in press at *Cognition*

Causal judgments about atypical actions are influenced by agents' epistemic states

Lara Kirfel*
University College London

David Lagnado
University College London

Abstract

A prominent finding in causal cognition research is people's tendency to attribute increased causality to atypical actions. If two agents jointly cause an outcome (conjunctive causation) but differ in how frequently they have performed the causal action before, people judge the atypically acting agent to have caused the outcome to a greater extent. In this paper, we argue that it is the *epistemic state* of an abnormally acting agent, rather than the abnormality of their action, that is driving people's causal judgments. Given the predictability of the normally acting agent's behaviour, the abnormal agent is in a better position to foresee the consequences of their action. We put this hypothesis to test in four experiments. In Experiment 1, we show that people judge the atypical agent as more causal than the normally acting agent, but also judge the atypical agent to have an epistemic advantage. In Experiment 2, we find that people do not judge a causal difference if no epistemic advantage for the abnormal agent arises. In Experiment 3, we replicate these findings in a scenario in which the abnormal agent's epistemic advantage generalises to a novel context. In Experiment 4, we extend these findings to mental states more broadly construed and develop a Bayesian network model that predicts the degree of outcome-oriented mental states based on action normality and epistemic states. We find that people infer mental states like desire and intention to a greater extent from abnormal behaviour if this behaviour is accompanied by an epistemic advantage. We discuss these results in light of current theories and research on people's preference for abnormal causes.

Keywords: causal judgment, normality, epistemic states, mental states

*Corresponding author: Lara Kirfel (ucjulki@ucl.ac.uk), 26 Bedford Way, London WC1H 0AP

"*The person who lit the match ought to have anticipated the presence of oxygen, whereas nobody is generally expected to pump all the oxygen out of the house in anticipation of a match-striking ceremony.*" (Pearl & Mackenzie, 2018, "The Book of Why", p. 291)

Introduction

Dust explosions, caused by the presence of combustible dust particles and a source of ignition, present a real danger in a variety of workplaces and industries. In 1983, damage of $1.2 million was caused by a fire in a furniture manufacturer in Walkertown, North Carolina. An employee who was carrying out repair works on the rooftop of the factory had dropped a cigarette in wood dust accumulated from the ventilation systems on the roof. Although both the wood dust and the lit cigarette were necessary for the fire to occur, the employee's smoking was reported as the main cause of the fire (U.S. Chemical Safety and Hazard Investigation Board, 2006). However, when a dust explosion occurs, the source of ignition is not automatically determined as the major cause. At an outdoor music festival in a Taiwanese fun-park in 2015, a festival-like colour powder was released over an area where party-goers were dancing and smoking, eventually leading to a dust explosion. The local fire department reported the spray of the combustible colour powder to have caused the incident ("Taiwan Formosa Water Park explosion injures hundreds", 2015, June 28).

Accident and incident reports give a unique insight into our judgments about actual causation. Which factors are determined to be crucial for an event and which are deemed peripheral shows how we attribute causality among a set of multiple causal factors. Why is the lighting of a cigarette determined as "the cause" for the dust explosion in the furniture factory, but not for the explosion in the water park? A prominent pattern found in causal cognition research is people's tendency to single out atypical or abnormal events as causes (Hart & Honoré, 1959/1985). And in fact, while it is fairly common to smoke at an outdoor festival, spreading combustible dust is rather atypical. In contrast, woodworking is a typical activity in a wood factory, while smoking, even in designated places, is comparatively rare. People's preference for atypical actions, objects or events as causes is a well-studied phenomenon in philosophy and psychology (Cheng & Novick, 1991; Hart & Honoré, 1959/1985; Hesslow, 1988; Hilton & Slugoski, 1986; Mackie, 1974), and might well explain why causal selection in the two dust explosion cases differs.

Aim of this paper

In this paper, we aim to provide an alternative account of people's increased causal attributions to 'abnormally' acting agents ("abnormal inflation" effect, Icard, Kominsky, and Knobe (2017)). In addition to acting atypically, both the smoking employee as well as the event organiser who instructed the colour powder release could have anticipated that the other would have acted 'typically'. In consequence, both could have foreseen the outcome of their action. We argue that it is the *epistemic state* of an abnormally acting agent, – what the agent foresees or expects – rather than the mere abnormality of their action, that is driving people's causal judgments in such cases. In four experiments, we show that the difference in causal judgements about atypical and typical actions is driven by the difference in agents' epistemic states. Furthermore, this epistemic asymmetry influences

people's inferences about the agents' mental states towards the outcome. These findings shed light on the crucial role of agentive epistemic states for causal attributions, and their role in people's preferences for atypical causes.

Abnormal Agents

How people perceive and judge atypical actions and events, often also termed "exceptional" (Kahneman & Miller, 1986) or (statistically) "abnormal" (Icard et al., 2017; Knobe, 2009), is of central interest to psychologists and philosophers. In particular, the atypicality an action has been shown to influence a variety of properties that people assign to the acting agent. In an experimental scenario by Kahneman and Miller (1986), Mr. Jones decides to give a stranger a ride and gets robbed as a consequence. People were more likely to attribute to Mr. Jones a feeling of regret over his action when Mr. Jones usually would not take hitch-hikers in his car, compared to when he frequently did so (Kahneman & Miller, 1986; Kutscher & Feldman, 2019; Miller, Turnbull, & McFarland, 1990; Roese, 1997). Similarly, in the case of negative personal consequences, an abnormally acting agent is perceived as more unlucky than an agent who acted as usual (Kutscher & Feldman, 2019), and as more deserving of compensation (Kutscher & Feldman, 2019; Miller & McFarland, 1986, see the former for a failure of replication). At the same time, people attribute more free will and free choice to an agent who deviates from a usual routine, and judge them more responsible for a harmful outcome (Fillon, Lantian, Feldman, & N'gbala, 2019; Miller & McFarland, 1986).

Of the various cognitive domains that have been shown to be influenced by normality, causal cognition is perhaps the most surprising one. Among a set of multiple causes, people systematically select the factor that is most abnormal as "the cause" for an outcome (Cheng & Novick, 1991; Hart & Honoré, 1959/1985), or rate this factor as having caused the outcome to a greater extent (Icard et al., 2017; Knobe, 2009; Kominsky, Phillips, Knobe, Gerstenberg, & Lagnado, 2014). Crucially, this causal preference prevails even when all factors are necessary for the outcome to occur, exemplified by the common practice to cite the lit match as cause of a fire, but neglect to mention the equally necessary oxygen in the air (Pearl, 2009). Consider the following example, adapted from Icard et al. (2017):

> A designer and a travel agent work in the same building. The building's climate control system is a new design that saves energy by keeping track of the number of people in the building, and only turns on when the designer AND the travel agent enter the building. The travel agent almost always arrives at work at 8:45am, but the designer almost always arrives at 10am. One day, the travel agent arrives at 8.45am, and, unexpectedly, the designer also arrives at 8.45am. As a result, the climate system turned on at 8:45am.
>
> (Icard et al., 2017, Vignette 'Building')

When asked about the extent to which each of the two people caused the climate control system to turn on at 8:45, Icard et al. (2017) found that participants agreed substantially more with the claim that the designer, the atypically acting agent, caused the outcome. Their study demonstrates how, despite equal causal contribution, the frequency

or "normality" with which an agent performs an action influences how causal they are perceived for an outcome (Icard et al., 2017).

Normality in Causal Cognition

People's tendency to attribute increased causality to abnormal rather than normal causes has been shown for atypical or unexpected events ("statistical" or "descriptive normality"), but also for events that violate social or moral norms ("prescriptive normality") (Henne, Niemi, Pinillos, De Brigard, & Knobe, 2019; Icard et al., 2017; Kirfel & Lagnado, 2018; Knobe, 2009; Kominsky, Phillips, Gerstenberg, Lagnado, & Knobe, 2015). The abnormality of a cause affects judgments about its causal strength, but also about the causal strength of other causal co-factors. If two causes jointly cause an outcome, people increase causal attribution to the abnormal causal factor, and at the same time reduce how causal they judge the other, normal causal factor ("causal supersession", Kominsky et al. (2015)). Norms or normality have also been shown to play a crucial role for causal judgements about omissions, i.e. events that did *not* occur (Henne, Niemi, et al., 2019; Henne, Pinillos, & De Brigard, 2017; McGrath, 2005; Sartorio, 2009; Willemsen, 2018; Willemsen & Reuter, 2016). From a wide range of events that did not occur, people select those as causes that violated norms or prior expectations (Henne, Niemi, et al., 2019; Willemsen, 2018; Willemsen & Reuter, 2016) Similarly, the common tendency to pick most recent events as "the cause" for an outcome is overridden if prior events in a chain of causes are perceived as more abnormal (Reuter, Kirfel, Van Riel, & Barlassina, 2014).

To date, there is a variety of competing accounts aiming to address people's preference for abnormal causes. Normality has been suggested to influence people's mental representation of alternative possibilities, or *counterfactual reasoning* (Epstude & Roese, 2008; Gerstenberg & Icard, 2020; Henne, Kulesza, Perez, & Houcek, 2020; Icard et al., 2017; Kahneman & Miller, 1986; Knobe, 2009; Roese & Epstude, 2017). According to counterfactual accounts, people show a general preference to mentally undo an abnormal action, and generate more alternatives to an agent deviating from typical behaviour. As a result, more counterfactual alternatives come to mind in which the outcome would have been absent as a result of a change in the abnormal agent's behaviour, emphasising this agent's perceived causal strength.

Another family of accounts argues that these effects are driven by moral judgments (Alicke & Rose, 2012). On one account, prescriptive norms have been claimed to influence causal judgments by normative evaluations such as the attribution of blame or responsibility (Alicke & Rose, 2012; Driver, 2008; Sytsma, Livengood, & Rose, 2012). Alternatively, the presence of norms in these kind of experimental scenarios has been suggested to shift participants' interpretation of the causal test question about agents into the realm of accountability (Samland & Waldmann, 2016).

Additional theories have been suggested with reference to co-variation of cause and effect (Cheng & Novick, 1991; Harinen, 2017) or general pragmatic principles of communication (Grice, 1989; Hilton & Jaspars, 1987; Kirfel, Icard, & Gerstenberg, 2020). In sum, there is general consensus that deviations from normality influence people's causal attributions, yet there is an ongoing debate about why these effects persist.

Revisiting Atypicality

The dust explosion cases as well the "building" vignette demonstrate our tendency to select abnormal or atypical actions as causes. As indicated earlier, however, upon closer examination it becomes clear that the agents differ in yet another aspect. Let's consider again the 'Building' vignette from (Icard et al., 2017). In comparison to the travel agent who presumably did not anticipate the designer's earlier arrival, the designer can expect the travel agent will arrive at 8.45am. Assuming that employees are familiar with the climate control set up, by deciding to arrive at 8.45am, the designer hence is likely to foresee the the climate control turning on, or more so than the travel agent. This basic epistemic difference between the agents might give rise to further inferences from abnormal behaviour: Did the designer want the climate control to turn on that day? Did they intend to turn it on?

The influence of abnormality on causal judgments about agents has been predominantly shown for causal structures in which two causes are necessary for the occurrence of the outcome (Icard et al., 2017; Knobe, 2009; Kominsky et al., 2015; Sytsma et al., 2012). In such cases of "conjunctive causation", the causal consequences of one agent's action is dependent on the action of another agent. This causal co-dependence also maps onto the agent's knowledge about the consequences of their action. Foreseeing the consequences of one's own action is to some extent dependent on knowing what the other agent does. The frequency of past behaviour is one of many social cues that are used for predicting the behaviour of others (Danner, Aarts, & de Vries, 2008; Epstein, 1979): How 'normal' agents act hence influences the predictability of their actions. This relationship between normality and predictability of actions applies to various kinds of normality or typicality that have been discussed in the literature: *token - or agent - level typical behaviour* ("Ben usually smokes at partys"), *type - or group - typical behaviour* ("Ben's friends usually smoke at parties."), but also the *typicality of features or properties* ("Wood dust is usually combustible."). In sum, for two agents (groups) in a conjunctive causal structure whose actions differ in any of these kinds of typicality, it follows that they will also differ in the extent of knowing what the other agent does, and therefore in knowing the consequences of their actions.

To know or not to know: The role of epistemic states

Epistemic states, what an agent thinks or believes, and mental states more broadly, what an agents wants, feels or desires, play a crucial role for moral judgements, but also influence attributions of causation more directly (Kinderman, Dunbar, & Bentall, 1998; Lagnado & Channon, 2008; Sytsma, 2019a). Intentional actions are rated as more causal than unintentional actions with accidental outcomes (Alicke, Rose, & Bloom, 2012; Gilbert, Tenney, Holland, & Spellman, 2015; Wiener & Pritchard, 1994; Williams & Lombrozo, 2010). Lagnado and Channon (2008) tested causal and blame attributions to agents in causal chains and found that both intentionality and foreseeabilty increase these attributions to the agent. People judge an agent causing a foreseeable outcome, both from the agent's perspective ("subjective foreseeability") as well as from an objective point of view ("objective foreseeability"), as more causal than if the outcome was unforeseeable.

Epistemic states influence causal judgements, and recent studies suggests they also mediate the effect of normality on causal judgments. The influence of prescriptive norm violation on causal judgments has been shown to hinge largely on the knowledge state of

the norm-violating agent (Samland, Josephs, Waldmann, & Rakoczy, 2016; Samland & Waldmann, 2015, 2016). An agent who unknowingly performs a forbidden action, e.g. by being unaware about the rule or norm that they are violating, is not judged more causal than someone who abides by the norm. Sytsma et al. (2012) show that typical, rather than atypical behaviour, is judged more causal if repeated behaviour increases the agent's ability to foresee or anticipate an outcome (Kirfel & Lagnado, 2017; Sytsma, 2020; Sytsma et al., 2012).

However, epistemic states not only mediate causal attributions to (atypical) behaviour; there is evidence that deviations from normal behaviour trigger inferences about a wider class of mental states of the agent (Alicke, 2000; Gerstenberg et al., 2018; Knobe, 2003; Monroe & Ysidron, 2021; Sytsma, 2019a). Research in attribution theory has traditionally argued that unexpected or odd behaviour is diagnostic of that agent having certain mental states, or dispositional and internal attributes (Jones, Davis, & Gergen, 1961; Jones & Harris, 1967; Kelley, 1967, 1973; Lucas, Griffiths, Xu, & Fawcett, 2009; Uttich & Lombrozo, 2010). People draw strong inferences about an individual's motives, intentions or character when their present action deviates from past behaviour (Heider & Simmel, 1944) or general expectations (Jones et al., 1961; Jones & Harris, 1967). Engaging in a prescriptively or statistically 'abnormal' behaviour receives higher attributions of free will (Clark, Baumeister, & Ditto, 2017; Clark et al., 2014), mediated by an inference about the agent's particularly strong personal desire for and choice of the abnormal action (Monroe & Ysidron, 2021).

Despite the often crucial role of epistemic states in causal attributions and inferences from abnormal behaviour, they are rarely controlled for experimentally. Studies on the effects of normality on causal judgments have predominantly used descriptive vignettes with human causal agents (but see Gerstenberg & Icard, 2020; Kirfel et al., 2020; Kirfel & Lagnado, 2019). The short verbal description of these causal scenarios often lack in-depth information about what the causal agents think, believe or know. As Sytsma (Sytsma, 2019a, p. 25) points out: "This [the lack of control for mental states] raises an important methodological issue for empirical work on ordinary causal attributions: researchers need to carefully consider and control for the inferences that participants might draw concerning the agents' mental states and motivations".

