; Wiegmann, Samland, & Waldmann, 2016), the *accountability hypothesis* assumes that subjects form hypotheses about the intended meaning of the causal test question. Due to the ambiguity of causal queries, they either interpret this question as a request to assess accountability or as a request to assess causality (in the narrow sense). Which of the two meanings is accessed depends on pragmatic contextual factors in the test situation; subjects will choose the one that makes more sense in the present context. This relation between causal test questions, causality, and accountability is presented in Fig. 1.

Causality in the narrow sense refers to contingent dependency relations between causes and effects that are generated by causal mechanisms (Fig. 1, left). A popular account of how causal dependencies are represented is the *counterfactual view* that claims that an event qualifies as a cause if the effect had not happened in the counterfactual absence of the cause (Lewis, 1973). This view has been extended to account for more complex causal networks (see, for example, Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2014; Gerstenberg & Tenenbaum, in press; Lagnado & Gerstenberg, in press; Spellman & Kincannon, 2001). In the pen vignette both agents are equally causal; had either the professor or the secretary not taken a pen, the problem would not have occurred. Thus, both agents equally contributed to the outcome. Note that the counterfactual theory of causal selection (Hitchcock & Knobe, 2009) primarily addresses a separate counterfactual reasoning process in a later phase after the initial phase of establishing a causal model. Thus, a critique of the assumption that counterfactual reasoning underlies causal selection is compatible with the view that causal model representations are based on counterfactual intuitions.

Queries targeting accountability are more general than queries referring to causal relations in the narrow sense (see Fig. 1, right). Accountability assessments include the identification of causal relations between acts and outcomes (hence the possibility to use the term "cause") (Fig. 1, right, bottom layer) but there are numerous additional factors that determine accountability judgments (Fig. 1, right, top layer): Accountability in social contexts requires that causal effects of the actions are positively or negatively valued. Moreover, accountability judgments are sensitive to mental state factors, such as the agent's intentionality, the foreseeability of the outcome, or the agent's knowledge about the existence and applicability of a prescriptive norm (see, e.g., Cushman, 2008; Gailey & Falk, 2008; Lagnado & Channon, 2008; Malle, Guglielmo, & Monroe, 2014; Young & Saxe, 2011). Thus, whether an agent is held accountable for an outcome is not only dependent on her causal contribution but also on these additional factors. For example, an agent who caused a negative outcome unintentionally, did not anticipate the outcome, or was unaware that the act was forbidden will be held less accountable than an agent who caused the outcome intentionally and with full knowledge. The abnormality of the behavior alone does not suffice for assessing accountability; the additional boundary conditions have to be checked as well.

Initial evidence for the relevance of such additional features in causal queries comes from a recent developmental study investigating a child-friendly variant of the pen vignette in children and adults (Samland, Josephs, Waldmann, & Rakoczy, 2016). Adult subjects in this study were more likely to select a norm-violating agent, a hedgehog, as the cause if it knew about the norm than when it was ignorant.

In sum, the key difference between the accountability hypothesis and its competitors is that the accountability hypothesis does not assume that prescriptive norms change causal representations or the way causal representations are accessed but that pragmatic contextual features steer subjects toward an accountability understanding of the causal test question. In the pen vignette, for instance, it seems far more plausible that the causal query addresses accountability than causal mechanisms because the causal relations are trivial. That the act of taking a pen removes a pen is obvious so that it is unlikely that subjects will think that they are supposed to solely judge this causal relation. The aspect of norm

[1] In some cases, causal responsibility may be indirect, such as in situations in which parents are held responsible for the actions of their children. In such situations, the underlying assumption seems to be that parents are in control of their children's behavior. We test an example of such an indirect accountability relation in Experiment 4.

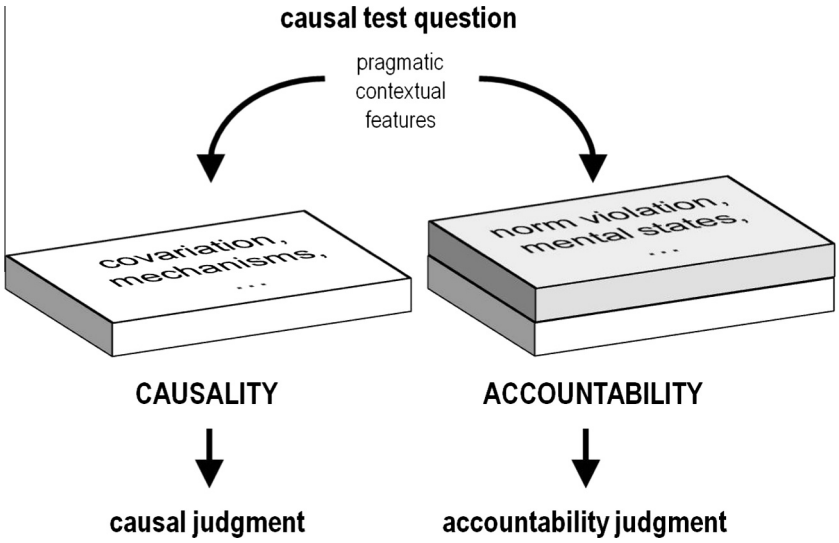

Fig. 1. The relation between causal test questions, causality, and accountability (see text for explanations).

violation is highlighted in the cover story and the causal test question asks about the *agents* (e.g., Professor Smith) who are the primary target of accountability assessments. All these factors converge on making an accountability interpretation far more plausible than a narrow causal interpretation.

1.3. The present studies

The culpable control model of blame (Alicke, 2000), the counterfactual reasoning account of causal selection (Hitchcock & Knobe, 2009), and our accountability hypothesis postulate different mechanisms for the norm effect in causal selection. Experiment 1 tests the assumption of the culpable control model that the wish to blame a norm-violating agent leads to an exaggeration of causal strength. We test this hypothesis by using an unambiguous strength measure. Experiment 2 focuses on the counterfactual account of causal selection, which views counterfactual reasoning as the mediator between abnormality and causal selection. To test this account we manipulate the salience of counterfactual alternatives by other means than norm violations. If the salience or relevance of counterfactual alternatives triggered causal selection, this manipulation should also be effective. Finally, Experiments 3 and 4 focus on the accountability hypothesis, which assumes that causal test questions are ambiguous. In Experiment 3 we test whether it is possible to influence the understanding of the test question by creating pragmatic contexts that either suggest a narrow causality or an accountability interpretation of the test question. Experiment 4 is motivated by what we consider the major shortcoming of the counterfactual reasoning account of causal selection, its sole focus on the abnormality of behavior. The experiment tests the prediction that causal selection in social scenarios is mediated by accountability judgments, which depend on a number of factors besides norm-violating behavior, such as intentionality and knowledge of norms.

2. Experiment 1

One possibility how norms may affect causal representations is by altering the size of causal parameters. For example, Alicke (2000) has claimed that one way to justify blaming the norm-violating agent is to exaggerate her causal contribution to the effect. The goal of Experiment 1 is to test whether norms indeed influence the size of the assumed causal parameters.

Most causal theories focusing on causal structure and strength belong to the heterogeneous class of dependency theories that view causes as *difference makers*: a factor C is a cause of its effect E if E depends upon C (see Paul & Hall, 2013; Waldmann & Hagmayer, 2013; Waldmann & Mayrhofer, 2016, for overviews). These theories generally contrast the case in which both cause and effect are present with the counterfactual case in which the cause is absent. To measure subjects' intuitions about this contrast, we chose probability estimates in the presence and absence of the target causes. Since in the pen vignette and related stories the effects are generated by the joint presence of two causes, the proper contrast for each target cause are cases in which the co-factor is present (see Halpern & Hitchcock, 2014). Thus, for example, the causal impact of the action of the professor can be seen by contrasting the estimated probability of a lack of pens in the presence versus the absence of the professor's act given that the second cause, the action of the administrative assistant, is simultaneously occurring.

The present experiment contrasts responses to this contrast measure with the responses to the standard question how strongly the action caused the outcome. Previous research has shown that this question is ambiguous so that we expect to see the usual norm effect. Unlike the cause question, the more indirect contrast measure avoids the ambiguity and allows for an unconfounded test of the hypothesis that norms influence causal parameters. We are going to test whether abnormality affects causal strength estimations using two different cover stories from the literature.

2.1. Method

2.1.1. Participants

162 subjects (mean age = 30.01; SD = 9.31), recruited via a crowdsourcing platform with participants from many countries, took part in the online study. All subjects earned 50 British pence for their participation. 28 participants were excluded due to their wrong answers to a manipulation check question which checked whether participants could correctly remember each agent's normative status. We thus made use of the data of 134 participants.[2]

[2] In our experience, dropout and exclusion rates depend on the online site. Typically, these numbers are lower in experiments run in the highly selected M-Turk community, which is not accessible to researchers outside the U.S.

2.1.2. Design and procedure

The design of the experiment is based on a 2 (scenario: pen vs. computer crash) × 2 (question-type: cause vs. contrast) × 2 (normality: normal vs. abnormal) structure with the last factor being manipulated within subject. Each participant was randomly assigned to one of two scenarios and one question type (see Appendix A for the complete materials).

