**Forthcoming in *Advances in Experimental Philosophy of Science***

**Causal Judgment: What Can Philosophy Learn from Experiment? What Can It Contribute to Experiment[[1]](#footnote-1)?**

James Woodward

**1. Introduction**.

The topic of this volume is experimental approaches to philosophy of science. It is uncontroversial that empirical research, broadly speaking, is relevant to many topics in philosophy of science—philosophers interested in space and time should make use of the best empirical theorizing on these topics, and similarly for philosophers of biology interested in the structure of evolutionary theory, philosophers of psychology interested in perceptual processing and so on. Issues having to do with the relevance of empirical or more specifically experimental research become more controversial, however, when we move to topics having to do with scientific method, or more generally, to issues that are within the purview of general philosophy of science—issues having to do with theory testing, confirmation, evidence, causal reasoning, explanation and so on. A natural thought (and one I endorse below) is that these issues have an important *normative* component: they have to do with how we *ought* to think and reason. If so, one might then wonder how discoveries about how people in fact reason could be relevant to these normative concerns.

My aim in this essay is to address this issue in a particular context: causal reasoning. But I also have some broader aims; I will also compare appeals to experimental results with two alternatives: the traditional philosophical method of appealing to “intuitions” about cases and more recent developments in “experimental philosophy” some (but by no means all) of which appeal to survey-like results to address philosophical questions. My overall goal is to use this work to explore some more general questions having to do with how empirical results might best be brought to bear on issues in philosophy of science. Causation and causal explanation (hereafter I will sometimes just write “causation”) are of course topics of long-standing importance in philosophy of science and in philosophy more generally. These topics have also recently undergone a rich theoretical development in disciplines like statistics, econometrics and computer science. At the same time, the past several decades have seen an explosion of research on the empirical psychology of causal and explanatory reasoning, much of it conducted by psychologists, but also involving contributions from researchers in philosophy, primatology, and animal behavior, among other disciplines. It is natural to suppose that these bodies of work can usefully inform one another, but the details of how this might work are far from obvious.

Insofar as my focus is empirical, it will be on features of causal reasoning (such as the role of invariance) that are reflected directly in scientific practice. Much of the work having this character has been conducted by psychologists such as Gopnik, Cheng, and Tenenbaum. By comparison, there has been much less work on causal judgment and cognition by experimental philosophers and the research that has been conducted (e.g. Hitchcock and Knobe, 2009, ) has tended to focus on actual cause judgments and on the role of norms in causal selection. Although important for understanding aspects of lay causal thinking, these issues are (in many cases) less central to causal thinking in science, which tends to focus on type-causal claims[[2]](#footnote-2). Hence my focus on empirical research conducted by psychologists.

**2. Normative and Descriptive Theorizing.**

Normative theories of causal reasoning purport to describe how we ought to think about causation or what good or correct causal reasoning consists in; descriptive theories describe how we do reason and the causal factors (including the computations and information processing) that underlie such reasoning. A normative focus is explicit in work on causation in disciplines like statistics and machine learning. It is also the dominant way of thinking about theories of causation in empirical psychology that have descriptive aspirations—a point that is less paradoxical than it may initially seem, for reasons to be explained below. It is less common to view the standard philosophical theories of causation as normative proposals but I suggest this is a very natural and illuminating way of understanding their significance. Here the normative content has less directly to do with inference but instead with how we should conceptualize causation (including distinctions we ought to draw among various kinds of causal claims), which relationships we should treat as causal and how causal claims should be connected to other notions of interest, such as counterfactual dependence and probability. For example, David Lewis’ well-known theory might be regarded as a normative proposal to the effect that we should regard only those events that stand in a certain relationship of counterfactual dependence as causally related, that we should reason about causal relationships by means of such counterfactuals, assessing these in turn in terms of the particular similarity measure across possible worlds advocated by Lewis and so on. Theories according to which causal relationships must involve the transference of energy and momentum (so that, for example, absences cannot be causes) can be regarded as proposals that we ought to regard certain relationships and not others as causal.

The characterization of the above theories as “normative” raises the obvious question of where their normativity comes from. I address this issue in more detail in Section 3, where I argue that this is a matter of means/ ends justification: there are certain ends or goals distinctively associated causal reasoning and we justify claims that this reasoning should possess various features by showing that these features conduce effectively to these goals[[3]](#footnote-3). .