Hypotheses

In this paper, we aim to investigate what role agents' epistemic states play for causal judgments when the statistical normality of agents' actions varies. Previous research has explained the difference in causal judgements about an abnormal and normal agent in a conjunctive causal structure with reference to the difference in normality of behaviour (Hitchcock & Knobe, 2009; Icard et al., 2017; Knobe, 2009; Kominsky et al., 2015). Here, we explore the question whether the co-occurring epistemic asymmetry between abnormal and normal agent is the factor that influences people's causal attributions. In addition, we also aim to investigate what inferences people make about the causal agents' mental states from the normality of their behaviour.

Causal Judgements. Our main hypothesis derives from the close connection between normality and predictability of behaviour (Danner et al., 2008). In a conjunctive causal structure, expectations about the other agent's actions influence the relative foreseeability of the consequences of one's own action. An agent whose co-agent acts typically will

hence be in a better position to foresee whether their action will cause the outcome, compared to someone whose co-agent acts atypically. Comparing a normal and abnormal agent, the latter is provided with an epistemic advantage. At the most basic level, our hypothesis is that in a conjunctive causal structure, it is the abnormal agent's foreseeability of the outcome (via the foreseeability of their normal co-agent's actions) that leads to an increase in people's causal contributions to the abnormal agent.[1] Crucially, however, whether this epistemic advantage arises is dependent on whether the two agents know about each other and each others' action frequency. If agents do not know about how often the other one acts, no asymmetry in forseeability of the outcome arises. Our hypotheses with regards to causal judgments about intentional agents in a conjunctive causal structure are as follows:

i When the agents know about each other's actions, people will judge the abnormal agent as more causal for the outcome than the normal agent (*epistemic advantage*).

ii When the agents have no, or limited knowledge about each other's actions, people will not judge the abnormal agent as more causal for the outcome the normal agent (*no epistemic advantage*).

Outcome-Oriented Mental State Inference. In the last part of the paper, we return to the idea that the normality of behaviour not only influences causal judgments, but also gives rise to further inferences about the agents' mental states (Jones & Harris, 1967). We develop a Bayesian network model that allows us to predict the probability of an agent's mental state toward an outcome based on the normality of the agents' behaviour, as well as their epistemic states. With the help of a simplified example, we illustrate the model's prediction for a case in which the agents differ in the frequency of their actions and know/don't know about each other. We then test the qualitative predictions of the model for a case in which the agents know about each other, and a case in which the agents do not know about each other. In line with our previous hypotheses, we predict that people will infer outcome-oriented mental states to a greater degree from abnormal behaviour, but only when the abnormal agent has an epistemic advantage:

iii People infer outcome-oriented mental states to a greater degree from abnormal behaviour than from normal behaviour, but only if the agents know about each other.

We conducted four experiments to test these hypotheses. Experiment 1 will test hypothesis i). Experiment 2 will test hypothesis ii). Experiment 3 will test both hypotheses

[1]The example of the dust explosion cases suggests that the influence of epistemic states might go beyond occurrently entertained beliefs and expectations (Zimmerman, 1997). Even if party organiser and smoker did not currently or consciously expected their action to have a certain consequence, they could or should have (reasonably) been expected to do so (or more so than others). In this sense, non-actual *dispositional* epistemic states (FitzPatrick, 2017; Murray & Vargas, 2018; Sher, 2009), and perhaps even normative epistemic states (FitzPatrick, 2008) might have an equal impact on causal judgments. In this paper, our aim is to establish the influence of epistemic states in causal judgments about abnormal causal agents. As a test case, we will use actual and prevailing beliefs and expectations. We leave open the possibility of this assumed influence to expand to weaker or normative modes of epistemic states as well.

i) and ii) for a case in which the agent's expectations about each other generalise to a novel context. In Experiment 4, we put hypothesis iii) to test by assessing people's inferences about outcome-oriented mental states of agents who acted normally or abnormally. In sum, our experiments explore the influence of epistemic states for causal attributions to and inferences from abnormal behaviour. While the hypothesis we put forward here is neutral with regards to the exact mechanism by which epistemic states influence causal judgements, we will address this point by returning to some of the accounts of normality in causal cognition in the General Discussion.

Experiment 1: Abnormality and knowledge about each other

In the first experiment, we aimed to replicate previous research on the influence of atypicality on causal judgements. For this, we followed the general experimental paradigm involving two intentional agents in a conjunctive causal structure. The study was designed to explicitly control for participants' assumptions about agents' epistemic states. In Experiment 1, we aimed to test the effect of atypicality when both agents clearly know about each other and each other's actions.

Participants and Design. We recruited 159 participants via Amazon Mechanical Turk (∅ 80 participants per condition). Five participants were excluded for failing half of comprehension questions, i.e. five or more of the ten comprehension check questions in the entire study (see Appendix A), leaving a final sample size of $N = 154$ ($M_{age} = 33.28$, $SD_{age} = 9.72$, $N_{female} = 54$). The experiment has a 2 normality ("Both agents normal" vs. "Agent 1 normal, Agent 2 abnormal") × 2 agent (Agent 1: fixed agent vs. Agent 2: varied agent) × 2 scenario ("kettle" vs. "microwave") mixed design. The factors 'normality' and 'agent' were manipulated within participants, 'scenario' was manipulated between participants. All materials, data analyses, model code and power analyses can be found at `https://github.com/LaraKirfel/Atypicality`.

Material and Procedure

Participants completed both experimental conditions, i.e. the *'Both agents normal'* condition, and the *'Agent 1 normal, Agent 2 abnormal'* condition. It was randomised which condition participants completed first. Each of the two experimental conditions was structured in a similar manner. First, participants received a short introductory text about the scenario, together with a picture of the scene. The scenario involved two co-workers who share an office together (Figure 1). The agents were male with different names and looks across all conditions. Depending on the scenario type, the office either has two kettles or two microwaves that the employees can use whenever they want. For energy saving purposes, the company introduces the "Green Friday" on which the building is switched into a power-saving mode. As a result of this power-saving mode, the use of *more than one* kettle [microwave] in the office on Fridays will lead to a power cutout in the company, and agents are aware of this policy. This power-saving mode was introduced as a purely causal mechanism, with no particular prohibition or ban of using these items. Any inference about this mechanism as suggestive for a norm would still apply equally to all agents. The introduction text also stated explicitly that the agents share an office, know each other very well, and usually know what the other is doing during the day. The introduction was

Figure 1. **Illustration of two scenes from the video clips in Experiment**. In each scenario, two agents work together in a joint office. Depending on the scenario condition, the office has two kettles, or two microwaves that the agents can use. In the scenario pictured, only one agent uses the item ('Agent 1 normal, Agent 2 abnormal' condition).

followed by four comprehension questions, asking about the office situation (1), the agents' knowledge about each other (2), and the underlying causal structure (1) (see Appendix A).

Causal Structure. Studies on the role of norms in causal cognition have predominantly used vignette studies. In these vignettes, the described time frame of the causal event focuses on the narrow time point at which the singular causal outcome occurs. Prior causal history, e.g. whether the outcome has occurred before, is ambiguous. The co-variation between cause and effect has been shown to influence judgments of causal strength (Cheng & Novick, 1991; Harinen, 2017; Kirfel & Lagnado, 2018; Lagnado, Waldmann, Hagmayer, & Sloman, 2007). By introducing a restriction of kitchen devices on Fridays, we implemented a conjunctive causal structure that is temporally limited to a particular weekday. This kind of causal structure allows us to control for the frequency of outcome and for cause effect co-variation across the two normality conditions described below.

Normality. In order to manipulate action normality more naturally, we used animated video clips. After having read the introductory text, participants proceeded to watch a clip that shows a week in the office, from Monday to Friday (ca. 40s). In the *'Both agents normal'* condition, both agents Agent 1 and Agent 2 use the kettles (microwaves) from Monday to Thursday every day. In the *'Agent 1 normal, Agent 2 abnormal'* condition, only Agent 1 uses a kettle [microwave] from Monday to Friday, and Agent 2 only on Friday. In both conditions, both agents use the devices on Friday *at the same time* and as a result, the power stops. In our design, Agent 1 acts as the fixed agent: Agent 1's action is always 'normal' or typical, while Agent 2 acts as varied agents, with the normality of Agent 2's action varying across conditions.

Causal Question. At the end of the clip, participants were asked to what extent they agree with the following two causal rating questions about the agents on 7-point Likert

scales (1-'strongly disagree', 7-'strongly agree'): "Agent 1 [2] caused the power failure" (with one scale for each agent)[2], testing a graded notion of causality (Halpern & Hitchcock, 2015). The order of the questions was randomised across participants.

After completing the clip and the causal agreement questions, participants had to answer one more comprehension check question concerning the frequency of the agents' action in the clip. After having watched both clips for each normality condition and responded to the causal rating questions, they proceeded to a final question about the agents' epistemic states.

Expectation Question. In the final part of the experiment, participant had to rate their agreement about the agents' epistemic states in both clips. This question served as a control question to assess whether the difference in the agents' action frequency also corresponded to a difference in people's judgments about the agents' expectations about each other's actions. We used retrospective ratings of the agents' expectations about the other's actions as a proxy for how likely the agent was to foresee the consequences of their action. People were asked to rate their agreement with two statements for each of the two video clips they saw on 7-point Likert scale (1-'strongly disagree', 7-'strongly agree'): "Agent 1 (2) expected Agent 2 (1) to use the kettle (microwave) on Friday.". The order of the questions was randomized.

Participants completed the experiment by providing demographic information. On average, it took participants 11 minutes ($SD = 10.14$) to complete Experiment 1.

Results

Causal Rating. We analysed participants' agreement with the causal statements by comparing a series of linear mixed models using the `lme` package `lme` package in `R`, with participants as random effects.

The analysis revealed a significant main effect for agent, $\chi^2(1) = 39.99$, $p < .001$, $R_c^2 = .22$, normality $\chi^2(1) = 34.68$, $p < .001$, $R_c^2 = .28$, and a significant interaction of normality and agent $\chi^2(1) = 49.67$, $p < .001$, $R_c^2 = .35$.[3] Analysing the effect of agent for both normality conditions showed that people judge the agent who has not acted frequently before ($M = 5.56$, $SD = 1.68$, 95% CI [5.29, 5.82]) as more causal than the frequently acting agent ($M = 3.82$, $SD = 2.11$, 95% CI [3.48, 4.15]), $t(465) = 10.10$, $p < .001$ (see Figure 2). When both agents act frequently, there is no difference between Agent 1 ($M = 5.47$, $SD = 1.53$, 95% CI [5.23, 5.71]) and Agent 2 ($M = 5.45$, $SD = 1.54$, 95% CI [5.21, 5.69]), $t(465) = .23$, $p = .91$. The interaction of normality and agent was independent of the scenario type, $\chi^2(4) = .63$, $p = .96$.

[2]While the majority of studies in this area have focused on *intra-agent* comparisons (Icard et al., 2017; Kominsky et al., 2015), i.e. participants evaluate either fixed or varied agent only, our studies employ an *inter-agent* comparison design. We let participants rate both fixed [Agent 1] and varied [Agent 2] agent and compare the ratings of both agents in each experimental condition. See Sytsma (2019b) for a systematic review of effects of inter vs. intra-agent comparison contrasts.

[3]As effect size, we report the conditional R-squared for a full mixed linear effect model. R_c^2 provides the variance explained by a model including both fixed effects and random effects. R_c^2 values for mixed-effects models are calculated using the r.squaredGLMM function of the MuMIn package (Barton & Barton, 2015) that implements a method developed by Nakagawa and Schielzeth (2013).

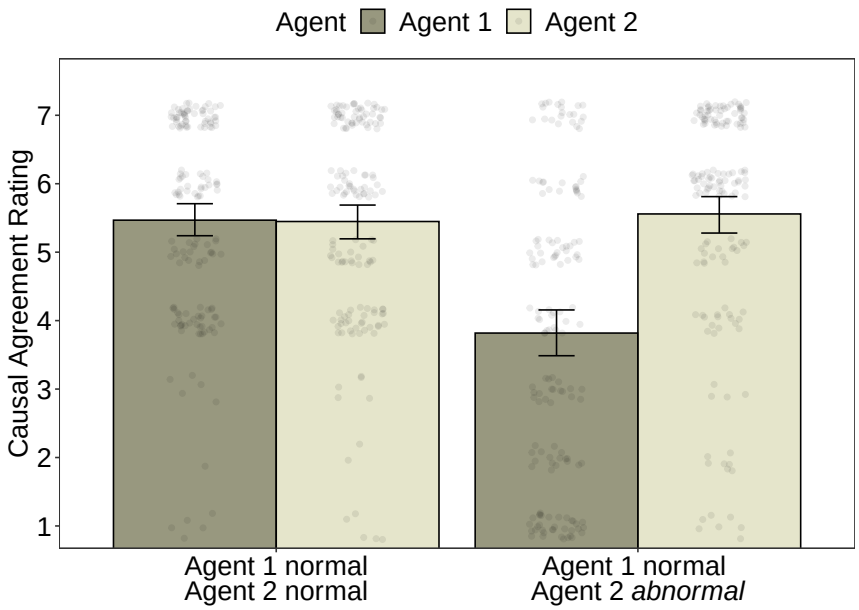

Figure 2. **Experiment 1: Causal Ratings**. Error bars depict 95% Confidence Intervals. Grey dots are individual participants' judgments jittered for visibility.

Expectation Rating. The analysis of expectation ratings revealed a significant main effect for agent $\chi^2(1) = 89$, $p < .001$, $R_c^2 = .25$, normality $\chi^2(1) = 94.55$, $p < .001$, $R_c^2 = .39$, and a significant interaction of normality and agent $\chi^2(1) = 101.85$, $p < .001$, $R_c^2 = .51$ (see Figure 3).

Analysing the effect of agent for both 'normality' conditions showed no significant difference between the varied ($M = 5.89$, $SD = 1.51$, 95% CI [5.64, 6.13]) and fixed agent ($M = 5.70$, $SD = 1.69$, 95% CI [5.43, 5.64]) when both act frequently $t(465) = 1.11$, $p < .01$, and crucially, a difference between the frequently ($M = 3.06$, $SD = 2.16$, 95% *CI* [2.72, 3.40]) and the infrequently acting agent ($M = 5.80$, $SD = 1.58$, 95% CI [5.55, 6.05]), $\chi^2(1) = 129.97$, $p < .001$. There was a significant interaction effect of the scenario type, $t(465) = 16.16$, $p < .001$, $R_c^2 = .02$. A decomposition of effects shows that the reduction in causal attribution to the frequently acting agent in the 'abnormal' condition is lower in the 'kettle' ($M = 2.56$, $SD = 1.91$, 95% CI [2.21, 3.10]) vs. 'microwave' scenario ($M = 3.43$, $SD = 2.30$, 95% CI [2.93, 3.93]), $\chi^2(4) = 6.21$, $p = .04$.