We tested two scenarios, the *pen vignette* described in the introduction (Knobe & Fraser, 2008), and a *computer crash scenario* in which two agents simultaneously log on a computer although only one of them has the permission to do so. As a consequence, the computer crashes (Knobe, 2005). Both scenarios have a conjunctive causal structure with multiple necessary causes with a norm-violating (abnormal) and a norm-conforming (normal) cause equally contributing to the effect.

After having read the instruction, subjects in the cause question conditions were presented with two test questions similar to the ones used in previous studies: "How strongly did agent A/agent B cause X (the outcome)?" Responses were given on an 11-point Likert Scale ranging from "not at all" (0) to "completely" (100). Subjects in the contrast measure conditions were asked to estimate the probabilities of the effect in the presence of both causes, in the absence of each of the two causes (while the other was still present), and in the absence of both causes. In the pen scenario, for instance, we asked subjects (i) "How likely is it that there are no pens left given that there had been only two pens on the desk and both Professor Smith and the administrative assistant took one pen each?". Subsequently we asked for each agent in randomized order (ii [iii]) "How likely would it have been that there are no pens left given that there had been only two pens on the desk but only Professor Smith [the administrative assistant] had taken one pen (and the administrative assistant [Professor Smith] had not taken any pen)?" Finally, we asked them to (iv) estimate the likelihood of no pens being left "(...) given that there had been only two pens on the desk but neither Professor Smith nor the administrative assistant had taken one pen?" To express their judgments, we gave subjects an 11-point Likert Scale ranging from "impossible" (0) to "certain" (100).

As a measure of causal strength, we took the difference between the two estimates for the presence of both causes (i) and the absence of each single cause (ii and iii). The (conditional) contrast measure for the abnormal cause, for instance, is the difference between the hypothetical situation in which the abnormal cause is present and the one in which it is absent given that the normal cause is constantly present in these two situations (i–iii). At the end of the study, participants were given two control questions to identify participants who did not remember correctly who the norm-violating and norm-conforming agents were.

Table 1
Descriptive statistics for the estimates of the probability in the presence or absence of the normal or abnormal cause (i–iv).

Scenario	Computer crash	Pen	Overall
(i) Both causes present			
Mean (SD)	82.86 (20.88)	85.94 (22.12)	84.50 (21.43)
(ii) Abnormal cause only			
Mean (SD)	30.71 (27.34)	27.19 (29.86)	28.83 (28.53)
(iii) Normal cause only			
Mean (SD)	27.14 (24.47)	30.31 (31.67)	28.83 (28.35)
(iv) Both causes absent			
Mean (SD)	18.57 (27.58)	21.88 (25.46)	20.33 (26.29)

2.2. Results and discussion

The means and standard deviations of the likelihood estimations can be seen in Table 1. In general, the probability of the effect in the presence of the two causes, that is, the joint action of both the norm-violating and the norm-conforming agent, was estimated highest, near the ceiling. We attribute the small deviations from the ceiling either to an unwillingness to use the extremes of the scale or to background assumptions that may question the sufficiency of these two causes in similar situations. Consistent with the conjunctive causal structure with multiple necessary causes the estimates were near the other end of the scale when one or both causes were absent.

The results of the conditional contrast measure and the cause question can be seen in Fig. 2. We replicated the norm effect with the cause question: the norm-conforming (i.e., normal) agent was viewed as considerably less causal than the norm-violating (i.e., abnormal) agent, $F(1,130) = 548{,}43$, $p < 0.001$, $d = 3.341$. By contrast, no effect of normality is seen with the unambiguous conditional contrast measure, $F(1,130) = 0.005$, $p = 0.944$. Both the normal and the abnormal cause were viewed as equally causal. The overall ANOVA reveals a main effect for normality, $F(1,130) = 248.19$, $p < 0.001$, $\eta p^2 = 0.66$, and a significant interaction between normality and question-type, $F(1,130) = 244.93$, $p < 0.001$, $\eta p^2 = 0.65$.

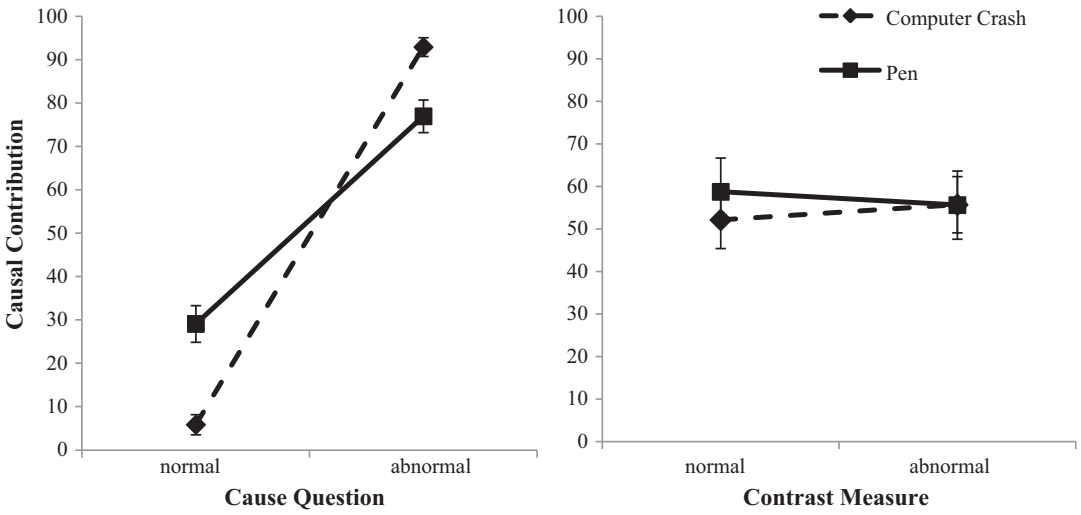

Fig. 2. Results of Experiment 1. The judgments for the cause question can take values between 0 and 100; the contrast measure, which is based on differences, can in principle take values between −100 and 100. However, only generative causes were used that entail positive values. Error bars represent standard errors of means (*SE*).

In sum, the results of Experiment 1 show that norms do not alter the size of the strength parameters of causal models: when participants were confronted with a measure that specifically targets intuitions about causal strength, causal judgments were unaffected by whether the protagonist violated a norm or conformed to it. The findings therefore cast doubt on versions of blame-based accounts (e.g., culpable control model) that assume that the causal strength of the norm-violating agent tends to be exaggerated.

3. Experiment 2

In Experiment 1 we have shown that prescriptive norms do not alter causal parameters. However, there is still the alternative that they may guide causal selection. According to Hitchcock and Knobe's (2009) counterfactual reasoning account of causal selection, norm effects can be explained by a two-stage counterfactual analysis (see also Halpern & Hitchcock, 2014). The first stage involves the setting up of a causal network that consists of causal variables referring to states of causes and effects (e.g., presence vs. absence). These variables refer to causal events, not objects or persons (see Waldmann & Mayrhofer, 2016). So, for instance, in the pen vignette, the cause variables may refer to the presence or absence of the event that a specific person took a pen. Causal relations can be expressed by referring to counterfactuals. For example, each agent in the pen vignette is assumed to be causal because the effect (lack of pens) would not have occurred in the absence of the action of the person. We do not question the counterfactual account of this initial stage; in fact, we have used it ourselves in Experiment 1.

Causal selection occurs in a second stage in which subjects pragmatically choose one of the causes as "the cause." This stage presupposes that each candidate cause has been already established as causal in the first stage. To explain why, for example, a lightning beam is chosen over oxygen in a forest fire, a well-established theory claims that it is covariation within the focal set of considered events that drives the selection (Cheng & Novick, 1991). Because oxygen is constantly present in the typically considered cases of forest fires, it is backgrounded. By contrast, the lightning beam is rare and covarying, and is therefore picked as "the cause." We agree with this account which is practically identical with Hitchcock and Knobe's (2009) analysis of such cases predicting the preferential selection of *statistically abnormal* events. Where we disagree is with the claim that different kinds of abnormality play the same role in causal selection, and that the effect established for statistical abnormality similarly applies to moral abnormality.

Experiments 2a and 2b test the prediction of the counterfactual reasoning account of causal selection that norms lead to causal selection through selective activation of relevant counterfactuals to morally abnormal events. This prediction can be divided into two assumptions (Hitchcock & Knobe, 2009; Kominsky et al., 2015; Phillips et al., 2015): (1) prescriptive norm violations trigger counterfactual thinking about the abnormal factor, and (2) a factor is chosen in a causal selection task when the contrast between the actual and the counterfactual value of the factor is relevant and reveals a difference (i.e., when the abnormal factor covaries with the effect).