Thinking about philosophical theories of causation as normative in this way contrasts with two more standard ways of thinking about their significance, according to which they are either attempts to capture the content of “our concept” of causation or else (for those who are more metaphysically inclined) attempts to describe the “nature” of causation or “causation as it is in itself”. From my perspective, such construals efface the normative content of these theories and make it sound as though their intent is purely descriptive or reportorial (of our concept of causation or causation itself)[[4]](#footnote-4). An additional problem is that it is dubious that there is any such thing as “our concept” of causation if by this is meant a single “concept” shared by all competent users. Instead, even from the point of view of a single overarching treatment of causation, such as a counterfactual or interventionist theory, there are a number of distinct varieties of causal relationship that should be distinguished -- actual cause, various type causal notions, notions of direct and indirect effect, as well as different notions of indirect effect[[5]](#footnote-5) . In addition to this, empirical research (some described below) shows that people exhibit non-trivial heterogeneity in some casual judgments. If such judgments are interpreted as evidence for some concept of causation that “we” possess, this raises the question of whether those who judge differently have different concepts. In interpreting theories of causation as claims about concepts of causation, we saddle ourselves with the problem of distinguishing those features of practices of causal reasoning and judgment that are reflective of or constitutive of the concept (or concepts) of causation from those that are not. On any reasonable account of concepts, there will be many interesting and important features of casual reasoning that are not constitutive of our concept of causation. —examples will be provided below. In fact, experimental results in the cognitive psychology of causal reasoning are rarely presented as claims about the (or a) concept or concepts of causation—instead researchers talk about different sorts of representations of causal relations and the computations associated with these, different strategies for causal learning and judgment and so on. Philosophers would do well to follow this practice whether their goals are descriptive or normative[[6]](#footnote-6).

The other standard construal associated with philosophical theories is that these aim to characterize “what causation is” or to capture its underlying “nature”, as distinct from our concept of it. A sharp version of this distinction (between concept and underlying nature) only makes sense if how we think about or conceptualize causation can come apart very sharply from what the causal relation “really is”, with the latter being revealed by some science (e.g., physics) distinct from common sense reasoning or perhaps revealed by some form of metaphysical insight. (An analogy would be the distinction between how we think about gold, as reflected in its stereotype as a valuable, yellowish metal and what gold really is—an element with atomic number 79.) Whatever one thinks of this possibility of uncovering the underlying nature of causation, it involves a project to which the cognitive psychology of causal reasoning as well as the methods of consulting intuitions about particular cases which are common in the philosophical literature appear to be largely irrelevant (or if relevant at all, only in a negative way)[[7]](#footnote-7)—instead it will be physics or metaphysics that tell us what causation really is. I will accordingly ignore this project in what follows[[8]](#footnote-8).

Thinking about philosophical theories of causation as (in part) normative proposals fits naturally with another important idea which is commonly assumed in the psychological literature and which I think ought to guide philosophical inquiry as well. This is that human causal reasoning is often very successful in enabling us to cope with and get around in the world. Different theories of causation can be associated with different views about the coping abilities that causal reasoning provides—as discussed in more detail below, for interventionists (cf. Woodward, 2003) these distinctively have to do with manipulation and control[[9]](#footnote-9). As an empirical matter, our practices of causal cognition are relatively “rational” or well-adapted to the circumstances in which we find ourselves—something that becomes particularly salient when we compare the abilities of humans (even young children) with other primates. (See, for example, Woodward, 2007.) Just as one of the goals of a theory of the operation of the human visual system should be to understand how it is that this system can successfully extract from the visual array reliable enough information about the structure of the external environment, so also theories of causation and causal reasoning should give us some insight into how we are able to successfully learn about and reason concerning the causal structure of our environment. As I note below, this is one reason (among many) why the most successful descriptive investigations into causal cognition are often those in which some normative theory plays an important role—one of the things we want from a good descriptive theory is to explain how we are successful to the extent that we are and this requires normative theorizing (as well as a conception of what success consists in). This focus on explaining success provides one of a number of points at which the normative and the descriptive elements in theorizing can fruitfully come together.

An emphasis on explaining success and well-adaptedness in connection with causal cognition also points to an additional limitation of the idea that the goal of theories of causation should just be to characterize our concept of causation: even putting aside the misgivings expressed earlier what we want to understand is not just what concepts we have but why those concepts work or lead to successful reasoning and inference to the extent that they do and, along with this, what their limitations might be. When limitations are present, we also ought to consider how our current ways of reasoning about causation (including our “concepts”) might be improved so that they are better adapted to our goals, an enterprise that may involve rethinking or re-engineering those ways of reasoning. By itself a description of our current concepts does not accomplish this.

