Cognition 161 (2017) 80–93

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Normality and actual causal strength

Thomas F. Icard [a,*], Jonathan F. Kominsky [b], Joshua Knobe [c]

[a] Department of Philosophy and Symbolic Systems Program, Stanford University, United States
[b] Department of Psychology, Harvard University, United States
[c] Program in Cognitive Science and Department of Philosophy, Yale University, United States

ARTICLE INFO

Article history:
Received 1 July 2016
Revised 4 January 2017
Accepted 9 January 2017
Available online 01 February 2017

Keywords:
Causal reasoning
Normality
Counterfactuals
Actual causation
Sampling
Bayes nets

ABSTRACT

Existing research suggests that people's judgments of actual causation can be influenced by the degree to which they regard certain events as normal. We develop an explanation for this phenomenon that draws on standard tools from the literature on graphical causal models and, in particular, on the idea of probabilistic sampling. Using these tools, we propose a new measure of actual causal strength. This measure accurately captures three effects of normality on causal judgment that have been observed in existing studies. More importantly, the measure predicts a new effect ("abnormal deflation"). Two studies show that people's judgments do, in fact, show this new effect. Taken together, the patterns of people's causal judgments thereby provide support for the proposed explanation.

1. Introduction

Judgments of actual causation—concerning the extent to which a given event or factor caused some outcome—have been at the center of attention in work on causal cognition. One intriguing phenomenon that has long been recognized is that people's judgments of actual causation can be influenced by the degree to which they regard certain events as *normal*. In recent years, this effect has been explored both in experimental studies and in formal models (e.g., Halpern & Hitchcock, 2015; Kominsky, Phillips, Gerstenberg, Lagnado, & Knobe, 2015; Phillips, Luguri, & Knobe, 2015).

Considerable debate remains about how to explain the effect. One approach would be to posit some independent factor, outside the core processes involved in causal cognition, that explains the impact of normality. For example, one might hypothesize that the impact of normality is the result of a motivational bias or of conversational pragmatics (e.g., Alicke, Rose, & Bloom, 2011; Driver, 2008; Samland & Waldmann, 2016). Our aim is to explore a different approach. We suggest that the impact of normality might be explained by basic facts about how causal cognition works. Our explanation will rely on standard tools familiar from the literature on graphical causal models and, in particular, on a specific way of thinking about computations over these causal models involving probabilistic sampling. Drawing on these ideas, we propose a *measure of actual causal strength*. Our hypothesis is that this actual causal strength measure will help to explain the impact of normality.

The key evidence for this hypothesis comes from facts about the precise pattern of the impact of normality on causal judgment. The most well-studied effect in this domain is the tendency whereby people are inclined to regard abnormal events as more causal in certain cases. However, as we will see, the actual pattern is considerably more complex. There are also cases in which people's judgments about the causal status of a given event depend on the normality of *other* events, and these effects in turn depend on the details of the causal structure in question (Section 2). It can be shown that the causal strength measure proposed below accurately captures the details of these patterns (Section 4). More importantly, this measure generates a novel prediction, namely, that there should be cases in which abnormal events are systematically regarded as *less* causal. Two new experiments show that this prediction is in fact borne out (Section 5). Taken together, the patterns thereby provide support for the present approach.

2. Three effects of normality on actual causation judgments

Before discussing the impact of normality on people's actual causation judgments, it may helpful to clarify the notion of normality itself. To begin with, we need to distinguish two kinds of norms. First, there are purely *statistical norms*. For example, winter months in Oregon generally tend to be cloudy and overcast, so if Oregon ever had a sunny winter, this weather could be said to be violating a statistical norm. Second, there are *prescriptive norms*. These norms are constituted not by purely statistical tendencies but by the way things ought to be or are supposed to be. Suppose we believe that the police ought to accord criminal defendants cer-

* Corresponding author.
E-mail address: icard@stanford.edu (T.F. Icard).

http://dx.doi.org/10.1016/j.cognition.2017.01.010
0010-0277/

tain rights. Even if we do not believe that the police actually do tend to accord defendants these rights, we might think that failing to do so is a violation of a prescriptive norm.

A question arises as to which of these two types of norms are reflected in ordinary judgments of actual causation. As explained below, existing research suggests that actual causation judgments are influenced by both kinds of norms. More strikingly, these two kinds of norms show the same pattern of impact on such judgments. As a result, researchers have suggested that it might be helpful to posit a single undifferentiated notion of normality that integrates both statistical and prescriptive considerations (Halpern & Hitchcock, 2015; Kominsky et al., 2015). On this approach, an event counts as "abnormal" to the extent that it either violates a statistical norm or violates a prescriptive norm, and as "normal" to the extent that it follows both of these types of norms. Difficult questions arise about precisely how statistical and prescriptive considerations are integrated into an undifferentiated notion, but we will not be resolving those questions here (cf. Bear & Knobe, in press). Instead, we focus on three specific ways in which normality—both statistical and prescriptive—impacts people's intuitions about actual causation.