The first assumption has been investigated in many different ways and there are a number of findings supporting increased reasoning about counterfactual states of abnormal events (e.g., Kahneman & Miller, 1986; McCloy & Byrne, 2000; N'gbala & Branscombe, 1995). Regarding the second assumption, by contrast, it is more difficult to draw clear conclusions from the existing literature. There is plenty of evidence showing that people use counterfactual contrasts to determine the existence and strength of causal relations, which we above designated the initial stage of a counterfactual analysis of causal relations (e.g., Gerstenberg & Tenenbaum, in press; Lagnado & Gerstenberg, in press; Spellman & Kincannon, 2001). However, it is less clear whether counterfactuals triggered by abnormality intuitions mediate causal selection. Walsh and Sloman (2009), for instance, argue that "(...) although the availability of a counterfactual alternative to a particular event may increase the perceived causal role of that event, counterfactual availability does not influence the likelihood that an event will be selected as "the cause" from a set of necessary conditions" (p. 189) (see also Mandel & Lehman, 1996, for a similar view). Moreover, it is questionable whether all kinds of abnormality are equivalent. Whereas Cheng and Novick (1991) present evidence demonstrating the role of statistical abnormality (i.e., covariation within a focal set) as a factor influencing causal selection, this does not imply that moral abnormality plays an equivalent role.

The best evidence for the potential role of prescriptive norms in influencing causal selection through counterfactual reasoning comes from a recent study by Phillips et al. (2015) that was published after we ran our Experiment 2a. In their Experiment 2, the authors focused on the pen vignette. After having replicated the well-established norm effect in a between-subjects design with a professor either conforming to or violating a norm, Phillips and colleagues ran a control experiment studying a variant of the pen vignette in which both the professor and the secretary were permitted to take pens. To study the role of counterfactual reasoning, Phillips et al. (2015) had subjects in one condition either reflect about counterfactual alternatives of the professor's action of taking pens or in a control condition summarize the cover story. The results demonstrate higher causal ratings for the professor when subjects were requested to reflect on counterfactuals than in the control condition.

Although these results seem to confirm the view that counterfactual reasoning about moral abnormality underlies causal selection, there are some problems with the study. First, as Phillips et al. (2015) acknowledge, the effect size in the counterfactual control study (their Experiment 2b) is about a third of the norm effect obtained in their Experiment 2a. The authors argue that instructing subjects to consider alternatives may not be as effective in triggering counterfactuals as the norm manipulation, but an alternative possibility is that the norm effect is triggered by a different mechanism (e.g., accountability assessments). A second problem of the study is that causal selection has only been indirectly tested. Only ratings for one of the agents, the professor, were reported, not for the other agent, the administrative assistant. Thus, although the ratings for the professor varied across conditions, it is not clear whether the ratings for the secretary were unaffected by the manipulations. Moreover, instructing subjects to focus on one of two factors may alert them to the possibility that the experimenter considers this factor particularly important, which could have introduced a demand characteristic. Finally, mediation analyses showed that the data equally fit the favored model according to which counterfactual reasoning precedes causal selection and a model in which counterfactual reasoning follows causal selection. We will argue in the General Discussion that the latter possibility is in fact consistent with the accountability hypothesis. Because of these problems we felt that it would be helpful to run further tests of the potential role of counterfactual reasoning in causal selection.

3.1. Experiment 2a

In this experiment we crossed the factors normality and counterfactual salience and measured causal selection by obtaining

ratings for both relevant agents. Experiment 2b follows up on the control study of Phillips et al. (2015) with additional questions that more comprehensively measure causal selection.

3.1.1. Method

3.1.1.1. Participants. 211 participants (mean age = 40.72, *SD* = 12.77) were randomly assigned to four conditions that were part of a larger study that was run online in the U.K. 78 subjects who did not correctly remember the normative status of the mentioned agents or the causal structure described in the story were excluded. Thus, the results are based on the remaining 133 subjects. Subjects earned 50 British pence for their participation.

3.1.1.2. Design and procedure. The design of the experiment was based on a 2 (normality: both agents normal vs. one abnormal) × 2 (counterfactual salience: high vs. low) × 2 (agent: Anna vs. Sue) structure with the last factor being manipulated within subject. Each participant was randomly assigned to one of the four resulting conditions: the baseline condition (both agents normal and low salience of counterfactuals), the normal counterfactual condition (both agents normal and a salient counterfactual for one of them), the abnormality condition (one agent abnormal and no salient counterfactuals), and the abnormal counterfactual condition (one agent abnormal and a salient counterfactual for the other one).

Participants in all conditions were presented with a scenario about a company's elevator that has a malfunction. If it is called simultaneously on the two different floors of the building, the system breaks down. One day, two employees, Anna and Sue, press the two elevator buttons on the ground and first floor simultaneously, therefore the elevator system breaks down. Participants in the baseline condition were presented with a version of the scenario in which it was made clear that both the agent on the ground floor (Anna) and the agent on the first floor (Sue) were allowed to call the elevator. In the normal counterfactual condition, both agents were allowed to press the respective elevator button, but for one of them, Anna, the counterfactual outcome was explicitly pointed out in the scenario and thereby made salient ("If she had not pressed the button, the system would not have broken down."). In the abnormality condition, a norm was introduced. The rule says that it is allowed to press the elevator button on the ground floor (where Anna pressed the button), but it is not allowed to use the elevator on the first floor (where Sue pressed the button). Sue, however, ignores the rule (as most employees do), which leads to the malfunction. The fourth condition is the abnormal counterfactual condition in which both the norm was mentioned and the counterfactual for the norm-conforming agent was made salient.

The test question was presented below the description of the scenario. Subjects were asked how much they agree with the statements about the two agents: "Anna (Sue) caused the collapse of the system." Responses were given on a 7-point Likert Scale ranging from "not at all" (1) to "completely" (7). At the end of the experiment, participants were requested to answer two comprehension questions (similar to those in Experiment 1) to demonstrate their understanding of the causal structure and of the norm (if applicable).

3.1.2. Results and discussion

Fig. 3 shows the results. As can be seen there, the norm effect was replicated. The norm-violating agent was rated as more causal than the norm-conforming one. When both agents did not violate norms, no difference can be seen. However, the salience of the counterfactual did not affect ratings, neither in the norm-violation conditions nor in the conditions in which no norm was violated. This pattern is borne out in the statistical analyses. An overall ANOVA reveals main effects for the factor normality, $F(1,129) = 14.9$, $p < 0.001$, $\eta p^2 = 0.1$, and the factor agent, $F(1,129) = 69.56$, $p < 0.001$, $\eta p^2 = 0.35$, as well as a significant interaction between these two factors, $F(1,129) = 89.44$, $p < 0.001$, $\eta p^2 = 0.41$. There is a significant interaction between the factors agent and counterfactual salience, but the effect size is very small, $F(1,129) = 4.13$, $p = 0.044$, $\eta p^2 = 0.03$, and practically of no relevance (see Fig. 3). More focused t-tests show that the two agents were rated differently in the conditions in which the norm was introduced, regardless of the presence, $t(29) = 4.96$, $p < 0.001$, $d = 1.281$, or the absence, $t(35) = 9.37$, $p < 0.001$, $d = 2.209$, of a salient counterfactual. However, the ratings for the two agents did not differ in the condition without a norm in which the counterfactual for one agent was mentioned, $t(27) = 1.79$, $p = 0.08$. Likewise, no significant difference between the agents can be seen in the baseline condition, $t(38) = 1.43$, $p = 0.16$.

Focusing on each individual agent the results show that the ratings for Sue were substantially higher when she violated a norm (in the abnormality condition) compared to not violating a norm (in the baseline condition), $t(73) = 6.20$, $p < 0.001$, $d = 1.433$. In the comparison between these two conditions, the ratings for Anna differed significantly, $t(73) = 2.66$, $p = 0.01$, $d = 0.614$. This effect is consistent with the superseding effect, discovered by

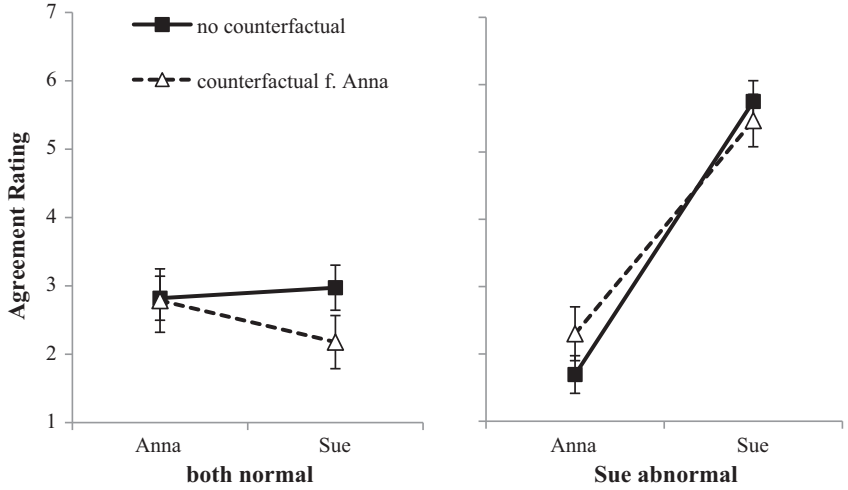

Fig. 3. Results of Experiment 2a. Error bars represent standard errors of means (SE).