I intend these claims about the adaptiveness or rationality of much common sense causal reasoning as a high-level empirical claim—one that I think is supported by a great deal of current psychological research (see, e.g., Holyoak and Cheng, 2011). A detailed defense of this claim is beyond the scope of this essay but some brief clarificatory remarks may be helpful. First, what about various well-known results that seem to show that humans are prone to all sorts of errors when reasoning about probability and other matters? Some of these results seem to me to be infected with the very same methodological problems that often infect philosophical accounts that trade heavily on appeals to intuitions as well as survey-style X-phi research—e.g., failures to control for ambiguities in the question the philosopher (or psychologist) poses to herself or others, with the result that subjects look irrational when they are simply answering a different question from what the researcher intends. Some of these methodological problems are discussed below. Second, there is a crucial distinction between people’s explicit reasoning capacities about probabilities and other matters when presented with verbally described problems and their abilities to make use of probabilistic and other sorts of information when this is presented in formats and contexts which do not require such explicit reasoning. To a very substantial extent people are better at tasks of the latter sort than the former. Many non-human animals as well as humans are, as an empirical matter, very good at tracking frequency and contingency information when this presented in an ecologically natural way and adjusting their behavior in the light of this in ways that are rational, given their goals. Brains appear to do lots of Bayesian updating even though many humans are unable to successfully reason explicitly with Bayes theorem. Focusing on people’s explicit reasoning abilities in responding to verbal questions underestimates the rationality of much that they do[[10]](#footnote-10).

Finally let me make explicit another assumption that goes along with these claims about adaptiveness. I focus in this essay on aspects of causal cognition among ordinary, lay subjects. I think, however, (this is another empirical claim) that there is a great deal of continuity between such ordinary causal cognition when successful and more sophisticated forms of causal reasoning found in the sciences[[11]](#footnote-11). Consider the assumptions about the invariance of causal relationships that underlie Cheng’s causal power theory, discussed in section 7. Very similar assumptions play an important role in causal and explanatory reasoning in many areas of science (Woodward, 2010, 2018). This is one of several reasons why we can learn things that are relevant to science and philosophy of science by studying common-sense causal reasoning.

**3. Sources of Normativity**.

I’ve been talking so far of philosophical and other theories of causation as normative proposals. What is the source of this normativity and how might we assess whether one proposal is normatively superior to another? In this essay, I will treat this as entirely a matter of means/ends justification. Inquirers have certain ends or goals, including epistemic goals—achievement of these goals is what constitutes success. Proposals about causation – whether these are proposals about how to infer to causal conclusions from certain kinds of data or which patterns of causal reasoning are correct, or how we should conceptualize or think about causation or make distinctions among different sorts of causal relationships—are to be evaluated in terms of whether or not they are effective means to these goals. In principle this approach to justification might be associated with a number of different goals. As noted above, interventionists think that one of the distinctive goals associated with causal thinking is the discovery of relationships that are exploitable for purposes of manipulation and control but (as far as this approach to justification goes) one might instead associate causal reasoning with other sorts of goals—for example, with the simple and non-redundant representation of information about regularities. In any case, various causal concepts and strategies for causal inference are to be evaluated in terms of how well they conduce to this goal. For example, a number of arguments show that randomized experiments are a particularly good way of identifying relationships that support manipulation. Possible procedures for inferring causal conclusions from non-experimental data can then be evaluated in terms of the extent to which they yield information that is like the manipulation –supporting information that would result from a properly conducted experiment—this is the generally accepted rationale for employing techniques like instrumental variables and regression discontinuity designs (See, e.g, Angrist and Pischke, 2009). This amounts to a means/ends justification for adoption of these techniques.

As another illustration, given our general interests in identifying relationships relevant to manipulation and control, we have a more specific interest in formulating manipulation-supporting relationships that generalize successfully to contexts that are different from those in which they were originally discovered. We thus value the discovery of causal relationships that are relatively invariant (where this means, among other considerations, that they will continue to hold under changes in background circumstances-- see below) and it makes sense that we adopt ways of thinking about causation that reflect invariance-linked features, such as Cheng’s causal power concept (See Sections 7- 8)[[12]](#footnote-12).