Discussion

Experiment 1 replicated the influence of action typicality on causal attributions to agents in a conjunctive causal structure. People judge a difference in the causality of two agents if these agents differ in how often they have performed the causal action. More precisely, the agent who has not performed this action before, i.e. who acts atypically or 'abnormally', is judged as more causal for the outcome. While previous literature has demonstrated an increase in causal attributions to the abnormal agent, we find in our experiment that people reduce their causal attributions to the normal agent in order to

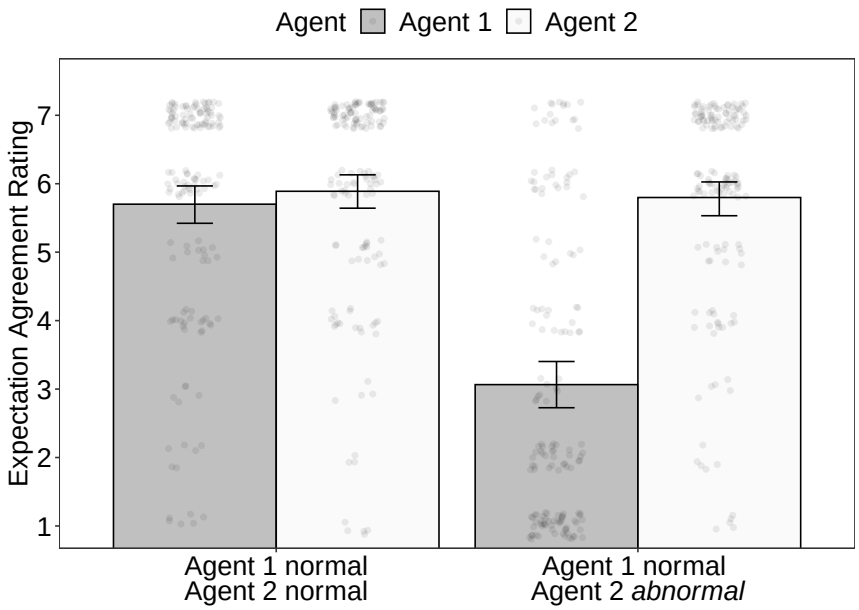

Figure 3. **Experiment 1: Expectation Ratings**. Error bars depict 95% Confidence Intervals. Grey dots are individual participants' judgments jittered for visibility.

express a difference in perceived causality between abnormal and normal agent (see "causal supersession" Kominsky et al., 2014). By using animated video clips, we were able to show this effect when the 'normality' or frequency of actions is manipulated in a sequential manner. Crucially, we found that the manipulation of action normality also has an impact on how people judge the agents' epistemic states. In hindsight, people judge that the abnormally acting agent expected their co-worker to act on Friday to a greater extent than vice-versa.

Experiment 1 confirmed the influence of abnormality on causal attributions, but also showed that this difference in causal judgments corresponds to a perceived difference in the agents' expectations about each other. This raises the question of what is driving people's causal perception of abnormal actions. Does statistical abnormality influence causal judgments, with epistemic states merely being a by-product, or is the epistemic state pivotal for people's causal attributions? The difference in action frequency corresponds to an epistemic advantage for the agent who usually does not engage in the respective action. At the most basic level, this epistemic advantage consists of foreseeing that one's action will cause the outcome. Experimental paradigms manipulating statistical normality hence often co-manipulate the agents' epistemic states. In the second experiment, we therefore wanted to test whether normality influences causal judgment when the agents do not know about each other's actions.

Experiment 2: Abnormality without knowledge about each other

In Experiment 2, we aimed to investigate the influence of normality of actions on people's causal judgments when the normality of actions does not change the agents' epistemic

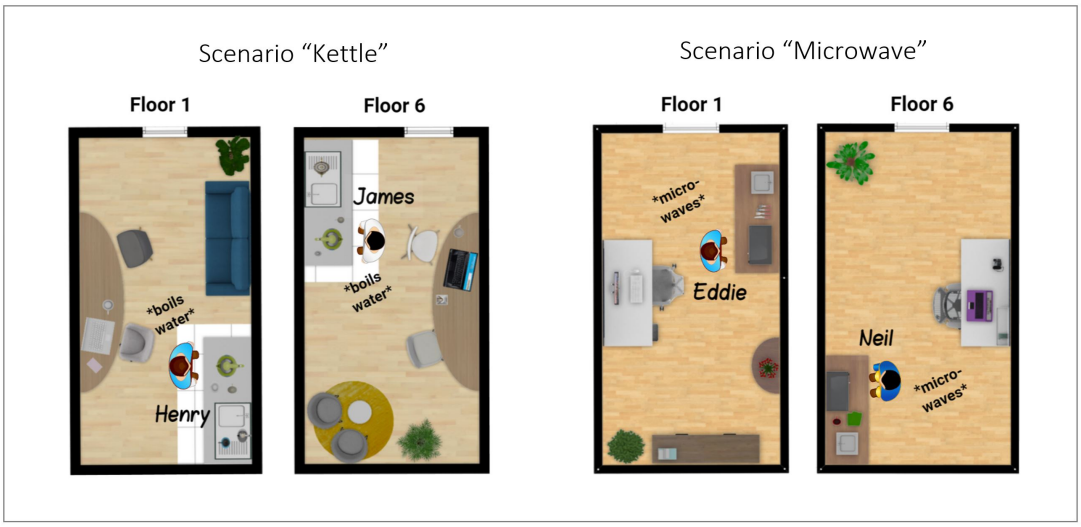

Figure 4. **Illustration of two scenes from the video clips in Experiment 2.** In each scenario, two agents work in separate offices on different floors. Depending on the scenario condition, each of the two offices has a kettle, or a microwave that the agents can use. In the scenario pictured, both agents use the item at the same time ('both agents normal' condition).

states.

Participants and Design

For Experiment 2, 149 participants were recruited via Amazon Mechanical Turk. 19 participants were excluded for failing five or more out of ten comprehension check questions (see Appendix B), leaving a final sample size of $N = 130$ ($M_{\text{age}} = 36.15$, $SD_{\text{age}} = 10.81$, $N_{\text{female}} = 43$).[4] As in Experiment 1, we adopted a 2 normality ("Both agents normal" vs. "Agent 1 normal, Agent 2 abnormal") × 2 agent (Agent 1: fixed agent vs. Agent 2: varied agent) × 2 scenario ("kettle" vs. "microwave") mixed design. The factors 'normality' and 'agent' were manipulated within participants, 'scenario' was manipulated between participants.

Material and Procedure

The material and procedure closely followed Experiment 1, with one crucial difference regarding the agents' knowledge about each other. The two co-workers work in separate offices on different floors, and the participants were informed that the agents do not know each other and have never met (Figure 4). As in Experiment 1, the company has introduced 'Green Friday' on which the use of more than one kettle [microwave] in the offices will lead

[4] We performed a power analysis using the 'SimR' package (Green & MacLeod, 2016) based on the effect size estimates from Experiment 1. Our Experiment 2 with $N = 130$ had an observed power of 1 CI [96.3; 100] to detect a significant interaction of normality × agent for both causal and expectation judgments at $p<0.05$.

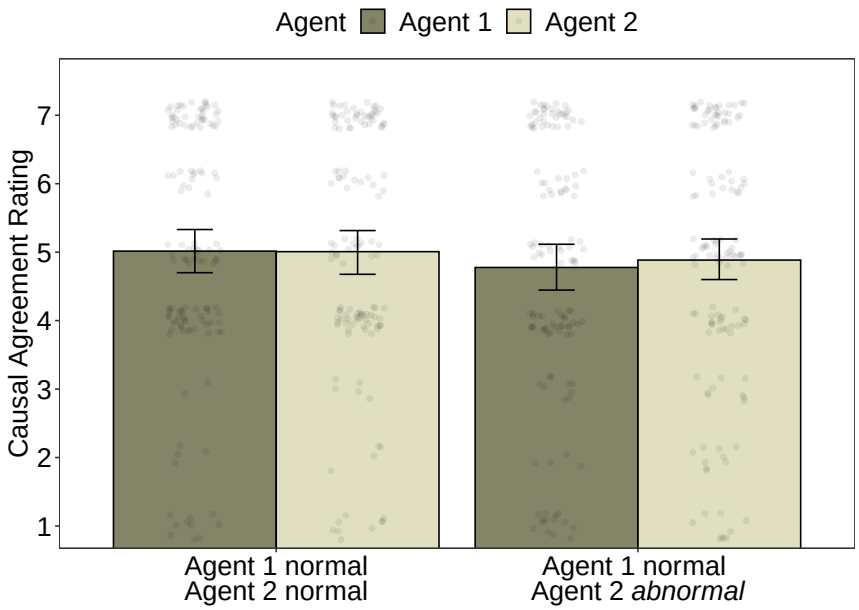

Figure 5. **Experiment 2: Causal Ratings**. Error bars depict 95% Confidence Intervals. Grey dots are individual participants' judgments jittered for visibility.

to a power cutout in the company. Frequency of actions was manipulated as in Experiment 1, with both agents being located in separate offices. On Friday, both agents make use of the kettle [microwave] and a power failure occurs.

Cause and Expectation Question. At the end of each clip, participants were asked to what extent they agree with the following two questions about Friday on 7-point Likert scale (1-'strongly disagree', 7-'strongly agree'): "Agent 1 (2) caused the power failure." The order of the questions was randomised across participants. At the end of the experiment, participants had to rate their agreement about the agents' epistemic states in both clips. We kept the epistemic rating question from Experiment 1 as a control question for our manipulation. People were asked to rate their agreement with two statements for each of the two video clip they saw on 7-point Likert scale (1-'strongly disagree', 7-'strongly agree'): "Agent 1 (2) expected Agent 2 (1) to use the kettle (microwave) on Friday." The order of the questions was randomised.

Results

Causal Rating. A Mixed Linear Model analysis on causal ratings revealed no significant main effect for agent, $\chi^2(1) = 0.21$, $p = .64$, normality $\chi^2(1) = 2.81$, $p = .09$, nor for the interaction between normality and agent $\chi^2(1) = .29$, $p = .59$. The infrequently agent ($M = 4.88$, $SD = 1.85$, 95% CI [4.57, 5.20]) is judged equally causal as the agent who has acted frequently before ($M = 4.78$, $SD = 1.93$, 95% CI [4.45, 5.11]) (see Figure 5). There was no significant interaction effect with scenario type, $\chi^2(4) = 7.65$, $p = .11$.

Expectation Rating. There was a significant main effect for agent $\chi^2(1) = 4.80$, $p = .03$, $R^2_c = .81$, normality $\chi^2(1) = 11.90$, $p < .001$, $R^2_c = .81$, and for the interaction

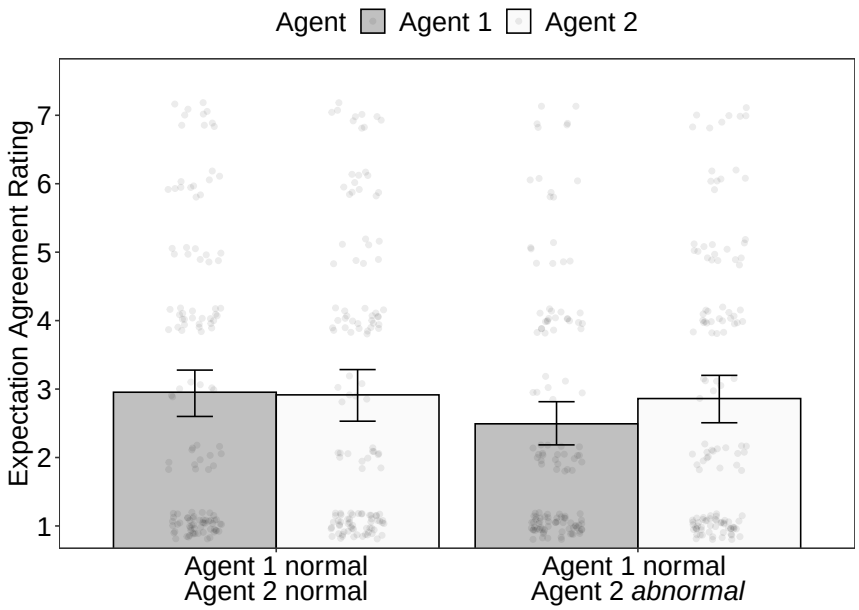

Figure 6. **Experiment 2: Expectation Ratings**. Error bars depict 95% Confidence Intervals. Grey dots are individual participants' judgments jittered for visibility.

between normality and agent $\chi^2(1) = 7.64$, $p < .001$, $R_c^2 = .82$.

While overall disagreeing that (M < 4) the agents expected each other's behaviour, people disagree slightly less with the statement that the frequently acting agent expected the infrequently acting co-worker to act ($M = 2.49$, $SD = 1.82$, 95% CI [2.18, 2.81]), than vice versa ($M = 2.86$, $SD = 1.96$, 95% CI [2.52, 3.20]), $t(393) = -3.54$, $p < .001$ (see Figure 6). We also found a significant three way interaction of normality, agent and scenario type $\chi^2(4) = 10.84$, $p = .02$, $R_c^2 = .83$. Decomposing the three-way interaction revealed that in the 'both agents normal' condition, participants disagree less with statements about the agents' expectations in the kettle scenario ($M = 3.27$, $SD = 2.11$, 95% CI [2.92, 3.63]) compared with the microwave scenario ($M = 2.56$, $SD = 1.96$, 95% CI [2.22, 2.91]), $\chi^2(1) = 4.02$, $p = .04$.

Discussion

Experiment 2 showed that the normality of actions does not influence people's causal judgements when the agents do not know about each other. An abnormally acting agent is judged equally causal for the outcome as a normally acting agent when both are unaware of each other's actions. Experiment 2 suggests that people do not consider the abnormality of an agent's action for causal judgments if the abnormal agent does not have an advantage in foreseeing the outcome. In general, causal ratings in Experiment 2 were slightly lower than in Experiment, but still above mid-point ($M > 4$), indicating that people generally do judge the agents to be causal. In light of the fact that participants still identified the agents as causal for the outcome, we take this as evidence that the lack of a difference between the normal and abnormal agent is not due to a lack of causal attribution in general. Rather,

participants do not perceive a causal difference between the two agents because there is no epistemic difference.

While participants overall tended to disagree with the statement that each agent expected their colleague to perform the causal action on Friday, there was a very small but significant difference between the agents in the abnormality condition. In general, retrospective evaluation of the agents' expectations allows people to attribute certain knowledge states to the agents in hindsight in order to make sense of their behaviour. One way to address this potential issue is to gather people's on-line judgements about the agents' knowledge about each other prior to the final outcome scenario. In Experiment 1 and 2, we investigated whether retrospective epistemic ratings correspond to the causal judgments that participants have given before. In Experiment 3, we wanted reverse the order of these two rating types, investigating whether people's prospective evaluation of the agents' epistemic states prior to the outcome corresponds to their causal judgments after the outcome has occurred. In addition, we were interested in the scope of influence of epistemic states. For Experiment 3, we wanted to test whether the asymmetry in expectations also influences causal judgements when the agents' expectations about each other need to generalise to a novel context.

Experiment 3: Do epistemic states influence causal judgments about agents in novel contexts?

In Experiment 3, we tested whether an asymmetry in agents' epistemic states influences people's causal judgments even if the agents' expectation about each other's actions have to be transferred to a novel context.