Kominsky et al. (2015), which predicts that causal ratings for the norm-conforming agent should be lowered when the other agent has violated a norm. No effect can be seen for the norm-conforming Anna regardless of whether a counterfactual for her behavior was mentioned (normal counterfactual condition) or not (baseline condition), $t(65) = 0.06$, $p = 0.949$. Moreover, the ratings for Sue did not differ between the two conditions, $t(65) = 1.58$, $p = 0.118$. Thus, the counterfactual condition did not lead to a superseding effect, which is consistent with the finding that increasing the salience of the counterfactual did not influence causal selection. Thus, there is no evidence in the present experiment that the strong norm effect we observed is mediated by counterfactual reasoning.

3.2. Experiment 2b

Whereas in Experiment 2a we emphasized the salience of a counterfactual alternative by explicitly describing it, it cannot be guaranteed that participants considered this information relevant. It could be argued that it is not sufficient that a counterfactual alternative is just explicitly mentioned, but that subjects need to more actively reflect on counterfactuals to see their relevance (see Phillips et al., 2015).

Experiment 2b therefore uses a stronger manipulation by using the method Phillips et al. (2015) have used in their studies. Similar manipulations have been employed in several other studies before (Mandel, 2003; McCloy & Byrne, 2002). In their Experiment 2b, Phillips et al. (2015) presented two agents in a pen vignette who both were allowed to take pens. The counterfactual for one agent, the professor, was highlighted by asking subjects to reflect about possible alternative actions. The results of the experiment showed that the professor was rated more causal when subjects thought about counterfactuals than in the control condition. However, the effect was small (especially when compared to the norm effect). Moreover, since no ratings for the second agent were reported, it is unclear whether the manipulation indeed affected causal selection. Since in our Experiment 2a we did not find an effect of counterfactual highlighting in a design in which counterfactual reasoning was crossed with norm violation, we felt it would be interesting to see whether we can replicate the effect of Phillips et al. (2015) and test whether the manipulation indeed affects causal selection.

3.2.1. Method

3.2.1.1. Participants. 249 subjects (mean age = 27.82, $SD = 9.9$), recruited via a crowdsourcing platform with participants from many countries, were randomly assigned to two conditions. The number of subjects mirrors the statistical power in the experiment of Phillips et al. (2015). The experiment was run online and subjects earned 50 British pence for their participation. 11 participants did not correctly remember the normative status of the agents and were therefore excluded. Only the data of the remaining 238 subjects were used for the analyses.

3.2.1.2. Design and procedure. The design of the experiment is based on Study 2b of Phillips et al. (2015). It employs a 2 (consideration task: counterfactual vs. summary) × 2 (agent: Mrs. Smith vs. Mrs. Cooper) design with the last factor being manipulated within subject. Subjects first read a story that is similar to the neutral pen vignette used by Phillips et al. (2015) with the only difference being that both agents' names were introduced (Mrs. Smith and Mrs. Cooper) and that both agents work in different departments of the philosophical institute. This way we tried to control for possible effects of the difference of the professional status of the agents. After having read about the scenario, each participant was randomly assigned to one of two conditions: In the *counterfactual condition*, participants were asked to "(...) think about what other decisions Mrs. Smith (Department B) could have made, other than deciding to take a pen." In the *summary condition*, by contrast, no counterfactual reasoning was solicited. Instead, participants were asked to "(...) summarize and describe the events that actually happened in the vignette." Then in both conditions the question followed: "How much do you agree with the following statements?" In randomized order, subjects were asked to indicate their agreement to the two statements "Mrs. Cooper (Dept. A)/Mrs. Smith (Dept. B) caused the problem" using a scale ranging from 1 ("completely disagree") to 7 ("completely agree"). Finally, we tested whether participants remembered that both agents were allowed to take pens.

3.2.2. Results and discussion

Fig. 4 shows the results of the experiment. Replicating the findings by Phillips et al. (2015), we found a significant difference between the counterfactual condition ($M = 3.05$, $SD = 1.746$) and the summary condition ($M = 2.32$, $SD = 1.551$), $t(236) = 3.415$, $p < 0.001$, $d = 0.44$) for the ratings of Mrs. Smith. However, unlike in the original study we additionally tested whether this effect indeed reflects causal selection by also measuring ratings for the second agent, Mrs. Cooper. Interestingly, we found an effect for her in the same direction as for Mrs. Smith ($M = 2.88$, $SD = 1.673$ in the counterfactual condition; $M = 2.29$, $SD = 1.56$ in the summary condition), $t(236) = 2.80$, $p = 0.006$, $d = 0.36$. An overall ANOVA revealed a main effect for the consideration task, $F(1,236) = 10.38$, $p = 0.001$, $\eta p^2 = 0.04$, but, importantly, no significant interaction between agent and consideration task, $F(1,236) = 1.67$, $p = 0.197$.

This pattern shows that the selective activation of a counterfactual for one of the agents seemed to generally enhance ratings but did not lead to differential causal selection. Moreover, no causal superseding effect was found: the ratings for Mrs. Cooper were not diminished in view of Mrs. Smith's counterfactual alternative. A possible reason for the generally increased ratings might be that highlighting a counterfactual may have sensitized subjects to the causal roles of *both* agents.

In sum, none of the two studies demonstrates an effect of selective counterfactual reasoning on causal selection. In Experiment 2b we did find a small general effect of counterfactual reasoning which serves as a demonstration of the effectiveness of the manipulation. However, this manipulation did not affect causal selection, but influenced both causal factors. Experiment 2a replicated the norm effect showing again a far larger effect than what can be achieved by stimulating counterfactual reasoning. To account for this notable difference, Phillips et al. (2015) argue that norm manipulations may lead to more counterfactual reasoning than

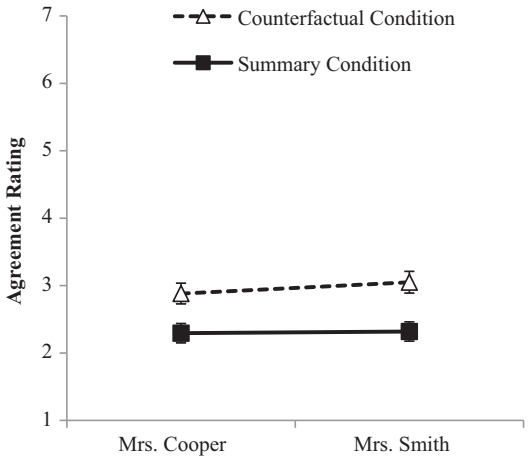

Fig. 4. Results of Experiment 2b. Error bars represent standard errors of means (SE).

the explicit instruction to reason about counterfactual alternatives. We find this implausible; there is no empirical evidence for this claim. In our view, the stronger effect of norms is due to a shift of judgments towards an evaluation of accountability.

4. Experiment 3

The accountability hypothesis attributes the norm effect to the ambiguity of the test question. Norm effects are only predicted when the pragmatic context invites an accountability interpretation of the test question. By contrast, unambiguous causal test questions should not be influenced by norms. This prediction was confirmed in Experiment 1 in which an unambiguous contrast measure was used to measure causal intuitions. Experiment 3 goes one step further and tests whether it is possible to disambiguate the intended meaning of the standard causal test question by manipulating pragmatic contextual features.

In most previous studies supporting the norm effect the causal test questions asked for the causal contribution of the agents (e.g., Professor Smith). The names of the agents were used as shorthand for the agents' actions. In the pen vignette it is Professor Smith's taking of a pen that causally contributes to the outcome of a shortage of pens, it is not the mere existence of Professor Smith that is the cause. Moreover, Professor Smith does many things in her life that conform to norms so that it is not the person Professor Smith who should be assigned an abnormal value in a causal model representation; it is her norm-violating action. Thus, in scenarios like the pen vignette the counterfactual reasoning account of causal selection predicts that subjects may think about alternatives to the action of an agent, not about the possible non-existence of the agent. However, using the name of an agent as a pointer to an abnormal action possibly creates a context that invites the interpretation of the test question as a request to assess accountability. It is agents and not events that are held accountable for an outcome.

To be able to manipulate the intended meaning of the test question, we used a scenario in Experiment 3 in which norms unambiguously regulated actions independent of the person who commits the actions. This way there could be no doubt that the relevant cause variables in the causal model refer to the presence or absence of action events, which either can be conceived of as conforming to a norm or violating it. In the cover story we presented a fertilizer scenario in which different gardeners use different chemicals on plants. Then a norm is introduced that forbids the use of one of the chemicals to prevent the plant from drying. Notably the norm directly refers to the employment of a specific chemical, not to a specific person. A counterfactual account should therefore specify the presence of the action of using a specific forbidden chemical as the abnormal value of the causal variable that encodes the relevant action event. The key manipulation of Experiment 3 was the framing of the test question. In the *person condition* we asked about each gardener whether his act of fertilizing the plant caused the drying. In the corresponding *chemical condition*, we asked whether the fertilization with the chemical led to the bad outcome. Both framings refer to the same event, and, if anything, the phrasing of the chemical condition is closer to what has been stated in the norm. Thus, the counterfactual reasoning account of causal selection should clearly predict a prescriptive norm effect in both conditions. By contrast, the accountability hypothesis predicts a strong norm effect in the person condition because it is people who are held accountable for bad outcomes, whereas questions about the causal role of chemicals should tend to be interpreted as queries about the causal mechanisms. Given that both chemicals, regardless of whether their use is permitted or forbidden, equally contribute to the effect, no difference is predicted in the chemical condition.