Other illustrations of the same basic means/ends justificatory strategy, not necessarily connected to causal inference, are provided by classical statistics. For example, the choice of an estimator for some quantity of interest proceeds by postulating certain goals or criteria that the estimator should satisfy—e.g., that it should be unbiased, and (among the class of unbiased estimators) have minimum variance. Justifying the choice of an estimator is then just a matter of identifying the estimator that best achieves these goals. As this example illustrates, showing that some concept or set of features is well-adapted to some goal often involves mathematical or conceptual analysis, although it may also have an empirical component. Again, there is a relatively close analogy with understanding the visual system, where this is taken to involve ascription of goals to that system (accurate enough representation of aspects of the environment relevant to action) and then an investigation (usually with a substantial mathematical component) of the means the system employs to achieve these goals[[13]](#footnote-13).

**4**. **Relating the Normative and Descriptive**.

The framework in 2-3 yields several ways in which descriptive and normative considerations can be related in studies of causal reasoning. One possibility is that we find that, as a matter of empirical fact, causal cognition among humans (or other subjects) exhibits feature F where F may be, e.g., the use of a causal concept with a certain structure or certain inferential or reasoning strategies involving causation. We can then ask whether feature F contributes to some goal G associated with casual reasoning (for interventionists, a goal such as manipulation or some subsidiary goal that follows from this. ). If so, we then have a partial explanation of those subjects “success” in their causal cognition to the extent that success is characterized in terms of the achievement of G: Subjects succeed because their causal cognition exhibits feature F. (The examples below involving the role of invariance considerations in causal reasoning have this character.) Note that in this reasoning a feature that is present as a descriptive matter is linked to a feature that involves a normative characterization, thus providing an is/ought connection. An additional possibility is that we may also have grounds for believing that some sort of selective process is at work that *explains* the presence of F in the sense that F is selected for *because* it leads to G—thus accounting for the presence of F via a functional explanation. Such selective processes might involve either natural selection or learning—e.g., subjects learn a certain way of thinking about causal relationships (rather than some alternative) through some feedback process involving the satisfaction of their goals with the feedback reinforcing that way of thinking.

In following either of the approaches described above, we identify features F of causal reasoning through empirical investigation and then ask, concerning them, whether they have a normative rationale in terms of contributing to goal G. I emphasize that the features F are *not* regarded as normatively justified or appropriate merely because they are found in people’s causal reasoning. Instead, to the extent that they are normatively justified this is because they can be given an independent justification in terms of contributing to G. I will add, though, that the empirical discovery that causal reasoning exhibits feature F can prompt us to consider the possibility that F may have some normative justification, where we might not think to consider this possibility prior to the discovery of F. For example, the tendency of subjects to assign higher causal strength ratings to contingencies in which the base rate of the effect is high, commonly thought to lack a rational justification, can be given a normative justification in terms of the invariance-linked ideas developed in Cheng’s causal power theory, as explained in Section 8 below.

Moreover, the direction of discovery can go, so to speak, in the opposite direction: given a normative theory which tells us that normatively good causal reasoning will contain feature F, we can then empirically investigate the reasoning of adult humans and other subjects to see if they exhibit this feature. In a number of cases, the answer to this question turns out to be “yes”[[14]](#footnote-14). Such cases illustrate what I will call the *motivating* or *enabling* role of normative theory in connection with empirical investigation; often it does not occur to researchers to do certain experiments or to look for whether certain features are present in causal cognition in the absence of a normative theory that tells us that normatively appropriate cognition will exhibit these features. This is another way in which normative and descriptive investigations can fruitfully interact[[15]](#footnote-15).

**5. Intuition and X-Phi**

The picture just sketched contrasts both with the usual forms of the intuition-based approaches to causation common in portions of the philosophical literature and some approaches that are common in X-phi. I begin with the former which I see as proceeding as follows: the philosopher describes various cases and then consults his or her “intuitions” about these, with the operative question usually being whether these are instances of causation. For example, a case in which a gardener fails to water flowers and the flowers die will be described and then an intuition or judgment will be advanced about whether the gardener’s omission caused the flowers to die. Similarly for cases involving double prevention, causal pre-emption and so on. The task of constructing a normative theory of causation is then taken to be, at least in part, the task of constructing a theory which captures or reproduces these intuitions—i.e., one part of the test for whether one has produce a good theory is whether it agrees with intuition. An obvious initial problem is that the intuitions of any single person are unlikely to be fully consistent with one another and, moreover, different people may hold inconsistent intuitions. In the absence of a single theory that fully captures everyone’s intuitions, a standard response is to appeal to some version of reflective equilibrium— one looks for a theory that captures as many intuitions as possible, but where this may involve rejecting some intuitions in favor of others in order to maximize overall systematic coherence[[16]](#footnote-16). Additional constraints/desiderata that are motivated on more general philosophical grounds such as the demand that the resulting theory be ‘reductive” or that it reflect certain requirements allegedly coming from “fundamental physics” or from “metaphysics” may be added to this mix—a prominent recent example is Paul and Hall, 2013, and a similar program seems to underlie Lewis’ well-known work on causation.