Participants and Design

We recruited 145 participants on Amazon Mechanical Turk. Three participants were excluded for failing five or more out of ten comprehension check questions (see Appendix C), leaving a final sample size of $N = 142$ ($M_{\text{age}} = 37.09$, $SD_{\text{age}} = 10.64$, $N_{\text{female}} = 58$). In Experiment 3, we adopted 2 knowledge (knowledge about each other vs. no knowledge about each other) $\times$ 2 agent (Agent 1: normal vs. Agent 2: abnormal) $\times$ 2 scenario ("coffee machine" vs. "microwave") mixed design. We replaced the kettle with a coffee machine. The factors 'knowledge' and 'agent' were manipulated within participants, 'scenario' was manipulated between participants.

Material and Procedure

The material of Experiment 3 closely matched the paradigm of Experiment 1 and 2. Participants read an introduction to the scenario, and completed a series of comprehension check questions. However, the scenario in Experiment 3 included a change in the office set up to allow for the manipulation of the agents' epistemic states (see Figure 7).

Normality & Knowledge. In Experiment 3, the agents need to move into two separate office spaces on different floors on Friday because their usual office is needed for meetings. We varied whether their usual office space from Mondays to Thursdays was a joint office, or whether they worked in separate offices. In the *knowledge condition*, the agents work together in one office from Monday to Thursday, and therefore know whether

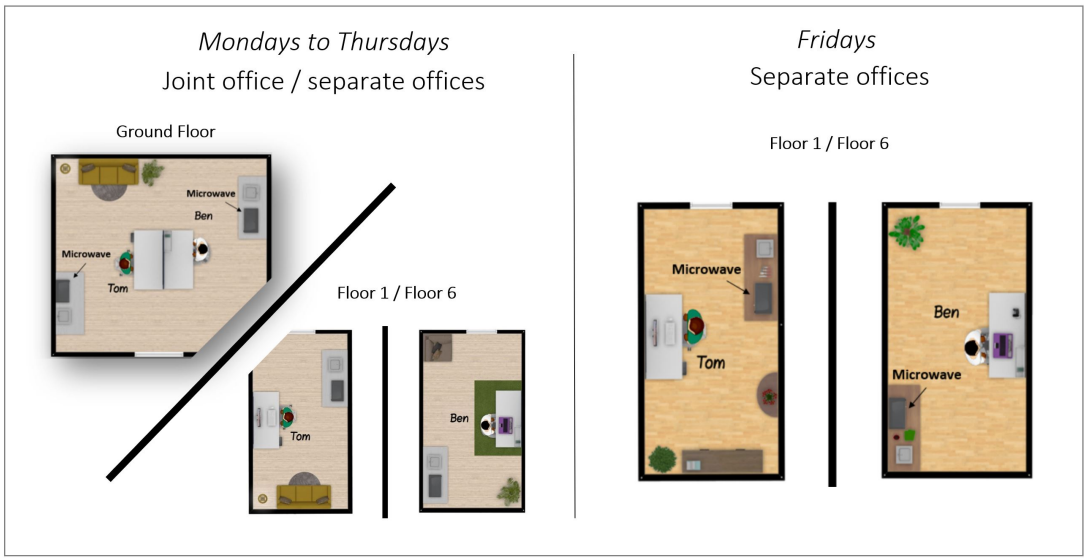

Figure 7. **Illustration of the two experimental conditions of the 'microwave' scenario in Experiment 3**. The two agents start off the week working either in a joint office ('knowledge condition') or in two separate offices on different floors ('no knowledge condition'). On Friday, in both knowledge conditions, the agents move into two different offices on separate floors.

the other agent is using the respective device each day. In the *no knowledge condition*, the agents work in different offices on separate floors, and do not know about each other. Agent 1 uses the device from Monday to Friday ("Normal Agent") and Agent 2 only uses the device on Friday ("Abnormal Agent").

Causal Structure. In line with the company policy, the use of two coffee machines [microwaves] on Fridays will lead to a power failure on Friday.

After having read the introduction and completed four comprehension check questions, participants proceeded to watch a first video clip. In this video clip, the weekdays from Monday to Thursday are shown, and on each of these day, only one of the two agents uses the coffee machine [microwave]. Depending on the knowledge condition, the agents are in a joint office space or in separate offices.

Expectation Rating. After having watched the first clip, participants are reminded that the agents need to move out of their current office and move into two separate offices the next day, Friday. They are then asked to rate the following statements about the agents' epistemic states on a 7-point Likert scale (1-'strongly disagree', 7-'strongly agree'): "Agent 1 (2) expects Agent 2 (1) to use the coffee machine (microwave) on Friday." As in the experiments before, the rating about what the agents expect about each other function as a proxy for how likely they think they might cause the outcome.

Participants then proceeded to watch a second video clip about the following Friday on which both agents have moved into their new offices. On this day, both of them use the coffee machine [microwave], and a power failure occurs.

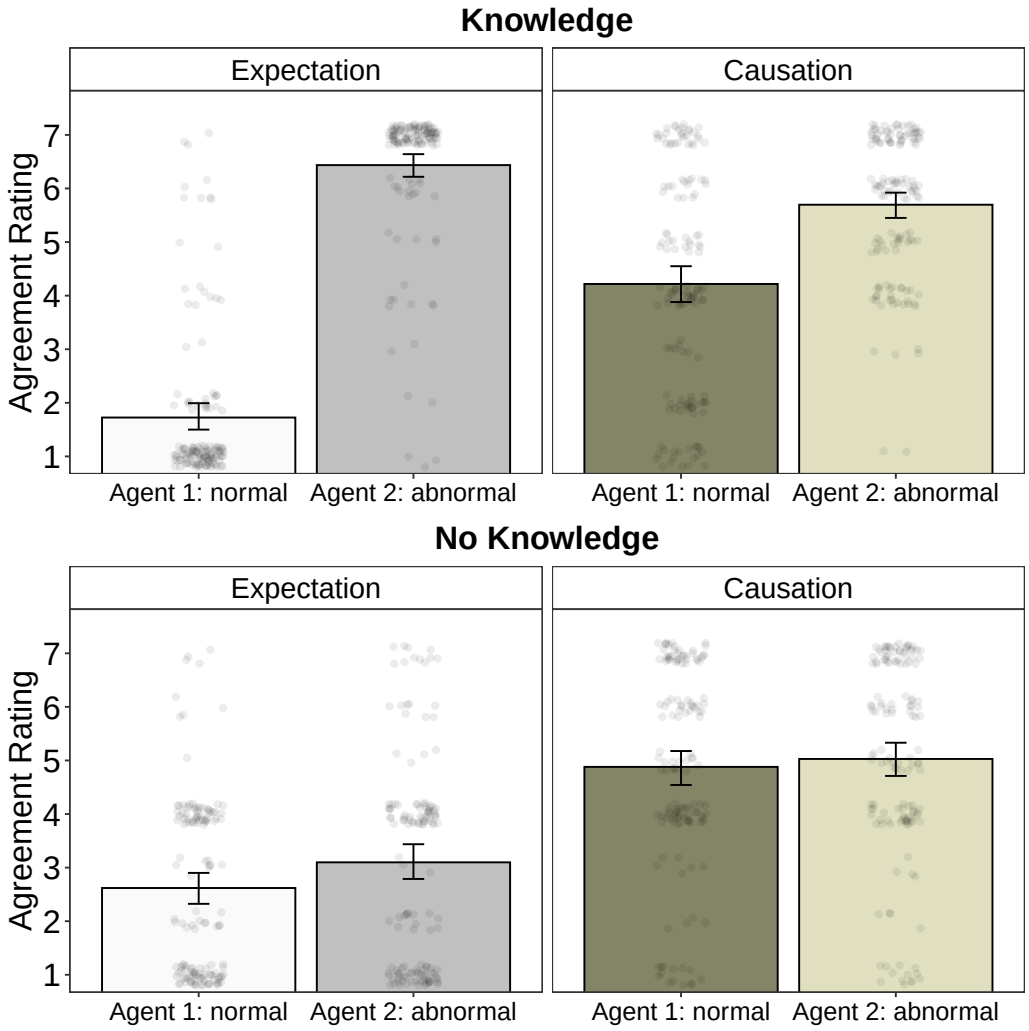

Figure 8. **Experiment 3: Expectation and Causal Ratings**. Agreement Ratings for the agents' epistemic states and causation of the outcome, depending on the 'knowledge' condition. Error bars depict 95% Confidence Intervals. Grey dots are individual participants' judgments jittered for visibility.

Causal Rating. Participants were then asked to what extent they agree with the following two questions about Friday on 7-point Likert scale (1-'strongly disagree', 7-'strongly agree'): "Agent 1 (2) caused the power failure." After having completed the causal judgment task, participants had to answer one more manipulation check question about the action frequency of both agents (see Appendix C).

Results

Expectation Ratings. A Mixed Linear Model analysis on expectation ratings revealed a significant main effect for the factor agent (Agent 1: normal vs. Agent 2: abnor-

mal), $\chi^2(1) = 196.1$, $p < .001$, $R_c^2 = .29$, knowledge $\chi^2(1) = 54.45$, $p < .001$, $R_c^2 = .36$, and a significant interaction for knowledge and agent $\chi^2(1) = 212.91$, $p < .001$, $R_c^2 = .60$.

When agents usually work in a joint office (Knowledge condition), participants judge the abnormal agent to expect the normal agent to act on Friday to a greater extent (M = 6.44, SD = 1.30, 95% CI [6.22, 6.51]) than vice versa (M = 1.73, SD = 1.49, 95% CI [1.48, 1.97]), $t(429) = 26.02$, $p < .001$ (see Figure 8). When both agents usually work in separate offices, agreement ratings are reduced and people agree slightly more about the prospective expectations of the abnormal agent (M = 3.10, SD = 1.94, 95% CI [2.78, 3.42]) compared to the normal agent (M = 2.62, SD = 1.66, 95% CI [2.35, 2.89]), $t(429) = 2.65$, $p < .01$. There was no interaction with scenario type, $t(429)$ = -27.03, $p = .35$.

Causal Ratings. The analysis of causal judgments revealed a significant main effect for the factor agent $\chi^2(1) = 44.84$, $p < .001$, $R_c^2 = .44$ and a significant interaction for knowledge and agent $\chi^2(1) = 32.98$, $p < .001$, $R_c^2 = .48$.

The abnormal agent is judged as more causal (M = 5.70, SD = 1.34, 95% CI [5.48, 5.92]) than the normally acting agent (M = 4.22, SD = 2.08, 95% CI [3.88, 4.56]) when both agents usually work in a joint office and know each other well, $t(429) = 9.16$, $p < .001$ (see Figure 8). In contrast, when the agents work in separate offices, people judge no causal difference between the abnormal (M = 5.03, SD = 1.87, 95% CI [4.72, 5.34]) and normal agent (M = 4.88, SD = 1.88, 95% CI [4.57, 5.19]) in causing the outcome on Friday, $t(429)$ = 0.92, $p = .23$. There was no interaction with scenario type, $\chi^2(4) = 5.86$, $p = .21$, $R_c^2 <$ .01.

Discussion

In Experiment 3, we gathered more evidence for our hypothesis that people's causal attributions to atypical actions are driven by the agents' epistemic states. The results show that in case of mutual knowledge about each others' habits, the abnormal agent is judged to expect a future action of the typically agent to a greater extent than vice versa. The abnormal agent is subsequently judged as more causal than the normal agent, even when the two causal actions occur in a context in which both agents need to predict the other's behaviour based on what they have learned about each other in the past. In contrast, if there is no such prior expectation, people do not perceive a causal difference between the abnormal and normal agent. Experiment 3 confirms our hypothesis in a paradigm in which the order of causal and epistemic ratings is reversed. Crucially, it shows that causal judgments can be influenced by agents' epistemic states in cases in which prior expectations generalise to novel contexts.

Mental state inference from (ab)normal behaviour

In Experiment 1 - 3, we show that causal judgments are influenced by the normality of the agents' actions, but only if the difference in action normality is paralleled by an epistemic advantage for the abnormal agent. A typical or statistically normal action is more predictable than an abnormal action. If the consequence of one's action is co-dependent on that of another person, the relative foreseeability of the outcome for an agent is higher when the second acts frequently, rather than rarely or atypically. At minimum, this enables

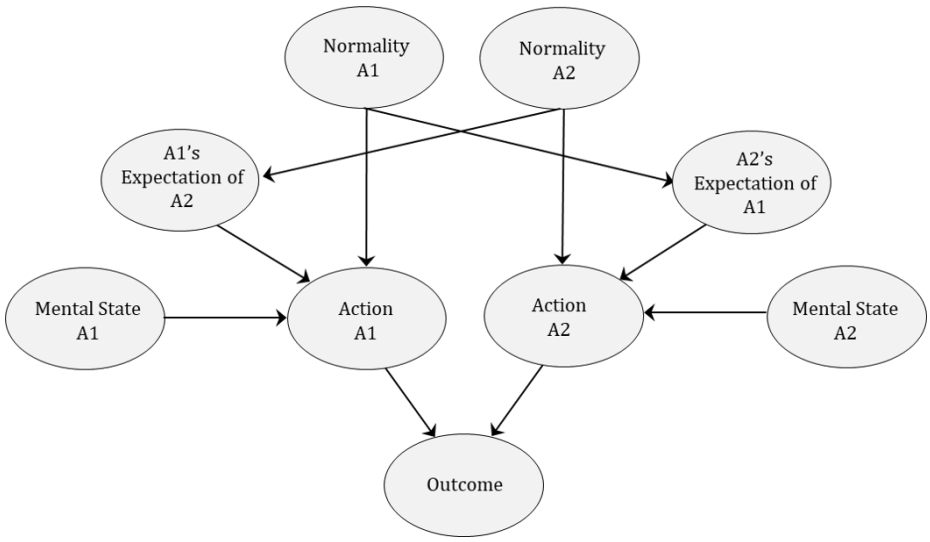

Figure 9. **Model 'Knowledge'**. Bayesian network Model for a conjunctive causal structure in which the outcome depends on the actions of two agents "A1" and "A2". Whether an agents acts depends on the normality of their action, their expectation about the other agent, and their mental state towards the outcome. In the "knowledge" condition, there is a link between the normality of an an agent's action is, and whether the other co-agent expects this agent's action.

the abnormal agent to foresee the consequences of their action to a greater extent than the normal agent.

While Experiments 1 - 3 demonstrated the agents' asymmetry in expectations about each other, we have yet to show that this asymmetry also translates into the predicted difference in foreseeability of the outcome. This was the first aim of Experiment 4. However, given that the abnormal agent acts despite being able to foresee a negative outcome, this raises further questions about the agent's mental state. More precisely, acting atypically evokes inferences about mental states that go beyond the agent's will or desire to use the microwave or coffee machine. We hypothesize that in such a scenario, people will make inferences from the (a)typicality of agents' overt behaviour about their "outcome"-oriented mental state, e.g. their intention or desire towards the outcome (Baker, Saxe, & Tenenbaum, 2011; Baker, 2012; Saxe & Houlihan, 2017). Given the asymmetry in epistemic states, we predict that an agent deviating from their usual routine of (non)-action will lead people to infer an increase in the degree of this agent's outcome-directed mental state, like a desire or intention to cause a power failure (Jones & Davis, 1965; Jones et al., 1961).