4.1. Method

4.1.1. Participants

50 subjects (mean age = 22.52, $SD = 3.61$) took part in the computer-based experiment that was run in a computer laboratory of the University of Göttingen. Subjects earned 5€ for their participation in a battery of experiments. Only subjects who correctly remembered the normative status of the mentioned agents and the causal relations described in the story were included, which left 43 participants for the analyses.

4.1.2. Design and procedure

We manipulated the framing of the test question (person vs. chemical) between subjects. Each participant read a story about a plant lover, Tom, who employs two gardeners, Alex and Benni. After Tom has read in a magazine that his plants would be even bigger and more beautiful if they were fertilized with chemicals, such as A X200® or B Y33®, he decides to let his gardeners use chemicals. However, because he has also learned that too many different chemicals can cause damage to the plants when used simultaneously, he tells his gardeners to only use one fertilizer, A X200®. When after several weeks he realizes that some of his plants have dried up while others have indeed grown and become more beautiful, he talks to his gardeners and finds out that one of them, Benni, has used the forbidden fertilizer B Y33® instead of A X200®. Tom discovers that the plants which were exposed to two different fertilizers have dried up, while the plants which were fertilized with only one single chemical have become bigger and more beautiful.

After reading this cover story, participants were asked the following question: "What caused the drying of the plants?" The answer options differed between the two conditions to which subjects were randomly assigned. In the *person* condition, participants were asked to choose between "the fertilization by Alex", "the fertilization by Benni" or they could opt for both. In the *chemical* condition, the answer options were "the application of chemical A X200®", "the application of chemical B Y33®" and again subjects could also choose both options. It is important to note that the framing of the causal query in the chemical condition refers to the same events that are being regulated by the prescriptive norm. To make sure that the chemicals were assigned to the right agents, two control questions followed on the next page ("Which fertilizer was used by Alex/Benni?"). Next, participants were asked whether it was allowed that Alex used A X200® and that Benni used B Y33®. Subjects who did not give the right answers to these four questions were excluded from the reported analyses. Subsequently, to test their understanding of the conjunctive causal structure participants were given a scale ranging from 0 to 100 to estimate the percentage of flower beds in which the plants had dried up given that (i) only A X200®, (ii) only B Y33®, or (iii) both A X200® and B Y33® had been applied.

4.2. Results and discussion

Fig. 5 shows the results. The answer pattern in the person condition is significantly different from the answer pattern in the chemical condition, $\chi 2(1, N = 43) = 12.22$, $p < 0.001$. Only in the person condition was the norm-violating action significantly selected over the norm-conforming action, $\chi 2(1, N = 11) = 11$, $p < 0.001$, compared to $\chi 2(1, N = 1) = 1$, $p = 0.317$, in the chemical condition.[3] In the person condition, responses were equally distributed across the option norm-violating agent and the option that

[3] Note that the expected cell frequency is less than 1 for this comparison so that the chi-square approximation may not be reliable (see Cochran, 1954).

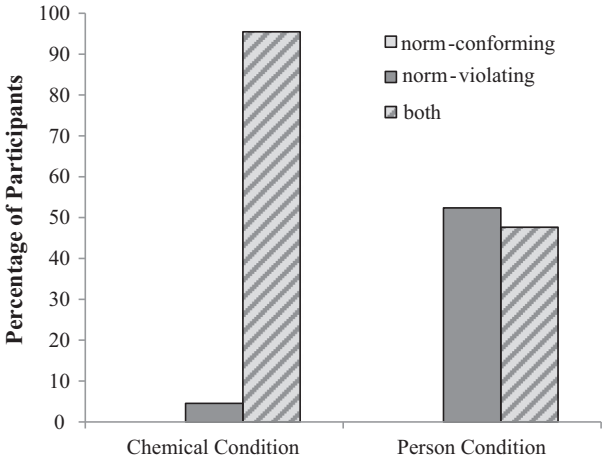

Fig. 5. Results of Experiment 3 (see text for explanations).

both gardeners contributed to the effect, $\chi 2(1, N = 21) = 0.048$, $p = 0.827$. In the chemical condition, by contrast, only one subject selected the forbidden chemical as the cause, with the clear majority choosing the option that both caused the drying of the plants, $\chi 2(1, N = 22) = 18.18$, $p < 0.001$.

In sum, in this experiment a norm was introduced that referred to the action of using a specific chemical. Thus, according to the counterfactual reasoning account of causal selection the use of the forbidden chemical should be seen as the abnormal value regardless of whether the causal query refers to the action by highlighting the agent or the chemical involved in the action. Nevertheless, we obtained a clear norm effect only when we asked about the actions of the gardener who either conformed to the norm or violated it, whereas similar queries about the use of the permitted or forbidden chemical did not show a norm effect. This pattern provides strong support for the accountability hypothesis.

5. Experiment 4

A major shortcoming of the counterfactual reasoning account of causal selection is that the concept of abnormality is only vaguely specified. In the experiments testing effects of social norms, forbidden actions (e.g., taking pens) have typically been used as test cases. In Experiment 3, however, we have shown that the abnormality of an action alone is not sufficient to generate a norm effect; apart from the norm-violating action the agent needed to be mentioned. The goal of the present experiment is to provide further support for the accountability hypothesis. One advantage of this hypothesis is that it suggests additional factors affecting accountability apart from the action and the outcomes they cause, such as the foreseeability of the outcome, the agents' intentions, or the agents' knowledge of relevant norms (see Cushman, 2008; Lagnado & Channon, 2008; Malle et al., 2014; Samland et al., 2016; Young & Saxe, 2011). All these factors influence the assessment of the degree of accountability, and should therefore, according to our theory, also affect causal selection.

To test whether these additional factors influence causal selection, the present experiment systematically varied the description of the mental states of the norm violator and, consequently, the reasons underlying the norm transgression. Judged accountability should be diminished if the agent did not intentionally perform the norm-violating action or if she did not know that she violated a norm. The key question that the present experiment addresses is whether these factors moderating accountability will also have an effect on causal selection.

The experiment tests these predictions by extending the methods used in Experiment 3 and in Samland et al. (2016). Samland et al. (2016) showed that norm-violating agents were only selected as the cause by adult subjects when the agents actually knew about the existence of the relevant norm. In the present experiment both the intentionality of one of the agents and several variants of lack of knowledge about the applicable norm were manipulated. We varied whether one of the agents intentionally violated a norm or just by accident although he knew about the norm. In further conditions we manipulated whether the norm-violating agent was ignorant about the stated norm because he was not informed or because another agent deceived him for selfish reasons. The key question motivating all these variations is whether factors known to reduce accountability also affect causal selection.

Following up on Experiment 3 we also manipulated the phrasing of the causal query. Again we expected a stronger norm effect when agents were mentioned compared to chemicals. To provide an additional test of this hypothesis we added a condition in which we offered only the names of the agents as candidates for causal selection without mentioning their actions.

5.1. Method

5.1.1. Participants

869 subjects (mean age = 32.14, $SD = 12.04$), recruited via a crowdsourcing platform, participated in the online study. All subjects earned 50 British pence for their participation. 87 participants were excluded because they could not remember the assignment of fertilizers to agents. An additional 71 participants were removed from the sample because they did not correctly indicate the agents' normative status, and finally 126 participants were excluded because they did not understand the conjunctive causal structure of the scenario (i.e., that the likelihood of drying up was higher when two different fertilizers were used compared to only one type). This left us with the data of 584 participants.

5.1.2. Design and procedure

In the experiment we used the cover story of Experiment 3. Thus, participants were presented with the story about Tom and his two gardeners who were only allowed to use one type of fertilizer because plants that were treated with two different fertilizers dried up. Again Benni did not follow the order and violated the prescriptive norm. Unlike in Experiment 3 we varied the reasons for the norm transgression of Benni. Furthermore, we added a condition with a test question that directly asked about the agents, using their names without mentioning the action. The experiment was based on a 4 (norm transgression: standard vs. unintended vs. ignorant vs. deception-based) $\times$ 3 (question type: chemical/event vs. person/event vs. agent) between-subjects design.

In four conditions we manipulated mental state features that should influence accountability assessments. In the first condition, *standard norm transgression*, the cover story was identical to the one used in Experiment 3. In this condition, the norm-violating agent, Benni, intentionally applies the chemical B Y33® although he knows that he is not allowed to use this fertilizer. In a second condition, *unintended norm transgression*, Benni intends to conform to Tom's rule and to use the chemical A X200® but he accidentally uses the wrong can which contains the forbidden chemical B Y33®. In a third condition, *ignorant norm transgression*, Benni intentionally uses the chemical B Y33® but he does not know Tom's rule because Alex forgot to tell him that they are only allowed to utilize chemical A X200®. In a fourth condition, *deception-based norm transgression*, Benni intentionally uses the chemical B Y33® but again he is not aware that this chemical has been forbidden by Tom. The reason for his lack of knowledge is that Alex wants him

to get fired and therefore deceives him by telling him that Tom wants them to use chemical B Y33®.