2.1. First effect: abnormal inflation

Abnormal inflation is the simplest of the three effects. We will eventually be introducing a formal framework in which it can be described more precisely, but for the moment, we offer the following rough characterization:

> *Suppose that an outcome depends on a causal factor C as well as an alternative causal factor A, such that the outcome will only occur if both C and A occur. Then people will be more inclined to say that C caused the outcome when they regard C as abnormal than when they regard C as normal.*

This basic effect appears to arise both for statistical norms and for prescriptive norms.

It has been known for decades that actual causation judgments can be influenced by statistical norms (Hilton & Slugoski, 1986). Suppose that a person leaves a lit match on the ground and thereby starts a forest fire. In such a case, the fire would not have begun if there had been no oxygen in the atmosphere, and yet we would not ordinarily say that the oxygen caused the fire. Why is this? The answer appears to involve the fact that it is so (statistically) normal for the atmosphere to contain oxygen. Our intuitions should therefore be very different if we consider a case in which the presence of oxygen is abnormal. (Suppose that matches were struck on a regular basis but there is never a fire except on the very rare occasions when oxygen is present.) In such a case, people should be more inclined to regard the presence of oxygen as a cause.

Strikingly, this same effect arises for prescriptive norms. Consider the following case:

> The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take pens, but faculty members are supposed to buy their own.
>
> The administrative assistants typically do take the pens. Unfortunately, so do the faculty members. The receptionist has repeatedly e-mailed them reminders that only administrators are allowed to take the pens.
>
> On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later, that day, the receptionist needs to take an important message... but she has a problem. There are no pens left on her desk.

Faced with this case, participants tend to say that the professor caused the problem (Knobe & Fraser, 2008; Phillips et al., 2015). But now suppose that we change the first paragraph of the case in such a way as to make the professor's action not violate a prescriptive norm:

> The receptionist in the philosophy department keeps her desk stocked with pens. Both the administrative assistants and the faculty members are allowed to take the pens, and both the administrative assistants and the faculty members typically do take the pens. The receptionist has repeatedly e-mailed them reminders that both administrators and professors are allowed to take the pens.

Faced with this latter version, participants are significantly less inclined to say that the professor caused the problem (Phillips et al., 2015). Yet the two cases do not appear to differ from the perspective of statistical normality; the difference is rather in the degree to which the agent violates a prescriptive norm. The result thereby suggests that prescriptive norms impact causal judgments.

Within existing work, this first effect has been investigated in far more detail than the others we will discuss (see, e.g., Danks, Rose, & Machery, 2014; Phillips et al., 2015; Samland, Josephs, Waldmann, & Rakoczy, 2016). One of the most important findings to come out of this work is that the effect really does involve prescriptive considerations and cannot be reduced to a matter of purely statistical norms. First, one can explicitly pit the prescriptive against the statistical. In one study, participants were told that administrative assistants were allowed to take pens and faculty members were not (a prescriptive norm) but that in actual fact administrators never did take pens while faculty members always did (a statistical norm). People's judgments ended up being affected more by the prescriptive than by the statistical, with participants tending on the whole to say that the administrative assistant did not cause the problem while the faculty member did (Roxborough & Cumby, 2009). Second, one can look at cases in which different people have different prescriptive judgments. For example, one study looked at controversial political issues (abortion, euthanasia) and found that people who had opposing moral judgments about these issues arrived at correspondingly opposing causal judgments about people who performed the relevant actions (Cushman, Knobe, & Sinnott-Armstrong, 2008).

Yet, though existing work clearly shows that both statistical and prescriptive norms can lead to abnormal inflation, controversy remains regarding the explanation of this effect. Researchers have suggested that the effect might arise as a result of conversational pragmatics (Driver, 2008), motivational bias (Alicke et al., 2011), relativity to frameworks (Strevens, 2013), responsibility attributions (Sytsma, Livengood, & Rose, 2012), or people's understanding of the question (Samland & Waldmann, 2016). Here, we will be exploring a general approach that has been defended by a number of researchers in recent years, namely, that abnormal inflation reflects a process in which certain counterfactuals are treated as in some way more relevant than others (Blanchard & Schaffer, 2016; Halpern & Hitchcock, 2015; Knobe, 2010; Phillips et al., 2015).

If one simply looks at the abnormal inflation effect in isolation, it seems that any of these theories might be able to predict the experimental findings. The advantage of the account we will be offering emerges most clearly when we broaden the scope of our inquiry, looking at a number of different effects and trying to develop an account that predicts the pattern as a whole.

2.2. Second effect: supersession

Supersession is an effect whereby the apparent normality of one factor can actually influence the degree to which *other* factors are regarded as causes. The effect can be characterized roughly as follows:

> *Suppose an outcome depends on a causal factor C as well as an alternative causal factor A, such that the outcome will only occur*