Bayesian Network Model of Mental State Inference

Our hypothesis can be formalised using a causal Bayesian network Model (see Figure 9). A Bayesian network (Pearl, 2009) is a formalism that uses a directed acyclic graph to represent the probabilistic dependencies between variables. The qualitative side of a Bayesian network is the graph structure, where a link from X to Y corresponds to the claim

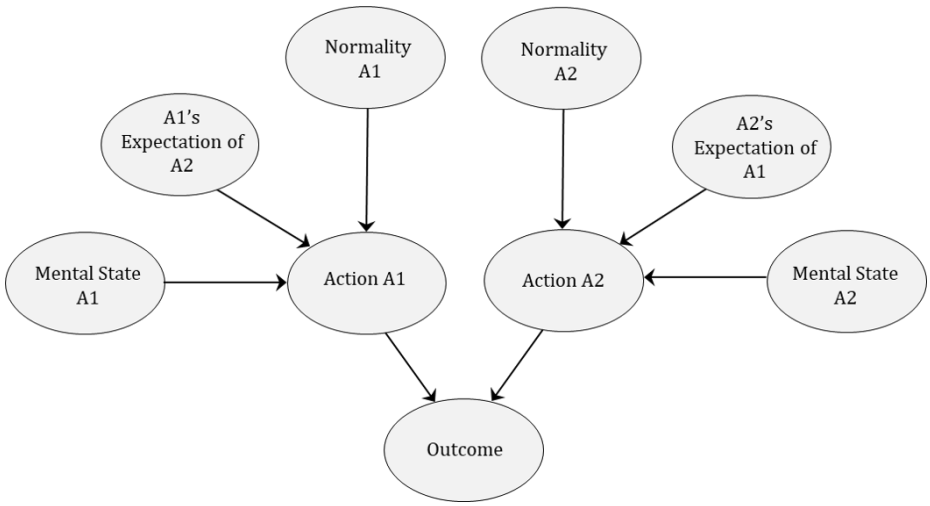

Figure 10. **Model 'No Knowledge'**. Bayesian network Model for a conjunctive causal structure in which the outcome depends on the actions of two agents "A1" and "A2". In the "No Knowledge" condition, there is no link between how normal an agent's action is, and whether the co-agent expects this agent's action

that Y depends on X. For the quantitative side, each variable has a conditional probability table (CPT) that specifies the probability of that variable given the possible values of its immediate causes (its parents in the graph). Variables with no parents are assigned prior probabilities.

When we acquire evidence about any of the variables in the model, we can use Bayes' rule to update the probabilities of the other variables. Bayesian networks are hence ideal for diagnostic inference: given some observed effect we can infer the probabilities of the possible causes of this effect. Here, we use Bayesian networks to model our scenarios, aiming to capture the key causes of the agents' actions, and how their probabilities should be updated given the manipulations in our experiments. Assuming that an agent's action is influenced by the normality of their actions, what they know about others and, crucially, their mental state, the model allows us to predict people's degree of inference about an agent's goal-directed mental state based on the other two factors. We develop this model for a scenario in which the two agents know about each other (see Figure 9), and a scenario in which neither agent knows about the other (see Figure 10).

Graph Structure. Central to our model is the assumption that an agent's action is influenced by three causal variables: how typical or "normal" their action is ("Normality"), what the agent expects other agents to do ("Expectation"), and finally, what kind of mental state the agent is in, that is, whether the agent intends or desires the outcome ("Outcome-Oriented Mental State"). These are represented as causal parents of the agent's action, and the same template is used for each agent, A1 and A2. In the "*Knowledge*" version of our model, whether A1 expects A2 to act is influenced by the normality of A2's behaviour (see Figure 9). In order to capture the assumption that normality of behaviour influences expectations, we add links from the normality of one agent's behaviour to the other agent's expectation about that behaviour. In the "*No Knowledge*" version of our model, the variable

Table 1

***Conditional Probability Tables for 'Knowledge' condition**: Conditional probability tables for the variables 'Normality', 'Mental State', 'Expectation about Agent 2' and 'Action' of agent A1. The probability tables are symmetrical for agents A1 and A2.*

(a) *'Normality'*

Normality A1	
true	0.8
false	0.2

(b) *'Mental State'*

Mental State A1	
true	0.1
false	0.9

(c) *'Expectation about A2'*

A1's Expectation about A2		
Normality of A2	*false*	*true*
false	0.9	0.1
true	0.1	0.9

(d) *'Action'*

Action of A1								
Mental State	*true*				*false*			
Normality of A1's Action	*true*		*false*		*true*		*false*	
Expectation about A2's Action	*true*	*false*	*true*	*false*	*true*	*false*	*true*	*false*
true	0.9	0.8	0.9	0.2	0.1	0.8	0.1	0.2
false	0.1	0.2	0.1	0.8	0.9	0.2	0.9	0.8

Table 2

***Conditional Probability Tables for 'No Knowledge' condition**: Conditional probability table for the variables 'Expectation about Agent 2' of agent A1. All other variables keep their values from the 'Knowledge' condition (see Table 1). The probability tables are symmetrical for agents A1 and A2.*

(a) *'A1's Expectation about A2'*

A1's Expectation about A2	
false	0.9
true	0.1

"A1's Expectation of A2" is independent of the normality of A2's behaviour (see Figure 10). Finally, whether the outcome in a conjunctive structure occurs depends on the action of

both agents A1 and A2.

Parametrising the model. We parameterise the network with illustrative probabilities. The conditional probabilities for each variable in this model serve as a rough approximation for our hypotheses, and can be flexibly adapted. For a start, we determine the probability for abnormal behaviour based on an agent acting on one out of five days, $P(A1\ abnormal) = 1/5 = 0.2$ (see Table 1a). Furthermore, in line with expectation ratings from Experiment 3, we assume that in the "Knowledge" condition, $A1$'s frequent, "normal" action will lead $A2$ to expect A1 to act with a likelihood of 90%, $P(A1\text{`}s\ Expectation\ about\ A2 \mid A2\ normal) = 0.9$ (see Table 1c). In the condition in which the agents do not know about each other ("*No Knowledge*"), we have set the prior probability of an agent knowing about the other to 10%, $P(A1\ Expectation\ about\ A2) = 0.1$. (see Table 2). In our study, the outcome in the conjunctive structure is a power failure. For an outcome with negative valence, we assume that the prior probability of having a certain mental state, e.g. a desire for bringing about a negative outcome, is low, $P(Mental\ State) = 0.1$ (see Table 1b) (see for example Chee & Murachver, 2012; Maselli & Altrocchi, 1969).

How does an agent's mental state with regards to a particular outcome influence their action? While there are various ways to model this influence, we make two basic stipulations. First, the probability of an agent acting who intends or desires the outcome and expects the other agent to act is high, i.e. $P(Action\ A1 \mid Mental\ State\ A1,\ A1\ expects\ A2) = 0.9$. (see Table 1d). Crucially, this is independent of how normal or abnormal the action is. Second, if the the agent intends the outcome but does not expect the other agent to act, we assume that the probability of this agent acting is simply how "normal" their action is, $P(Action\ A1 \mid Mental\ State\ A1, \neg\ A1\ expects\ A2) = P(Normality\ of\ A1)$.

Predictions

We now can compute the probability that an agent has an outcome-directed mental state given that both agents have acted in the 'knowledge' vs. 'no knowledge' condition using Bayesian updating. For our purposes, we will use a model that has set the variable "normality" to 'true' for $A1$, the normal agent, and 'false' for $A2$, the abnormal agent.

Knowledge. When both agents act but differ in how normal their actions are, our model predicts the normal agent to expect the abnormal agent's action with a probability of $P(A1\ expects\ A2 \mid A1\ normal,\ A2\ abnormal,\ Action\ A1, Action\ A2) = 0.89$, and the probablity of the abnormal agent expecting the normal agent's action as $P(A2\ expects\ A1 \mid A1\ normal, A2\ abnormal,\ Action\ A1, Action\ A2) = 0.02$ (see Figure 11 "Knowledge"). The probability of the normal agent having an outcome-directed mental state in such a scenario is $P(Mental\ State\ A1 \mid A1\ normal,\ A2\ abnormal,\ Action\ A1, Action\ A2) = 0.11$. In contrast, the probability for the abnormal agent to have an outcome-directed mental state is $P(Mental\ State\ A2 \mid A1\ normal, A2\ abnormal,\ Action\ A1, Action\ A2) = 0.46$.

No Knowledge. When neither agent knows about the other, our model predicts that the probability of the normal agent expecting the abnormal agent to act is 2%, and vice versa 9% (see Figure 11 "No Knowledge"). The probability for the abnormal agent having an outcome-directed mental state is 14%, and 11% for the normal agent .

Excursion: Positive and Neutral Outcomes. In contrast to the agent's expectations, the degree of predicted inference about an outcome-oriented mental states is to

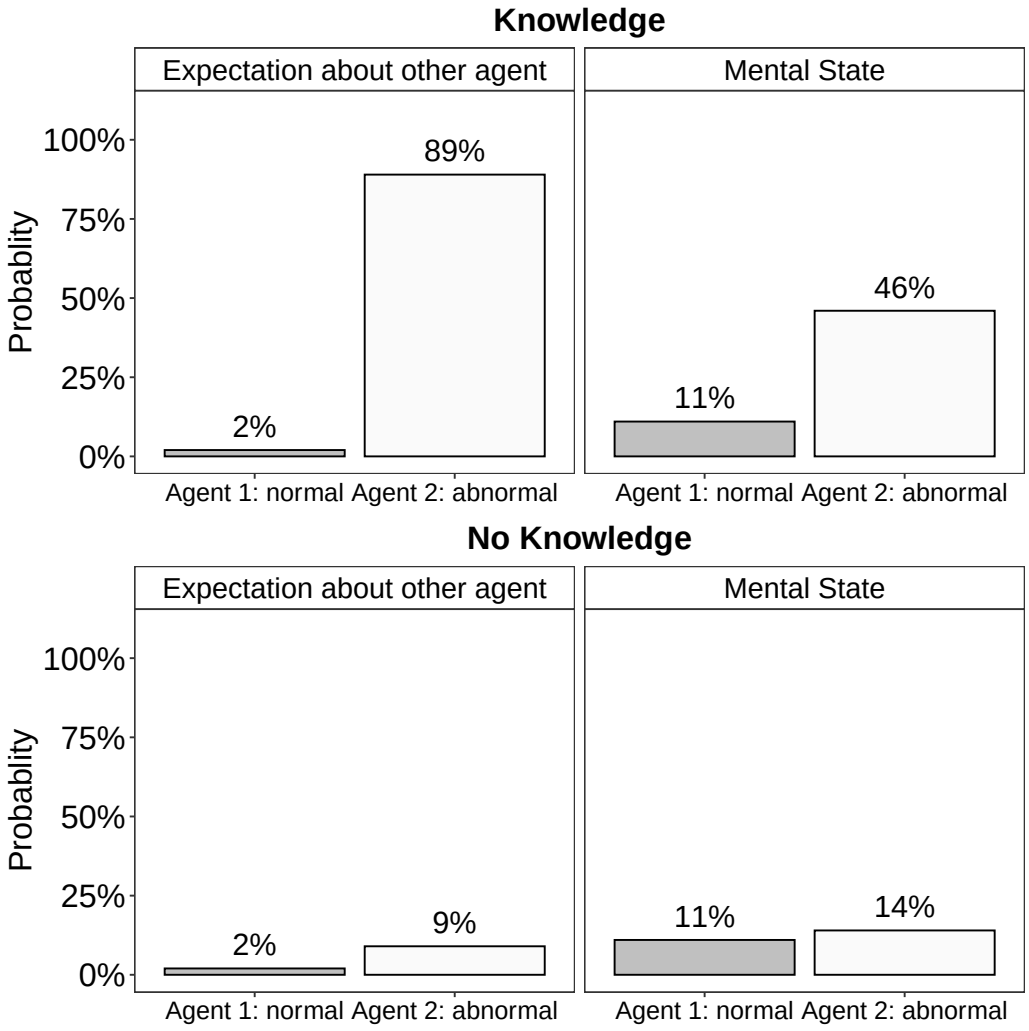

Figure 11. **Bayesian Model predictions in the "Knowledge" and "No Knowledge" scenario**. Bayesian model predictions for the probability of the normal [abnormal] agent i) expecting the abnormal [normal] to act, and ii) to have an outcome-directed mental state.

some extent dependent on their prior probability. If we would consider a case in which the outcome is positive, we would assume that the probability for having a mental state oriented at a positive outcome is high, e.g. $P(Mental\ State) = 0.9$. In such a case, our model predicts that the inferred probability of agents possessing this mental state is much higher, and the inferred difference smaller, 91% for the normal agent and 99% for the abnormal agent.

There is a lot of flexibility in how to specify the probabilities of each variable in this model, for example the assumed prior probability of having a mental state oriented at the outcome or how likely an agent with an outcome-oriented mental state is to act despite them usually not performing the causal action. Hence, exact values are not as important as comparative values. Crucially, the model in our example renders comparative differences

between the mental states of the normal and abnormal agent. Based on this model, we predict that people are more likely to infer an outcome-directed mental state from abnormal behaviour compared to normal behaviour in case of knowledge. When neither agent knows about the other, inference about an outcome-directed mental state for both abnormal and normal agent will be equally low.

Experiment 4: Inferences about Outcome-Oriented Mental States

In Experiment 4, we wanted to investigate people's inferences about agents' epistemic and outcome-oriented mental states based on the statistical normality of their actions and their knowledge about each other. First, we wanted to assess an intermediate step that has been tacitly assumed in our model, but not yet tested: that an asymmetry in knowledge about the other agent's action also corresponds to an asymmetry in expecting the outcome to occur as a result of one's action. At minimum, we expect people to attribute a higher expectation of the outcome to an abnormal compared to a normal agent when the agents know about each other, but not when they do not know about each other. Second, we wanted to put the qualitative predictions of the Bayesian network model about people's outcome-oriented mental state inferences to test.

Participants and Design

We recruited 163 participants on Amazon Mechanical Turk. One participant was excluded for failing five or more out of ten comprehension check questions (see Appendix D), leaving a final sample size of $N = 162$ ($M_{\text{age}} = 37.62$, $SD_{\text{age}} = 11.41$, $N_{\text{female}} = 71$)[5]. In Experiment 4, we adopted a 2 knowledge (knowledge about each other vs. no knowledge about each other) $\times$ 2 agent (Agent 1: normal vs. Agent 2: abnormal) $\times$ 2 scenario ("coffee machine" vs. "microwave") design. The factors 'knowledge' and 'agent' were manipulated within participants, 'scenario' was manipulated between participants.

Material and Procedure

In Experiment 4, we used the 'Agent 1: normal, Agent 2: abnormal' condition from Experiment 1 ("Agents know about each other") and Experiment 2 ("Agents don't know about each other"). As before, participants were introduced to the scenario, completed four manipulation check questions, and then proceeded to watch an animated video clip. In this clip, one agent frequently uses a coffee machine [microwave] from Monday to Thursday ("Agent 1: normal") while the other agent does not ("Agent 2: abnormal"). Both then cause a power failure by using both devices on Friday. Depending on the 'knowledge' condition, the agent work in one joint or two separate offices.