After having read one of these four cover stories, participants within each condition were randomly assigned to one of three conditions that varied the causal test question. The *chemical/event* and *person/event* questions were identical to the ones used in Experiment 3. Thus, subjects were asked "what caused the drying of the plants?", and were given either "the fertilization by Alex (Benni)", or both, or "the application of chemical A X200 (B Y33®)", or both as response options. In a third condition, the *agent* question condition, participants were just offered the options "Alex", "Benni", or both. We included this condition because this test question has been used most frequently in previous studies, and because the accountability hypothesis predicts the strongest norm effect when only the agents' names are offered as possible causes.

As in Experiment 3, several control questions served as comprehension checks. Subjects were asked about the assignment of chemicals to agents, about the normative status of the different actions, and about the conjunctive causal structure. To test their knowledge of the causal structure, participants were asked to estimate the percentage of flower beds in which the plants dried up given that (i) only A X200®, (ii) only B Y33®, or (iii) both A X200® and B Y33® had been applied. These questions were used to screen subjects who did not pay sufficient attention to the instructions (see Section 5.1.1).

5.2. Results and discussion

Fig. 6 shows the results of the experiment. Within each story condition the answers to the three question types differed significantly (standard condition: $\chi 2(2, N = 148) = 43.02$, $p < 0.001$; unintended norm transgression: $\chi 2(4, N = 124) = 26.89$, $p = 0.001$; ignorant norm transgression: $\chi 2(4, N = 161) = 48.8$, $p < 0.001$; deception-based norm transgression: $\chi 2(4, N = 151) = 64.42$, $p < 0.001$). In the standard norm transgression condition we replicated the results of Experiment 3. Thus, a norm effect was only seen when the agent was mentioned, either along with the action (person/event) or alone (agent). In the chemical/event condition the answer "both" was selected more frequently than the norm-violating action, $\chi 2(1, N = 55) = 27.66$, $p < 0.001$, while in the person/event condition responses were equally distributed across the option norm-violating event and the option that both actions contributed to the effect, $\chi 2(1, N = 51) = 1.59$, $p < 0.208$. In the agent question condition in which only the agents' names were presented as answer options, the majority of participants in the standard norm transgression condition selected the norm-violating agent, $\chi 2(1, N = 34) = 34$, $p < 0.001$. The frequency of "both"-answers was strictly monotonically decreasing from the chemical/event over the person/event to the agent condition while the frequency of selections of the norm-violating agent (Benni) was strictly monotonically increasing across these three response conditions ($T = 6.49$; see Pfanzagl, 1974). Thus, the more the agent was foregrounded in the causal query, the stronger the norm effect became.

For the chemical/event question, the answer pattern did not differ between the four stories ($\chi 2(6, N = 169) = 4.23$, $p = 0.646$). For the other two question types, by contrast, the answer pattern changed, with the agent question leading to the strongest effect ($\chi 2(6, N = 212) = 13.71$, $p = 0.03$, for the person/event question; $\chi 2(6, N = 203) = 79.72$, $p < 0.001$, for the agent question). Whereas the responses to the chemical/event question differed significantly from the answers to the person/event question in the standard norm transgression condition, $\chi 2(2, N = 106) = 9.44$, $p = 0.009$, the answers to these two types of causal queries did not significantly differ in the other three story versions ($\chi 2(2, N = 81) = 5.27$, $p = 0.07$ for the unintended norm transgression; $\chi 2(2, N = 95) = 2.78$, $p = 0.25$, for the ignorant norm transgression; $\chi 2(2, N = 99) = 5.23$, $p = 0.07$, for the deception-based norm transgression). Only in the three story versions that mention circumstances reducing Benni's accountability, participants chose the option "both" more often than the norm-violating event alone ($\chi 2(1, N = 47) = 4.79$, $p = 0.03$ for the unintended norm transgression; $\chi 2(1, N = 50) = 8$, $p = 0.005$, for the ignorant norm transgression; $\chi 2(1, N = 59) = 8.97$, $p = 0.003$, for the deception-based norm transgression condition).

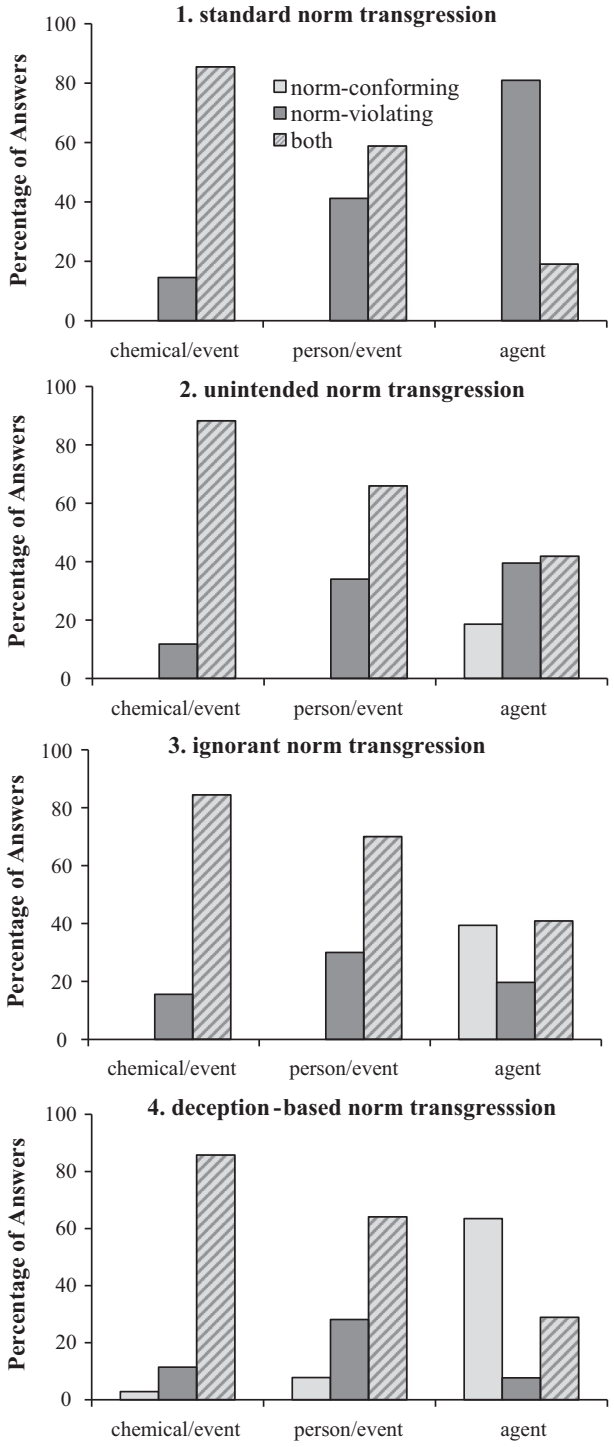

Fig. 6. Results of Experiment 4 (see text for explanations).

The responses to the agent question were affected most strongly by the manipulated mitigating factors. The frequency of

selections of the norm-violating Benni was strictly monotonically decreasing across the four cover stories in the agent condition ($T = 7.7$; see Pfanzagl, 1974) while the frequency of selections of the norm-conforming Alex was strictly monotonically increasing ($T = 6.88$; see Pfanzagl, 1974). In the standard norm transgression condition in which Benni violates the norm intentionally and knowingly he was preferentially selected over the option "both", $\chi 2(1, N = 42) = 16.1$, $p < 0.001$. In the unintended norm transgression condition in which Benni did not intend to violate the norm, the answer "both" was chosen equally often as Benni, $\chi 2(1, N = 35) = 0.029$, $p = 0.866$. Thus, the answer pattern differed significantly from the one in the standard norm transgression condition, $\chi 2(2, N = 85) = 17.5$, $p < 0.001$, demonstrating the relevance of the norm-violator's intentionality for causal selection. In the ignorant norm transgression condition, the answer "both" was chosen more often than the norm-violating agent, $\chi 2(1, N = 40) = 4.9$, $p = 0.03$, and equally often as the norm-conforming one, $\chi 2(1, N = 53) = 0.019$, $p = 0.891$. This answer pattern again differed significantly from the one in the standard norm transgression condition, $\chi 2(2, N = 108) = 42.46$, $p < 0.001$. Finally, in the deception-based norm transgression condition, Alex who deceived Benni was actually chosen more frequently than Benni whose behavior is actually violating the norm, $\chi 2(1, N = 37) = 22.73$, $p < 0.001$. This reversal relative to the standard norm transgression condition, $\chi 2(2, N = 94) = 58.41$, $p < 0.001$, provides strong evidence for the hypothesis that it is accountability and not the abnormality of the action that drives causal selection.

In sum, Experiment 4 strengthens the support for the accountability hypothesis. The key new finding is that causal selection was not only sensitive to the abnormality of the actions but also to mental state factors that are known to moderate accountability judgments. Despite invariant behavior, agents who violated the norm were chosen less frequently when they did not intend the norm violations or when they did not know that they broke a norm. When they were deceived it was actually the person who deceived them and not the norm-violator who was chosen most frequently as the cause in the condition in which the names of the agents were offered as response options. A second important finding is that we again found clear sensitivity to the phrasing of the causal test question. The strongest norm effects were seen when the names of agents were mentioned (either alone or along with the forbidden action), whereas no norm effect was observed when the chemicals were foregrounded in the description of the forbidden action, although it is the application of chemicals, not agents, that was subject to norm regulations.