Epistemic Questions. After watching the whole video clip including the final day, participants were first asked to rate the agents' expectations about each other: "Agent 1 (2) expected Agent 2 (1) to use the coffee machine (microwave) on Friday." (7 - point Likert scale; 1 - 'strongly disagree', 7 - 'strongly agree'). The order of the two questions was randomised across participants. Participants then completed a second follow up question:

[5] Power analysis based on the effect size from Experiment 3 showed that Experiment 4 with $N = 162$ had a power of 1 CI [96.3; 100] to detect a significant interaction of agent $\times$ knowledge on expectation judgments at $p<0.05$.

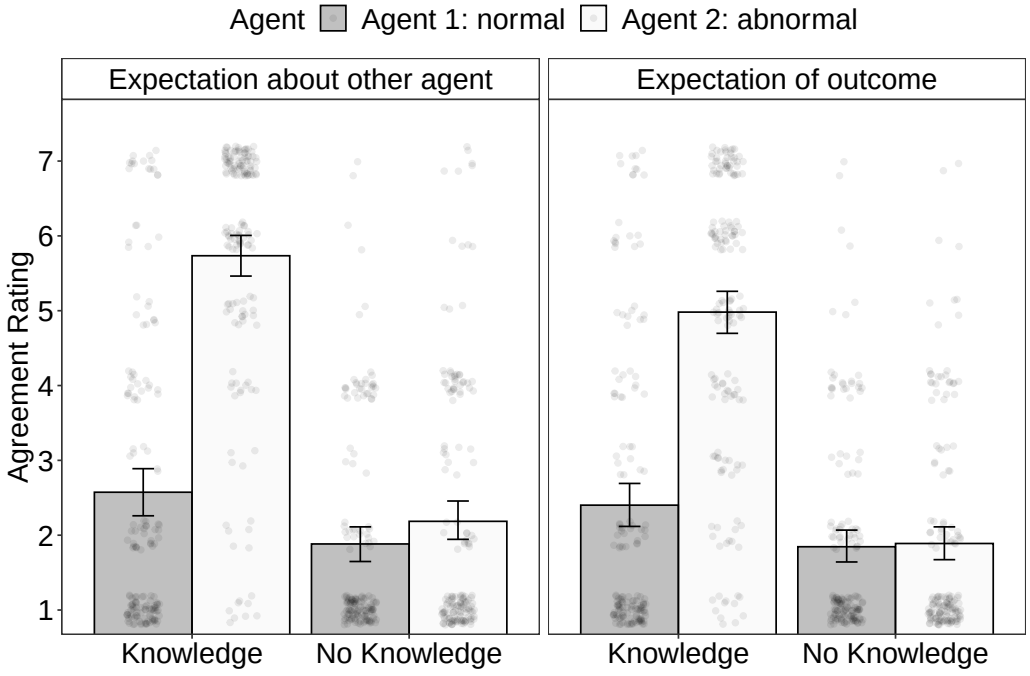

Figure 12. **Experiment 4: Epistemic State Ratings.** Ratings for the agent's expectation about i) the other agent and ii) the outcome are given with regards to the abnormal and normal agent. Error bars depict 95% Confidence Intervals. Red dots are individual participants' judgments jittered for visibility

"Given your answer to [question] (1), Agent 1 (2) _____ the outcome. ". The question was followed by a list of five outcome-oriented mental states, together with 7-point Likert scale (1 - 'strongly disagree', 7 - 'strongly agree') for each mental state: 1) "expected", 2) "did not mind", 3) "liked", 4) "desired", 5) "intended". Liking, desiring and intending qualify as being dispositional for an outcome-directed action, i.e. provide the mental condition for acting towards a goal or outcome (Brandstätter, Lengfelder, & Gollwitzer, 2001; Kuhlmeier, Wynn, & Bloom, 2003; Perugini & Bagozzi, 2001; Ryle, 2009). We included the mental state of 'being indifferent' because, while not necessarily being dispositional for an action, being indifferent towards the outcome does not prevent an agent from acting despite foreseeing it.

Participants had to rate their agreement with the insertion of each outcome-oriented mental state into the statement. This question was asked for both agents, and the order of each rating set for the agents was randomised across participants. After having answered all mental state rating questions, participant proceeded to the comprehension check question about the agents' action frequency in the video clip. Participants completed both the 'knowledge' and 'no knowledge' condition in randomized order.

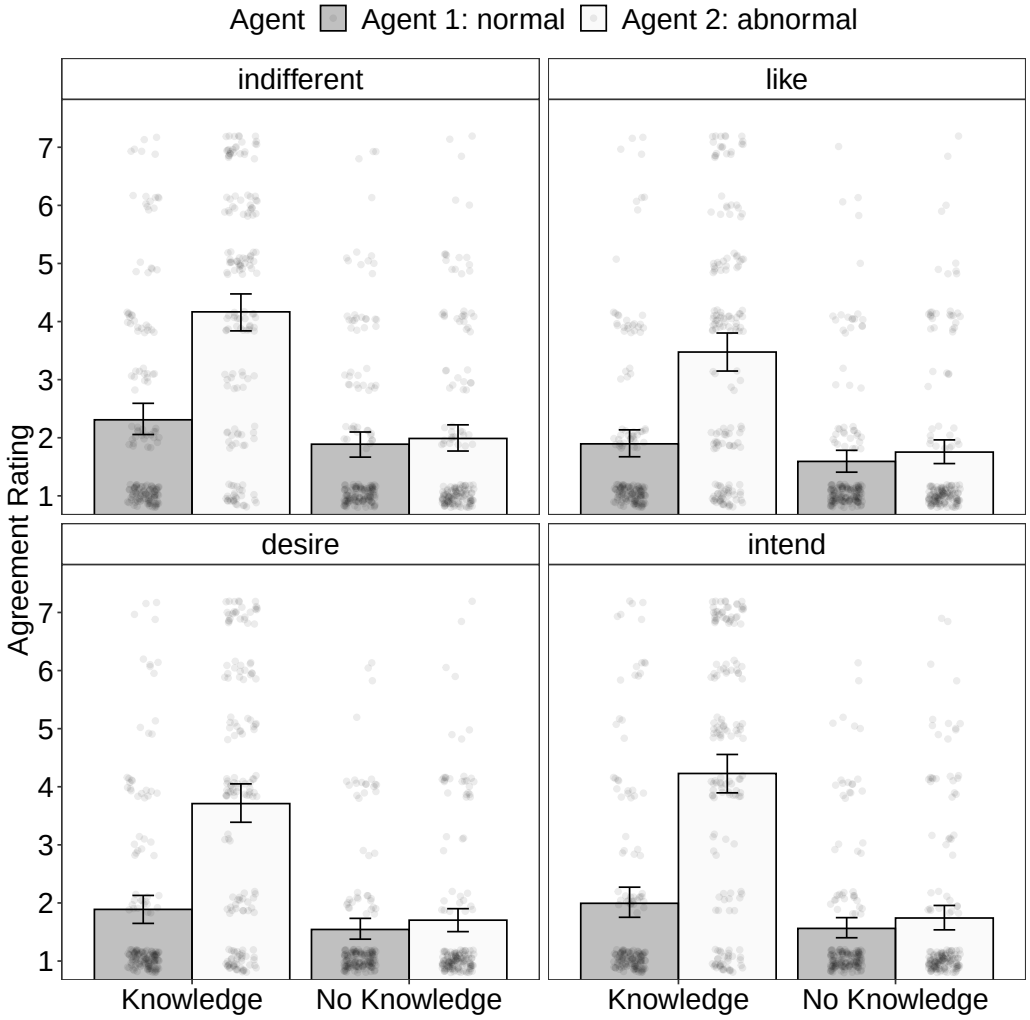

Figure 13. **Experiment 4: Outcome-Oriented Mental State Ratings**. Agreement ratings for four different mental states classes. Ratings for each outcome-oriented mental state are given with regards to the abnormal and normal agent. Error bars depict 95% Confidence Intervals. Grey dots are individual participants' judgments jittered for visibility.

Results

Expectation about each other. The analysis of agreement ratings with the statement "Agent 1 (2) expected Agent 2 (1) to use the coffee machine (microwave) on Friday." revealed a significant main effect for the factor agent $\chi^2(1) = 97.67, p < .001, R_c^2 = .14$, knowledge, $\chi^2(1) = 183.56, p < .001, R_c^2 = .38$, and a significant interaction for knowledge and agent $\chi^2(1) = 122, p < .001, R_c^2 = .51$.

The abnormal agent is judged to expect the other agent to act to a greater extent ($M = 5.73$, $SD = 1.74$, 95% CI [5.46, 6.00], $t(489) = 17.49, p < .001$, than the normally acting agent ($M = 2.57$, $SD = 2.02$, 95% CI [2.26, 2.89]) when both agents share an office,

compared to when abnormal ($M = 2.19$, $SD = 1.69$. 95% CI [1.92, 2.45]) and normal agent do not know each other ($M = 1.88$, $SD = 1.41$, 95% CI [1.67, 2.10]), $\chi^2(1) = 9.70, p < .01$, (see Figure 12). There was no interaction with scenario type $\chi^2(4) = 2.11, p = .71$.

Expectation about the outcome. The analysis of the agreement ratings for expectation of the outcome revealed a significant main affect for agent, $\chi^2(1) = 65.92$, $p < .001$, $R^2_c = .09$, knowledge, $\chi^2(1) = 155.82$, $p < .001$, $R^2_c = .34$, and a significant interaction between agent and knowledge $\chi^2(1) = 98.08$, $p < .001$, $R^2_c = .46$.

When both agents know about each other, the abnormal agent was judged to expect the outcome to a greater extent ($M = 4.98$, $SD = 1.90$, 95% CI [4.69, 5.27]) than the normal agent ($M = 2.40$, $SD = 1.95$, 95% CI [2.10, 2.70]), $t(489) = 14.95$, $p < .001$. There is no difference when the agents do not know about each other, $t(489) = -.25$, $p = .80$, (Abnormal agent: $M = 1.89$, $SD = 1.40$, 95% CI [1.67, 2.10]; Normal agent: $M = 1.85$, $SD = 1.34$, 95% CI [1.64, 2.05]) (see Figure 12). There was a significant interaction with scenario type $\chi^2(4) = 11.80$, $p = .02$, $R^2_c = .47$. Ratings for the abnormal agent in the 'no knowledge' were higher in the coffee scenario ($M = 2.26$, $SD = 1.69$, 95% CI [0.19, 1.69]) than in the microwave scenario, ($M = 1.55$, $SD = 0.97$, 95% CI [0.11, 0.96]), $t(119) = 3.22$, $p < .01$.

Outcome-Oriented Mental State Attribution. We aggregated the four mental state classes 'being indifferent', 'liking', 'desiring' and 'intending' into one measure of outcome-oriented mental state attribution. The analysis of the agreement ratings for a mental state towards the outcome revealed a significant main affect for agent, $\chi^2(1) = 66.63$, $p < .001$, $R^2_c = .25$, knowledge, $\chi^2(1) = 119.97$, $p < .001$, $R^2_c = .42$, and a significant interaction between agent and knowledge $\chi^2(1) = 71.01$, $p < .001$, $R^2_c = .50$.

When the agents know about each other, the abnormal agent was judged to posses a greater degree of the outcome-oriented mental state ($M = 3.90$, $SD = 1.97$, 95% CI [3.59, 4.20]), than the normal agent ($M = 2.02$, $SD = 1.51$, 95% CI [1.79, 2.25]), $t(489) = -13.40$, $p < .001$. People judge no difference when the agents do not know about each other, $t(489) = -1.07$, $p = .28$, with the abnormal agent judged to have this mental state to the same extent ($M = 1.80$, $SD = 1.33$, 95% CI [1.59, 2.00]) as the normal agent ($M = 1.64$, $SD = 1.14$, 95% CI [1.47, 1.82]) (see Figure 13). There was no interaction with scenario type, $\chi^2(4) = 6.89$, $p = .14$.

Discussion

Experiment 4 confirmed the qualitative predictions about outcome-oriented mental state inference from our Bayesian network model. First, in line with our previous studies, people judged the abnormal agent to expect the normal agent to act to a greater extent than vice versa, but only in the condition in which both agents know about each other. In addition, people also judged the abnormal agent to expect the outcome to a greater extent than the normal agent in the 'knowledge' condition, but not in the 'no knowledge' condition. Overall, participants inferred four types of outcome-oriented mental states – indifference, liking, desire and intention – to a greater extent from abnormal behaviour than from normal behaviour. However, this only mainly the case when both agents knew about each other.

The fact that epistemic state asymmetry also generates an asymmetry in inferences about outcome-oriented mental states raises the question whether causal attributions are partly driven by what people assume about the agents' intentions and desires, rather than

just by the difference in foreseeability. In our studies, we have shown that "normality" or action typicality influences causal judgments by creating an epistemic asymmetry between two agents who act with different frequency. The argument we make here is compatible with this influence being mediated by further inferences about mental states. However, as we have speculated above, inferences about mental states can vary quite substantially depending on the valence of the outcome, while expectations of the outcome likely remain unaffected by this factor. Future studies will need to test the exact parameters in this model and their influence on causal attributions more rigorously.

General Discussion

The phenomenon that people systematically choose atypical and abnormal actions, agents, and objects as causes has been the subject of debate in philosophy as well as psychology (cf. Hart & Honoré, 1959/1985). Our preference for abnormal causes manifests in causal explanations and language (Gerstenberg & Icard, 2020; Hilton & Slugoski, 1986), causal judgments (Icard & Knobe, 2016; Knobe & Fraser, 2008; Kominsky et al., 2015), causal intervention (Cheng & Novick, 1991) and prospective causal thinking (Henne, O'Neill, Bello, Khemlani, & De Brigard, 2021). Several theories have been proposed to explain why we have a tendency to prefer atypical and abnormal factors as causes (Alicke et al., 2012; Cheng & Novick, 1991; Hilton & Jaspars, 1987; Hitchcock & Knobe, 2009; Icard et al., 2017; Samland & Waldmann, 2016; Woodward, 2001).

In this paper, we did not aim to settle this debate. Rather, we have argued that, in terms of causal agents, an important and often overlooked factor driving people's causal attributions is the agents' epistemic states. In four experiments, we have shown that the tendency to judge an abnormally acting agent as more causal than a normally acting agent is influenced by the epistemic asymmetry between agents. People do not perceive a causal difference between an abnormal and normal agent if no epistemic asymmetry arises. In addition, we find that people are infer to a greater degree of an outcome-oriented mental state from an abnormal action, but again, only if this abnormal behaviour is accompanied by an epistemic advantage for the abnormal agent.

What are the implications of our findings? First, we discuss our results in the context of the current debate about norms in causal cognition. We will examine our findings in the light of prominent accounts explaining the influence of norms in causal cognition (Alicke & Rose, 2012; Alicke et al., 2012; Henne, Niemi, et al., 2019; Icard et al., 2017; Knobe, 2009; Kominsky & Phillips, 2019; Samland et al., 2016). Second, we will discuss how our findings relate to other research on the influence of abnormality on causal judgments, and by doing so, outline the limitations of our study and point out future directions of research. In particular, we will discuss research on atypically acting single agents (Kahneman & Miller, 1986), abnormal objects (Kominsky & Phillips, 2019) and a relatively novel finding in causal cognition research, the preference for normal causes in disjunctive structures (Gerstenberg & Icard, 2020; Icard et al., 2017).

Counterfactuals, Blame and Pragmatics

The influence of norms, or normality, on causal judgments has been the subject of recent debate in psychology and philosophy. Studies in this area show that agents who

violate prescriptive or statistical norms receive higher causal attributions than norm-abiding agents (Icard & Knobe, 2016; Kahneman & Miller, 1986; Knobe, 2009; Kominsky & Phillips, 2019; Kominsky et al., 2015). Several accounts have been suggested to explain this pattern.

Counterfactuals. The *counterfactual reasoning* account (Hitchcock & Knobe, 2009; Icard et al., 2017; Kahneman & Miller, 1986; Knobe, 2009; Kominsky & Phillips, 2019; Kominsky et al., 2015) aims to explain the influence of norms and normality by reference to people's thinking about alternative possibilities. Normality is assumed to make certain counterfactual possibilities more relevant than others. A norm-violating causal action is mentally replaced by a norm-conforming action, hence highlighting the counterfactual dependence of the outcome on that particular action (Hitchcock & Knobe, 2009; Knobe, 2009; Kominsky et al., 2015; Phillips, Luguri, & Knobe, 2015). Recent developments of this account have suggested an expectation-based account of normality (Kominsky & Phillips, 2019). According to this account, an action is only perceived as abnormal if the acting agent can expect their behaviour to be norm violating. This extension aims to explain why an agent who unknowingly violates a rule is not judged more causal than a norm-abiding agent (Samland & Waldmann, 2015, 2016).

How do our results perform in light of an expectation-based normality account? In terms of *group-level statistical normality* (Sytsma et al., 2012), an agent who acts in ignorance of what others do might not be aware which actions are frequently or typically performed, and is hence unaware of the atypicality of their own behaviour. In this sense, a counterfactual account would predict no causal difference between an abnormal and normal agent in the 'no knowledge' condition of our experiments. In terms of *agent-level statistical normality*, an agent can generally be expected to know when they deviate from their previous behaviour, independent of their knowledge about what others do. Whether an expectation-based account can explain the results in this paper depends largely on how knowledge about statistical normality is conceptualised.

Following up on the idea by Kominsky and Phillips (2019), we think there is potential for counterfactual accounts to incorporate epistemic states more generally (Kirfel & Lagnado, 2017, 2018, 2019). Adopting a *theory of mind* account of counterfactual reasoning, people's simulation of counterfactual alternatives to an agent's action might depend on the agent's knowledge and beliefs broadly construed. An agent's ignorance about relevant features of their action might influence to what extent people consider alternative situations in which the agent does not perform this action. If the causal agent does not know, or could not have reasonably been assumed to know that they will cause harm or violate a rule by a certain action, people might not consider (or be less likely to sample) counterfactuals in which this action is undone. In this respect, however, people's counterfactual reasoning (Icard et al., 2017) or simulation of possibilities (Johnson-Laird, Khemlani, & Goodwin, 2015; Khemlani, Byrne, & Johnson-Laird, 2018) is *not* directly influenced by the statistical normality of actions. Instead, normality of action influences causal judgements by changing agents' epistemic states about (the consequences of) their action. Future research will need to test the influence of agents' epistemic states on counterfactual reasoning more directly.

Blame and Responsibility. A different class of theories aims to explain the role of prescriptive normality in causal cognition in terms of moral judgements. According to the *Culpable Control Model*, people's causal judgments are biased towards a desire to assign blame to the abnormal factor (Alicke, 2000; Alicke & Rose, 2012). Increasing the perceived

causal contribution to a norm-violating causal agent allows people to validate a spontaneous blame response. Sytsma et al. (2012) argue that people's ordinary concept of causation is itself normatively enriched, and that it is used similarly to the concept of responsibility (Sytsma, 2019a; Sytsma & Livengood, 2019; Sytsma et al., 2012). Samland et al. (2016) suggest that norms shift people's pragmatic understanding of the verbal cause concept into the moral domain. In the context of norms and norm-violating agents, participants interpret the causal test question as a request to assign accountability (Samland & Waldmann, 2014, 2015, 2016).

The fact that action typicality influences causal judgments via epistemic states would allow blame-oriented accounts of causal judgments to extend their predictions to descriptive norms like action typicality. An agent is seen as more blameworthy, more responsible, or is held more accountable for a negative outcome if they could have foreseen the outcome (more) qua the typicality of their and other agents' actions. Sytsma (2019a) finds that an agent's character, such a being negligent, matters for causal attributions when this character trait warrants an inference about the agent's epistemic state about the outcome. In our studies, we show that people attribute high expectation of the outcome to the abnormal agent, yet are hesitant to attribute to them particularly strong intentions or desires for a negative outcome. In addition, according to our model, the degree of inferences about mental states like intentions and desires is sensitive to the nature of outcome, and might vary according to the perceived prior probability of an agent intending a good, bad, or neutral outcome. Studies, however, show a somewhat consistent difference in causal attributions to normal and abnormal causes for outcomes of different valence (Icard et al., 2017; Kominsky et al., 2015). In general, some blame-oriented accounts do not provide a fully fledged explanation of why people would attribute causality to an abnormal agent who causes a neutral or positive outcome (Icard et al., 2017). At the current stage, blame-oriented accounts are need to specify to what extent they take each of the epistemic and mental state components to influence attributions of blame, accountability or responsibility.

The findings from our experiments are in principle compatible with both accounts, and the studies in this paper do not speak in clear favour for either of the two theoretical lines. The role of epistemic states for causal attributions provides interesting new challenges for both lines of research – accounts that assume counterfactual reasoning, as well as those that stipulate an underlying 'normative' judgment. An agent's epistemic state might influence reasoning about alternatives to their action, but they are also a crucial factor in blame and moral judgments. Further studies on how epistemic states determine causal attributions with the addition of response measures assessing blame and perceived norm-violations might help decide between these two accounts.

Single actions

The experiments in this paper have exclusively focused on causal attributions to two agents in a conjunctive causal structure. As demonstrated by the 'hitchhiker case' (Kahneman & Miller, 1986), people are also more prone to generate counterfactual alternatives and attribute causality if a negative outcome results from a single agent performing an atypical vs. typical action (Fillon, Kutscher, & Feldman, 2020; Fillon et al., 2019; Hilton & Slugoski, 1986; Kahneman & Miller, 1986; Monroe & Ysidron, 2021). Crucially, in these scenarios, the action is not directly causal for the outcome, as the agent passively experiences an external

event (e.g. car crash, robbery). This breaks the usual link between typicality of actions and foreseeability of outcome, and makes assumptions and inferences about the agent's epistemic state speculative. Kutscher and Feldman (2019) find that people's anticipation of the likelihood of a negative incident is the same for atypical compared to typical actions (Macrae, 1992; Turley, Sanna, & Reiter, 1995).

However, Macrae (1992) presumes that an agent who deviates from a routine behaviour for which negative incidents have been absent in the past risks greater uncertainty towards the consequences of their abnormal action than someone who abides with past behaviour. They find that perpetrators were judged more negligent if their behaviour was preceded by exceptional circumstances. Monroe and Ysidron (2021) demonstrate that an agent's deviation from typical behaviour facilitates attributions of a particular desire and choice for this kind of behaviour. Spontaneous inference when information about mental states is absent have been argued to function as a tool to make sense of other people's past actions, promote accountability or predict future behavior (Young & Saxe, 2009; Young & Tsoi, 2013). While the exact link between mental state inference and causal judgments in cases of single abnormal actions is yet to be shown, there is mounting evidence for mental states to play a role here, too.

Abnormal objects

Deviations of 'normal behaviour' not only affect agents, but also causal attributions to inanimate objects (Cheng & Novick, 1991; Gerstenberg & Icard, 2020; Henne, Bello, Khemlani, & De Brigard, 2019; Knobe, 2009; Kominsky & Phillips, 2019). Gerstenberg and Icard (2020) investigate the effect of statistical normality in causal reasoning about inanimate objects in physical systems. They show their participants videos with physically realistic collisions of billiard balls, and find that participants select a statistically unlikely event as cause for an outcome in a conjunctive causal structure (see also Kirfel et al., 2020). Henne et al. (2021) extend this paradigm to a variety of different inanimate objects, and show that norm-violating object behaviour affects prospective causal judgments (i.e. before the outcome occurs), independent of the perceived agency of these objects. Likewise, malfunctioning artifacts are seen as more causal for an outcome than those which function normally (Kominsky & Phillips, 2019).

Normality incorporating accounts (Icard et al., 2017; Kominsky & Phillips, 2019; Phillips et al., 2015) have the advantage of predicting causal attributions to a variety of phenomena by reference to the same mechanism, without the need to introduce additional factors such as epistemic states. Based on our findings, we argue that theories of causal attribution need to include the epistemic dimension in order to make accurate predictions about human causal agents. People do not seem to prefer abnormal agentive behaviour as causes per se, but only in combination with epistemic states (Kirfel & Lagnado, 2019; Samland & Waldmann, 2015, 2016). This finding might shed further light on the fundamental differences in our causal thinking about inanimate objects and social agents (Fausey & Boroditsky, 2007; Kelemen, 1999; Saxe, Tzelnic, & Carey, 2007; Strickland, Silver, & Keil, 2017). As discussed earlier, counterfactual reasoning about alternatives to an action might depend on the acting agent's epistemic state. Likewise, it is possible that in case of human actions, people might prefer an agent's epistemic states as a locus of (causal) intervention (Halpern, 2016; Woodward, 2001). Given that inanimate objects do not have mental states,

this additional dimension for counterfactual reasoning or intervention does not apply. As a result, including epistemic states comes at the expense of a uniform normality-based theory of causal attributions, but might sharpen our theories of causal cognition with regards to the differences between various "kinds" of causes.

Normal causes in disjunctive structures

Recent studies show that the change from a conjunctive to disjunctive causal structure flips people's preferences to a normal, typical or frequent cause (Gerstenberg & Icard, 2020; Icard et al., 2017). If the motion detector scenario is triggered as soon as one person enters the building, people attribute more causality to the travel agent who frequently comes in at 8:45am compared to the design agent who comes in for the first time at 8:45 that day (Icard et al., 2017). The finding that the influence of normality on causal attributions is dependent on the underlying causal structure has challenged the idea of a uniform causal preference for abnormal factors. Icard et al. (2017) propose that the influence of normality is weighted differently by the sufficiency and necessity of a causal factor. Others have argued that the correspondence between the normality of an outcome and the normality of a cause is decisive for causal judgements (Harinen, 2017; Wells & Gavanski, 1989) (see Kirfel & Lagnado, 2018, for a comparison).

In light of our model, in a disjunctive causal structure, there is no link between normality of the other agent's action and the foreseeability of the outcome. The foreseeability of the outcome is high for each agent, independent of other people's actions. In this respect, the current version of our model would not predict an epistemic difference between an abnormal and normal agent. Taking into consideration the co-variation between normality of cause and outcome, however, raises new questions about the agents' epistemic states. On the one hand, it might be argued that it not only matters whether the agent expects the outcome to occur given their action, but also whether they expect their outcome to occur given their *non-action*. In a disjunctive structure, an agent will expect their action not to be necessary (or "counterfactually relevant") for the outcome if another agent acts normally or frequently – the outcome will be caused by someone else in any case. On the other hand, an agent who typically brings about a certain outcome in a disjunctive structure might know more about this outcome, be more familiar with it, or desire or intend the outcome to occur more often, than someone who has never caused it before (Alwin, 1973; Buss & Craik, 1983).

Hence, at the current stage, a variation of our Bayesian network model might be needed in order to account for increased causal attribution to typical actions in disjunctive structures. This variation would require the inference about a latent variable that is influenced by how typical, or normal an action is. Such a variation would still be compatible with the influence of epistemic or mental states, broadly considered. Under certain circumstances people are held more responsible for expected, 'normal' actions (Johnson & Rips, 2015). We take these and our findings to highlight the importance to monitor and control the various inferences people make about agents' mental state and dispositions, and their potential role in attributions of causality.

Conclusion

In the aftermath of the 2015 New Taipei Water park explosion, the event organiser was found guilty of the incident and was sentenced to five years in prison (Pan, 2016, April 26). Although modern wood factories are equipped with high tech dust collection systems, employees face hefty fines if they smoke outside of designated smoking areas. A causal analysis of incidents like these two dust explosion cases lays an important foundation for legal assessment, as well as the creation of prevention measures. Both cases are structurally similar – in both, the production of dust particles together with an act of ignition lead to a dust explosion. Yet, our judgment about the primary cause differs in these cases. We take the agent who arranged the spray of combustible dust particles to be the major cause in the outdoor festival incident, but judge the agent who ignited a cigarette to have caused the explosion in the wood factory. Crucially, in both cases, the action that is selected to be the main cause is, in the respective context, the more atypical one.

In this paper, we have argued that the typicality of actions changes how much an agent can foresee the consequences of their action. We have shown that it is this very epistemic asymmetry between a normally and abnormally acting agent that influences people's causal judgments. Both employee and party organiser acted abnormally, but when acting, they also could have foreseen the event of a dust explosion to a greater extent. While further research is needed on this topic, the connection between action typicality and epistemic states brings us one step closer to understanding the enigmatic role of normality in causal cognition.

Acknowledgments

We would like to thank Neele Engelmann, Tobias Gerstenberg, Paul Henne, Jonathan Kominsky, Toby Pilditch, Simon Stephan and Alex Wiegmann for helpful feedback.

References

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*(4), 556–574.

Alicke, M. D., & Rose, D. (2012). Culpable control and causal deviance. *Journal of Personality and Social Psychology Compass*, *6*, 723–725.

Alicke, M. D., Rose, D., & Bloom, D. (2012). Causation, norm violation, and culpable control. *The Journal of Philosophy*, *108*(12), 670–696.

Alwin, D. F. (1973). Making inferences from attitude-behavior correlations. *Sociometry*, 253–278.

Baker, C., Saxe, R., & Tenenbaum, J. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 33).

Baker, C. L. (2012). *Bayesian theory of mind: Modeling human reasoning about beliefs, desires, goals, and social relations* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Barton, K., & Barton, M. K. (2015). Package 'mumin'. *Version*, *1*, 18.

Brandstätter, V., Lengfelder, A., & Gollwitzer, P. M. (2001). Implementation intentions and efficient action initiation. *Journal of personality and social psychology*, *81*(5), 946.

Buss, D. M., & Craik, K. H. (1983). The act frequency approach to personality. *Psychological Review*, *90*(2), 105.

Chee, C. S., & Murachver, T. (2012). Intention attribution in theory of mind and moral judgment. *Psychological Studies*, *57*(1), 40–45.

Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition*, *40*, 83–120.

Clark, C. J., Baumeister, R. F., & Ditto, P. H. (2017). Making punishment palatable: Belief in free will alleviates punitive distress. *Consciousness and Cognition*, *51*, 193–211.

Clark, C. J., Luguri, J. B., Ditto, P. H., Knobe, J., Shariff, A. F., & Baumeister, R. F. (2014). Free to punish: A motivated account of free will belief. *Journal of personality and social psychology*, *106*(4), 501.

Danner, U. N., Aarts, H., & de Vries, N. K. (2008). Habit vs. intention in the prediction of future behaviour: The role of frequency, context stability and mental accessibility of past behaviour. *British Journal of Social Psychology*, *47*(2), 245–265.

Driver, J. (2008). Attributions of causation and moral responsibility.

Epstein, S. (1979). The stability of behavior: I. on predicting most of the people much of the time. *Journal of personality and social psychology*, *37*(7), 1097.

Epstude, K., & Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and social psychology review*, *12*(2), 168–192.