6. General discussion

Our main aim was to re-visit the question how prescriptive norms influence causal judgments. In the present set of studies, we tested three theories. The culpable control model of blame (Alicke, 2000) predicts that the wish to blame a norm-violating agent might lead to an exaggeration of her causal influence. Experiment 1 therefore investigated whether norms influence causal strength parameters but did not find any effect: judgments about causal strength were not influenced by prescriptive norms when the action and the outcome were otherwise held constant.

This leaves causal selection as the remaining possibility for a norm influence. Consistent with the counterfactual reasoning account of causal selection (Hitchcock & Knobe, 2009), causal selection can be regarded as a two-stage process. In the first stage counterfactuals are used to establish the factors that are causal. Thus, this theory is consistent with Experiment 1 which demonstrates that subjects were aware of the fact that both agents in the pen vignette are causal and jointly necessary for the effect. Causal selection occurs in the second stage in which the causal relevance of the variables has already been established. According to the counterfactual theory, in the second stage people tend to select an abnormal factor as the cause because they tend to consider counterfactual alternatives for abnormal but not for normal factors (see also Kominsky et al., 2015; Phillips et al., 2015). Contrary to the predictions of this account, we have found no convincing evidence for the claim that norms affect causal selection through counterfactual reasoning. Both of our attempts to induce counterfactual reasoning did not affect causal selection (see Experiment 2), thereby casting doubt on the conclusions of Phillips et al. (2015). Of course, one can always argue that our manipulations were too weak or ineffective. However, so far no manipulation of counterfactual reasoning has been presented in the literature on prescriptive norm violations that creates an effect on causal selection that in terms of size comes near to what can be accomplished by norm manipulations.

We have therefore proposed an alternative theory, the accountability hypothesis, which states that causal queries are generally ambiguous. They might refer either to causal relations in the narrow sense, or they might request assessments of moral accountability. Accountability entails causal relations, as agents are only held accountable for events they have actually directly or indirectly caused. But accountability judgments are sensitive to further factors: the intentionality of the agents, the relation of the acts to norms and values, and the expected utility of the outcomes for the agent and for society are also relevant.

Thus, there are commonalities and differences between the counterfactual reasoning account of causal selection and the accountability hypothesis. Both theories assume that in an initial stage the causal relevance of the factors under consideration need to be established. Counterfactual reasoning may well play a role in this phase in both theories. Whereas in the counterfactual reasoning account of causal selection this phase is simply a pre-condition of causal selection that takes place in a later phase, in the accountability theory causal relevance is part of a set of factors that jointly establish the degree of accountability. A second difference concerns the role of counterfactual reasoning in the causal selection phase. Whereas this process is crucial for explaining causal selection in the counterfactual theory, it does not play a role in the accountability theory. Finally, a crucial difference is that the counterfactual theory has so far used a fairly primitive concept of norm violation (or abnormality) which, at least in the studies on social norm violations published so far, simply refers to behavior that violates a prescriptive norm. By contrast, the accountability hypothesis claims that causal selection in tasks about social domains is often influenced by the degree of accountability ascribed to the agents, which in turn is sensitive to a complex set of potentially mitigating factors.

Experiment 3 represented an initial test of the accountability theory against the counterfactual reasoning account of causal selection (Hitchcock & Knobe, 2009). We presented subjects with a scenario in which two agents jointly caused a negative effect with one of the two actions being prohibited. In this experiment the norms referred to the use of specific fertilizers so that using a forbidden fertilizer would constitute the abnormal event within the counterfactual reasoning account of causal selection. To test this account against the accountability hypothesis, we asked subjects in the test phase about the causes but varied which component of the causal event was highlighted. We found that a norm effect was only seen when participants had to choose between event descriptions that included the agents' names – not when subjects had to choose between descriptions that focused on material components (i.e., chemicals) and hence the causal mechanisms. This

held true although it is these components that were regulated by the norm. The findings of the experiment support the accountability hypothesis which predicts that people, not objects, are held accountable for a negative outcome.

Experiment 4 provided further support for the accountability hypothesis, showing that factors additionally influencing accountability judgments also affect causal selection. If presented with a story in which mitigating circumstances explaining the norm transgression as unintended are introduced, the norm transgressor was less often selected as cause compared to a story in which he is fully accountable. Thus, it is not the abnormality of the committed action that drives causal selection but accountability. The most salient finding in this experiment was that in a situation in which the norm-violating agent was portrayed as an innocent victim who has been deceived by a second agent, it was this second agent who was selected over the norm-violating agent in a causal selection query.

A proponent of the counterfactual theory of causal selection could point out here that the deceiving agent also violated a norm, which may give rise to additional counterfactual reasoning. However, without a more elaborate account of abnormality that considers mitigating circumstances due to hierarchical relations between norm violations of different agents it is unclear what predictions this account makes for such cases. The hypothesis that subjects assess accountability of the two agents provides a more intuitive explanation of our findings.

A second important finding in the experiment, which replicates some of the results of Experiment 3, was that the more the agents were foregrounded in the answer options to the causal selection question, the stronger the norm effect became. If subjects had to choose between answer options in which the agents' names were not mentioned, both the norm-violating and the norm-conforming event were chosen as equally valid causes. The strongest norm effect was seen when only the names of the agents were mentioned.

Our work blends in with other research that supports the accountability hypothesis. Samland et al. (2016), for instance, showed that the agent's knowledge about the existence of the relevant norm is critical for obtaining the norm effect. Adult subjects only selected agents as causal who knew that they violated the norm. By contrast, children focused more on the behavior violating the norm and disregarded knowledge. These results fit with what we know about the development of blame and accountability judgments, but are hard to reconcile with the current version of the counterfactual theory of causal selection (Kalish & Cornelius, 2007; Killen, Mulvey, Richardson, Jampol, & Woodward, 2011; Yuill & Perner, 1988).

Another study supporting the accountability hypothesis was presented by Sytsma et al. (2012) who found that the norm-violating agent is only selected if it is typical for her to violate the norm, not when it is atypical. Sytsma and colleagues conclude that it is responsibility and not causality that is being evaluated.

Finally, Danks, Rose, and Machery (2014) found that normative evaluations cease to influence causal inferences if causal information is experienced in a trial-by-trial learning setting and not when it is described. Presenting causal information in an experience-based fashion is an alternative way of putting emphasis on the causality- rather than the accountability-meaning of the causal query (see also Samland & Waldmann, 2015, for similar results).

So far we have treated the counterfactual and the accountability theory as competitors. However, counterfactual reasoning may well be a component of determining accountability. It may be that accountability judgments trigger counterfactual thinking about what the blameworthy agent should have done instead, whereas we may not tend to think about alternative actions if the agent had done the right thing. However, it seems more likely that such counterfactual thoughts follow accountability judgments rather than precede them. This causal ordering is actually consistent with the results of Phillips et al. (2015).

A possible strategy to reconcile the counterfactual theory with our claim that causal queries are ambiguous is to postulate that different test questions invoke different norms. Whereas queries mentioning the norm-violating agent might highlight the prescriptive norm in the scenario, it is possible that a query asking about a component of the mechanism might instead activate norms of proper functioning (see Hitchcock & Knobe, 2009). This account seems possible as an explanation of results of Samland and Waldmann (2015, Experiment 2) in which we presented a variant of the pen vignette in which buttons needed to be pressed to obtain pens. We found the standard norm effect when we asked about the agents, but no such effect was seen when we asked about the buttons. It is possible that when asked about buttons subjects thought this was a query about the proper functioning of the buttons. Since neither button violated a norm of proper functioning (both worked as they were supposed to), a counterfactual account referring to these norms would also predict the absence of causal selection. To remove this possibility in the present Experiment 3, we used a norm that directly prohibited the use of a chemical as a fertilizer. Thus, both the prescriptive norm and the test question asking about the causal role of the chemicals referred to the same event of using a forbidden chemical. The counterfactual theory clearly predicts in this case that the prescriptive norm should lead to causal selection of the forbidden chemical.

Although norms of proper functioning do not explain our findings in Experiment 3, it is an interesting topic for future research whether causal judgments are influenced by such norms, and whether it is necessary to separate norms of proper functioning from prescriptive and statistical norms. One general problem of the counterfactual theory of causal selection, which also applies to norms of proper functioning, is that we did not find evidence for the claim that counterfactual reasoning triggers causal selection (Experiment 2), which is the mechanism postulated as underlying effects of all kinds of norm violations. Moreover, to date no strong evidence has been put forward demonstrating that norms of proper functioning affect causal selection independently of statistical norms. In a study discussed by Hitchcock and Knobe (2009) ("machine vignette") subjects chose a defect red wire over a functioning black wire as the cause of the malfunction of a machine. In the scenario, the loose red wire occasionally touches a battery which in these cases leads to the malfunction, whereas the black wire works as expected in all cases. A plausible explanation of the causal selection of the red wire here might be that subjects chose the factor that covaries with the effect within the focal set over a constantly present co-factor (Cheng & Novick, 1991). This is the standard account of causal selection in tasks varying statistical relations, and therefore there is no need to postulate a separate mechanism for norms of proper functioning.

If counterfactual reasoning is modeled as a component of accountability judgments, it seems therefore necessary to give up the search for a domain-general version of a theory of causal selection. It is true that one attractive feature of the present version of the counterfactual reasoning account of causal selection is that it seems to apply to all kinds of abnormality in general. However, despite the general advantage of parsimony we think that it is necessary to distinguish between statistical abnormality (see Cheng & Novick, 1991) and prescriptive abnormality. Norms of proper functioning may or may not be added to this list depending on the outcomes of future studies. Using a highly abstract concept of abnormality in all these cases may make us miss important distinguishing features. The present studies just represent a first step in elucidating the complex set of factors underlying the relationship between norms and causal selection.

Appendices A and B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.07.007.

References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology, 63*, 368–378. http://dx.doi.org/10.1037/0022-3514.63.3.368.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*(4), 556–574. http://dx.doi.org/10.1037/0033-2909.126.4.556.

Alicke, M. D., Mandel, D. R., Hilton, D., 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. http://dx.doi.org/10.1177/1745691615601888.

Alicke, M. D., Rose, D., & Bloom, D. (2011). Causation, norm violation, and culpable control. *Journal of Philosophy, 108*, 670–696. http://dx.doi.org/10.5840/jphil20111081238.

Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition, 40*, 83–120. http://dx.doi.org/10.1016/0010-0277(91)90047-8.

Cochran, W. G. (1954). Some methods of strengthening the common chi-square tests. *Biometrics, 10*, 417–451. http://dx.doi.org/10.2307/3001616.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*(2), 353–380. http://dx.doi.org/10.1016/j.cognition.2008.03.006.

Danks, D., Rose, D., & Machery, E. (2014). Demoralizing causation. *Philosophical Studies, 171*, 251–277. http://dx.doi.org/10.1007/s11098-013-0266-8.

Deigh, J. (2008). Can you be morally responsible for someone's death if nothing you did caused it? In W. Sinnott-Armstrong (Ed.), *Moral psychology* (pp. 449–461). Massachusetts: MIT Press.

Gailey, J. A., & Falk, R. F. (2008). Attribution of responsibility as a multidimensional concept. *Sociological Spectrum, 28*, 659–680. http://dx.doi.org/10.1080/02732170802342958.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2014). From counterfactual simulation to causal judgment. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society* (pp. 523–528). Austin, TX: Cognitive Science Society.

Gerstenberg, T., & Tenenbaum, J. (in press). Intuitive theories. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning*. New York: Oxford University Press.

Halpern, J. Y., & Hitchcock, C. (2014). Graded causation and defaults. *The British Journal for the Philosophy of Science, 66*, 413–457. http://dx.doi.org/10.1093/bjps/axt050.

Hart, H. L. A., & Honoré, A. M. (1959). *Causation in the law*. Oxford: Oxford University Press.

Hesslow, G. (1988). The problem of causal selection. In D. Hilton (Ed.), *Contemporary science and natural explanation: Commonsense conceptions of causality* (pp. 11–32). Brighton: Harvester Press.

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal condition focus model. *Psychological Review, 93*, 75–88. http://dx.doi.org/10.1037/0033-295x.93.1.75.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy, 106*, 587–612. http://dx.doi.org/10.5840/jphil20091061128.

Kahneman, D., & Miller, D. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 80*, 136–153. http://dx.doi.org/10.1037/0033-295x.93.2.136.

Kalish, C. W., & Cornelius, R. (2007). What is to be done? Children's ascriptions of conventional obligations. *Child Development, 78*(3), 859–878. http://dx.doi.org/10.1111/j.1467-8624.2007.01037.x.

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition, 119*, 197–215. http://dx.doi.org/10.1016/j.cognition.2011.01.006.

Knobe, J. (2005). *Attribution and normativity: A problem in the philosophy of social psychology* Unpublished manuscript. University of North Carolina-Chapel Hill.

Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (pp. 441–447). Massachusetts: MIT Press.

Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D. A., & Knobe, J. (2015). Causal superseding. *Cognition, 137*, 196–209. http://dx.doi.org/10.1016/j.cognition.2015.01.013.

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition, 108*, 754–770. http://dx.doi.org/10.1016/j.cognition.2008.06.009.

Lagnado, D. A., & Gerstenberg, T. (in press). Causation in legal and moral reasoning. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning*. New York: Oxford University Press.

Lewis, D. (1973). Causation. *Journal of Philosophy, 70*, 556–567. http://dx.doi.org/10.2307/2025310.

Liu, B. S., & Ditto, P. H. (2013). What dilemma? Moral evaluation shapes factual belief. *Social Psychological and Personality Science, 4*, 316–323. http://dx.doi.org/10.1177/1948550612456045.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry, 25*, 147–186. http://dx.doi.org/10.1080/1047840X.2014.877340.

Mandel, D. R. (2003). Effect of counterfactual and factual thinking on causal judgments. *Thinking & Reasoning, 9*, 245–265. http://dx.doi.org/10.1080/13546780343000231.

Mandel, D. R., & Lehman, D. R. (1996). Counterfactual thinking and ascriptions of cause and preventability. *Journal of Personality and Social Psychology, 71*, 450–463. http://dx.doi.org/10.1037/0022-3514.71.3.450.

McCloy, R., & Byrne, R. M. (2000). Counterfactual thinking about controllable events. *Memory & Cognition, 28*(6), 1071–1078. http://dx.doi.org/10.3758/bf03209355.

McCloy, R., & Byrne, R. M. J. (2002). Semifacutal "even if" thinking. *Thinking and Reasoning, 8*(1), 41–67. http://dx.doi.org/10.1080/13546780143000125.

N'gbala, A., & Branscombe, N. R. (1995). Mental simulation and causal attribution: When simulating an event does not affect fault assignment. *Journal of Experimental Social Psychology, 31*, 139–162. http://dx.doi.org/10.1006/jesp.1995.1007.

Noveck, I. A., & Reboul, A. (2008). Experimental Pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences, 12*, 425–431. http://dx.doi.org/10.1016/j.tics.2008.07.009.

Paul, L. A., & Hall, N. (2013). *Causation: A user's guide*. Oxford, UK: Oxford University Press.

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. http://dx.doi.org/10.1016/j.cognition.2015.08.001.

Pfanzagl, J. (1974). *Allgemeine Methodenlehre der Statistik II*. Berlin: De Gruyter.

Samland, J., Josephs, M., Waldmann, M. R., & Rakoczy, H. (2016). The role of prescriptive norms and knowledge in children's and adults' causal selection. *Journal of Experimental Psychology: General, 145*, 125–130. http://dx.doi.org/10.1037/xge0000138.

Samland, J., & Waldmann, M. R. (2015). Highlighting the causal meaning of causal test questions in contexts of norm violations. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th annual conference of the cognitive science society* (pp. 2092–2097). Austin, TX: Cognitive Science Society.

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York: Springer-Verlag.

Sloman, S. A., Fernbach, P. M., & Ewing, S. (2009). Causal models: The representational infrastructure for moral judgment. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.). *Moral judgment and decision making* (Vol. 50, pp. 1–26). San Diego, CA, US: Elsevier Academic Press.

Spellman, B. A., & Kincannon, A. (2001). The relation between counterfactual ("but for") and causal reasoning: Experimental findings and implications for jurors' decisions. *Law and Contemporary Problems: Causation in Law and Science, 64*, 241–264. http://dx.doi.org/10.2307/1192297.

Suganami, H. (2011). Causal explanation and moral judgement: Undividing a division. *Millennium: Journal of International Studies, 39*(3), 717–734. http://dx.doi.org/10.1177/0305829811402809.

Sytsma, J., Livengood, J., & Rose, D. (2012). Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 43*, 814–820. http://dx.doi.org/10.1016/j.shpsc.2012.05.009.

Waldmann, M. R., & Hagmayer, Y. (2013). Causal reasoning. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 733–752). New York: Oxford University Press.

Waldmann, M. R., & Mayrhofer, R. (2016). Hybrid causal representations. *The Psychology of Learning and Motivation* (pp. 85–127). New York: Academic Press.

Walsh, C. R., & Sloman, S. A. (2009). Counterfactual and generative accounts of causal attribution. In P. McKay Illari, R. Russo, & J. Williamson (Eds.), *Causality in the science* (pp. 184–201). Oxford: University Press.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford Press.

Wiegmann, A., Samland, J., & Waldmann, M. R. (2016). Lying despite telling the truth. *Cognition, 150*, 37–42. http://dx.doi.org/10.1016/j.cognition.2016.01.017.

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition, 120*, 202–214. http://dx.doi.org/10.1016/j.cognition.2011.04.005.

Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children's judgments of actors responsibility and recipients emotional reaction. *Developmental Psychology, 24*(3), 358–365. http://dx.doi.org/10.1037/0012-1649.24.3.358.