Fausey, C. M., & Boroditsky, L. (2007). Language changes causal attributions about agents and objects. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 29).

Fillon, A., Kutscher, L., & Feldman, G. (2020). Impact of past behaviour normality: meta-analysis of exceptionality effect. *Cognition and Emotion*, 1–21.

Fillon, A., Lantian, A., Feldman, G., & N'gbala, A. (2019). Exceptionality effect in agency: Exceptional choices attributed higher free will than routine.

FitzPatrick, W. J. (2008). Moral responsibility and normative ignorance: Answering a new skeptical challenge. *Ethics*, *118*(4), 589–613.

FitzPatrick, W. J. (2017). Unwitting wrongdoing, reasonable expectations, and blameworthiness. *Responsibility: The epistemic condition*, 29–46.

Gerstenberg, T., & Icard, T. (2020). Expectations affect physical causation judgments. *Journal of Experimental Psychology: General*, *149*(3), 599.

Gerstenberg, T., Ullman, T. D., Nagel, J., Kleiman-Weiner, M., Lagnado, D. A., & Tenenbaum, J. B. (2018). Lucky or clever? from expectations to responsibility judgments. *Cognition*, *177*, 122–141.

Gilbert, E. A., Tenney, E. R., Holland, C. R., & Spellman, B. A. (2015). Counterfactuals, control, and causation: Why knowledgeable people get blamed more. *Personality and Social Psychology Bulletin*, *41*(5), 643–658.

Green, P., & MacLeod, C. J. (2016). Simr: an r package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498.

Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Halpern, J. Y. (2016). *Actual causality*. MIT Press.

Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *British Journal for the Philosophy of Science*, *66*, 413–457.

Harinen, T. (2017, jul). Normal causes for normal effects: Reinvigorating the correspondence hypothesis about judgments of actual causation. *Erkenntnis*. Retrieved from `https://doi.org/10.1007%2Fs10670-017-9876-4` doi: 10.1007/s10670-017-9876-4

Hart, H. L. A., & Honoré, T. (1959/1985). *Causation in the law*. New York: Oxford University Press.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American journal of psychology*, *57*(2), 243–259.

Henne, P., Bello, P., Khemlani, S., & De Brigard, F. (2019). Norms and the meaning of omissive enabling conditions.

Henne, P., Kulesza, A., Perez, K., & Houcek, A. (2020). Counterfactual thinking and recency effects in causal judgment.

Henne, P., Niemi, L., Pinillos, Á., De Brigard, F., & Knobe, J. (2019). A counterfactual explanation for the action effect in causal judgment. *Cognition*, *190*, 157–164.

Henne, P., O'Neill, K., Bello, P., Khemlani, S., & De Brigard, F. (2021). Norms affect prospective causal judgments. *Cognitive Science*, *45*(1), e12931.

Henne, P., Pinillos, Á., & De Brigard, F. (2017). Cause by omission and norm: Not watering plants. *Australasian Journal of Philosophy*, *95*(2), 270–283.

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

Brighton, UK: Harvester Press.

Hilton, D. J., & Jaspars, J. M. (1987). The explanation of occurrences and non-occurrences: A test of the inductive logic model of causal attribution. *British Journal of Social Psychology*, *26*(3), 189–201.

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, *93*(1), 75–88.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, *11*, 587–612.

Icard, T., & Knobe, J. (2016). Causality, normality, and sampling propensity. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 800–805). Austin, TX: Cognitive Science Society.

Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, *161*, 80–93. Retrieved from `https://doi.org/10.1016%2Fj.cognition.2017.01.010` doi: 10.1016/j.cognition.2017.01.010

Johnson, S. G., & Rips, L. J. (2015). Do the right thing: The assumption of optimality in lay decision theory and causal judgment. *Cognitive psychology*, *77*, 42–76.

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in cognitive sciences*, *19*(4), 201–214.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. In *Advances in experimental social psychology* (Vol. 2, pp. 219–266). Elsevier.

Jones, E. E., Davis, K. E., & Gergen, K. J. (1961). Role playing variations and their informational value for person perception. *The Journal of Abnormal and Social Psychology*, *63*(2), 302.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of experimental social psychology*, *3*(1), 1–24.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*(2), 136–153.

Kelemen, D. (1999). Function, goals and intention: Children's teleological reasoning about objects. *Trends in Cognitive Sciences*, *3*(12), 461–468.

Kelley, H. H. (1967). Attribution theory in social psychology. In *Nebraska symposium on motivation.*

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, *28*(2), 107–128.

Khemlani, S. S., Byrne, R. M., & Johnson-Laird, P. N. (2018). Facts and possibilities: A model-based theory of sentential reasoning. *Cognitive science*, *42*(6), 1887–1924.

Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, *89*(2), 191–204.

Kirfel, L., Icard, T., & Gerstenberg, T. (2020, May). *Inference from explanation.* PsyArXiv. Retrieved from `psyarxiv.com/x5mqc` doi: 10.31234/osf.io/x5mqc

Kirfel, L., & Lagnado, D. A. (2017). "Oops, I did it again." The impact of frequent behaviour on causal judgement. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 2420–2425). Austin, TX: Cognitive Science Society.

Kirfel, L., & Lagnado, D. A. (2018). Statistical norm effects in causal cognition. In

T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 615–620). Austin, TX: Cognitive Science Society.

Kirfel, L., & Lagnado, D. A. (2019). I know what you did last summer (and how often). epistemic states and statistical normality in causal judgments. *Proceedings of the 41st Annual Conference of the Cognitive Science Society*.

Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, *16*(2), 309–324.

Knobe, J. (2009). Folk judgments of causation. *Studies In History and Philosophy of Science Part A*, *40*(2), 238–242.

Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The cognitive science of morality: intuition and diversity* (Vol. 2). The MIT Press.

Kominsky, J. F., & Phillips, J. (2019, Oct). Immoral professors and malfunctioning tools: Counterfactual relevance accounts explain the effect of norm violations on causal selection. *Cognitive Science*, *43*(11). Retrieved from `http://dx.doi.org/10.1111/cogs.12792` doi: 10.1111/cogs.12792

Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D. A., & Knobe, J. (2015). Causal superseding. *Cognition*, *137*, 196–209.

Kominsky, J. F., Phillips, J., Knobe, J., Gerstenberg, T., & Lagnado, D. A. (2014). Causal supersession. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 761–766). Austin, TX: Cognitive Science Society.

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological science*, *14*(5), 402–408.

Kutscher, L., & Feldman, G. (2019). The impact of past behaviour normality on regret: replication and extension of three experiments of the exceptionality effect. *Cognition and Emotion*, *33*(5), 901–914.

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

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation. *Causal learning: Psychology, philosophy, and computation*, 154–172.

Lucas, C. G., Griffiths, T. L., Xu, F., & Fawcett, C. (2009). A rational model of preference learning and choice prediction by children. In *Advances in neural information processing systems* (pp. 985–992).

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford: Clarendon Press.

Macrae, C. N. (1992). A tale of two curries: Counterfactual thinking and accident-related judgments. *Personality and Social Psychology Bulletin*, *18*(1), 84–87.

Maselli, M. D., & Altrocchi, J. (1969). Attribution of intent. *Psychological Bulletin*, *71*(6), 445.

McGrath, S. (2005). Causation by omission: A dilemma. *Philosophical Studies*, *123*(1-2), 125–148.

Miller, D. T., & McFarland, C. (1986). Counterfactual thinking and victim compensation: A test of norm theory. *Personality and Social Psychology Bulletin*, *12*(4), 513–519.

Miller, D. T., Turnbull, W., & McFarland, C. (1990). Counterfactual thinking and social perception: Thinking about what might have been. In *Advances in experimental social psychology* (Vol. 23, pp. 305–331). Elsevier.

Monroe, A., & Ysidron, D. (2021). Not so motivated after all? three replication attempts and a theoretical challenge to a morally motivated belief in free will. *Journal of Experimental psychology. General*, *150(1):e1-e12*. Retrieved from `https:doi://10.1037/xge0000788`

Murray, S., & Vargas, M. (2018). Vigilance and control. *Philosophical Studies*, 1–19.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in ecology and evolution*, *4*(2), 133–142.

Pan, J. (2016, April 26). 'color play asia' organizer found guilty. *Taipei Times*. Retrieved from `https://www.bbc.co.uk/news/world-asia-33300970`

Pearl, J. (2009). *Causality*. Cambridge university press.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect* (1st ed.). USA: Basic Books, Inc.

Perugini, M., & Bagozzi, R. P. (2001). The role of desires and anticipated emotions in goal-directed behaviours: Broadening and deepening the theory of planned behaviour. *British Journal of Social Psychology*, *40*(1), 79–98.

Phillips, J., Luguri, J., & Knobe, J. (2015). Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition*, *145*, 30–42.

Reuter, K., Kirfel, L., Van Riel, R., & Barlassina, L. (2014). The good, the bad, and the timely: how temporal order and moral judgment influence causal selection. *Frontiers in psychology*, *5*, 1336.

Roese, N. J. (1997). Counterfactual thinking. *Psychological bulletin*, *121*(1), 133.

Roese, N. J., & Epstude, K. (2017). The functional theory of counterfactual thinking: New evidence, new challenges, new insights. In *Advances in experimental social psychology* (Vol. 56, pp. 1–79). Elsevier.

Ryle, G. (2009). *The concept of mind*. Routledge.

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*(2), 125–130. Retrieved from `https://doi.org/10.1037%2Fxge0000138` doi: 10.1037/xge0000138

Samland, J., & Waldmann, M. R. (2014). Do social norms influence causal inferences. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Samland, J., & Waldmann, M. R. (2015). Highlighting the causal meaning of causal test questions in contexts of norm violations. In D. C. Noelle et al. (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 2092–2097). Austin, TX: Cognitive Science Society.

Samland, J., & Waldmann, M. R. (2016). How prescriptive norms influence causal inferences. *Cognition*, *156*, 164–176. Retrieved from `https://doi.org/10.1016%2Fj.cognition.2016.07.007` doi: 10.1016/j.cognition.2016.07.007

Sartorio, C. (2009). Omissions and causalism. *Noûs*, *43*(3), 513–530.

Saxe, R., & Houlihan, S. D. (2017). Formalizing emotion concepts within a bayesian model of theory of mind. *Current opinion in Psychology*, *17*, 15–21.

Saxe, R., Tzelnic, T., & Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental psychology*, *43*(1), 149.

Sher, G. (2009). Who knew. *Oxford University Press USA. Smart, JJC (1961).'Free-Will, Praise and Blame'. Mind*, *70*, 291–306.

Strickland, B., Silver, I., & Keil, F. C. (2017). The texture of causal construals: Domain-specific biases shape causal inferences from discourse. *Memory & cognition*, *45*(3), 442–455.

Sytsma, J. (2019a). The character of causation: Investigating the impact of character, knowledge, and desire on causal attributions. *pre-print*.

Sytsma, J. (2019b). The effects of single versus joint evaluations on causal attributions.

Sytsma, J. (2020). Causation, responsibility, and typicality. *Review of Philosophy and Psychology*.

Sytsma, J., & Livengood, J. (2019). Causal attributions and the trolley problem. *pre-print*.

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*(4), 814–820.

"Taiwan Formosa Water Park explosion injures hundreds". (2015, June 28). *BBC World Asia*. Retrieved from `https://www.bbc.co.uk/news/world-asia-33300970`

Turley, K. J., Sanna, L. J., & Reiter, R. L. (1995). Counterfactual thinking and perceptions of rape. *Basic and Applied Social Psychology*, *17*(3), 285–303.

U.S. Chemical Safety and Hazard Investigation Board. (2006). Investigation report. combustible dust hazard study. Retrieved from `https://www.csb.gov/combustible-dust-hazard-investigation`

Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition*, *116*(1), 87–100.

Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of personality and social psychology*, *56*(2), 161.

Wiener, R. L., & Pritchard, C. C. (1994). Negligence law and mental mutation. In *Applications of heuristics and biases to social issues* (pp. 117–136). Springer.

Willemsen, P. (2018). Omissions and expectations: A new approach to the things we failed to do. *Synthese*, *195*(4), 1587–1614.

Willemsen, P., & Reuter, K. (2016). Is there really an omission effect? *Philosophical Psychology*, *29*(8), 1142–1159.

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science*, *34*(5), 776–806.

Woodward, J. (2001). Causation and manipulability.

Young, L., & Saxe, R. (2009). An fmri investigation of spontaneous mental state inference for moral judgment. *Journal of cognitive neuroscience*, *21*(7), 1396–1405.

Young, L., & Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. *Social and personality psychology compass*, *7*(8), 585–604.

Zimmerman, M. J. (1997). Moral responsibility and ignorance. *Ethics*, *107*(3), 410–426.

Appendix A
Comprehension Check Questions Experiment 1

Participants answered ten comprehension check questions in total, five for each *'normality'* condition.

[Part 1]

1. Tom and Ben ...
   - work in the same office.
   - work in separate offices.
2. Because of their office situation, Tom and Ben ...
   - know each other well.
   - do not know each other, and have never met or seen each other.
3. Because of their office situation, Tom and Ben ...
   - know what the other person is doing during the day.
   - do not know what the other person is doing during the day.
4. The use of how many microwaves does it take to produce a power failure on Friday?
   - one microwave.
   - two microwaves.

[Part 2]

5. Who used a microwave frequently (rather than infrequently) this week? (Multiple answers possible)
   - Tom.
   - Ben.

Appendix B
Comprehension Check Questions Experiment 2

[Part 1]

1. Tom and Ben ...
   - work in the same office.
   - work in separate offices.
2. Because of their office situation, Tom and Ben ...
   - know each other well.
   - do not know each other, and have never met or seen each other.
3. Because of their office situation, Tom and Ben ...
   - know what the other person is doing during the day.
   - do not know what the other person is doing during the day.
4. The use of how many microwaves does it take to produce a power failure on Friday?
   - one microwave.
   - two microwaves.

[Part 2]

5. Who used a microwave frequently (rather than infrequently) this week? (Multiple answers possible)
   - Tom.
   - Ben.

Appendix C
Comprehension Check Questions Experiment 3

[Part 1]

1. From Monday to Thursday, Tom and Ben ...
   - work in the same office.
   - work in separate offices.
2. Because of their office situation, Tom and Ben ...
   - know each other well.
   - do not know each other, and have never met or seen each other.
3. Because of their office situation, Tom and Ben ...
   - know what the other person is doing during the day.
   - do not know what the other person is doing during the day.
4. The use of how many microwaves does it take to produce a power failure on Friday?
   - one microwave.
   - two microwaves.

[Part 2]

5. Who used a microwave frequently (rather than infrequently) this week? (Multiple answers possible)
   - Tom.
   - Ben.

Appendix D
Comprehension Check Questions Experiment 4

[Part 1]

1. Tom and Ben ...
   - work in the same office.
   - work in separate offices.
2. Because of their office situation, Tom and Ben ...
   - know each other well.
   - do not know each other, and have never met or seen each other.
3. Because of their office situation, Tom and Ben ...
   - know what the other person is doing during the day.

- do not know what the other person is doing during the day.

4. The use of how many microwaves does it take to produce a power failure on Friday?
   - one microwave.
   - two microwaves.

[Part 2]

5. Who used a microwave frequently (rather than infrequently) this week? (Multiple answers possible)
   - Tom.
   - Ben.