A general problem for any approach that takes normative justification to involve appeal to reflective equilibrium is that there is no reason to suppose that there is a unique outcome which represents the best possible trade-off among the different intuitions and other desiderata (and no obvious way of telling whether we have found such an outcome, supposing that it exists). Instead there may be multiple equilibria, each corresponding to different ways of trading off or balancing the desiderata just described. Or there may be no equilibrium. Indeed, the variety of different theories that have been produced by investigators claiming to follow something like (some version of) the method of reflective equilibrium seems to support the conclusion that either there are multiple equilibria or that all but one of the theorists have misapplied the procedure (and we can’t tell which). The approach to normative justification that I favor does not suffer from these difficulties since it takes the basis for such justification to involve means/ends reasoning rather than appeal to reflective equilibrium[[17]](#footnote-17).

But even putting this aside there are several additional difficulties. One is that it is unclear what the goal or point of the enterprise just described is. It can’t be intended just as empirical psychology or description/explanation of aspects of causal judgment since even if we accept that intuition is a source of information about these (on this see immediately below), the project under discussion involves rejecting some of these intuitions in favor of others, settling cases in which intuition is uncertain in definite ways (rather than just reporting the uncertainty) and subjecting the whole investigation to additional philosophical or physics-based constraints that are not motivated by empirical psychology. A straightforwardly descriptive enterprise (with intuitions just taken to be reports or descriptions of how people judge) would not take this form. But at the same time, the account that will emerge from such a procedure will not have an obvious functional or normative rationale (in the sense described in Sections 2-3) either. One way of seeing this is simply to note that the reflective equilibrium procedure as described assigns no role to what the end product is to be used for or whether it is well or poorly designed to achieve goals associated with causal thinking. Instead, the product looks like a curious hybrid, partly constrained by the goal of capturing intuitions and partly constrained by other sorts of considerations coming from philosophy or the metaphysics of science which seem to have little to do with functional/normative considerations. The alternative approach to normative theory and its connection to empirical psychology described in Section 2 does not suffer from these limitations.

So far I have not addressed the question of what “intuition” itself can tell us (its “reliability” or what kind of information, if any, it provides) and how this consideration impacts assessment of traditional intuition-driven approaches. (Recent discussion of the role of intuition in philosophy has tended to focus on this question.) One possibility is that intuition involves a kind of rationalistic grasp of facts about the subject of the intuition —our intuitions about causation provide us with some sort of purely reason-based acquaintance with the nature of causation and so on. It is hard to construct a plausible story about how such a “faculty” could be a reliable source of information and I will not consider this possibility further.

Another, prima facie more plausible possibility, defended by Goldman (2007) among others, takes claims about intuitions to reflect (at least in the right circumstances) claims about concepts that ordinary speakers acquire as part of their linguistic competence. Thus intuitions about examples involving causation reflect the speaker’s knowledge of how causal concepts are correctly applied and hence information about the structure of those concepts One limitation of this view is that, for reasons described above, it assigns too central a role to concept talk in causal cognition, and in effect takes causal cognition to be just a matter of applying causal concepts to scenarios, thus neglecting many other aspects of causal cognition having to do with causal learning and inference. However, we might broaden Goldman’s suggestion in the following way: when a philosopher reports the intuition that such and such is a case of C (e.g. causation) , that person at least provides evidence about how she judges. If, as is very often the case, she believes or assumes that her intuition is fairly widely shared (as might be indicated by her use of words like “we think” or “we would say”) and if there is reason to assume that this assumption is accurate, then the report of the intuition can in principle provide us with evidence about shared practices of judgment . To the extent such claims are correct (a non-trivial condition of course), we can use them just as we use other sorts of empirical evidence, employing normative theory to assess their appropriateness in the manner described in Section 2[[18]](#footnote-18). Moreover, claims about the extent to which such judgments are widely shared can in principle be checked empirically, by survey-type methods that are commonly employed in X-phi.

Of course, the content of any particular intuitive judgment that, e. g., the relationship in some scenario involving double prevention is causal will be “about” that scenario and the nature of the relationship described in it. However, I assume, in virtue of earlier arguments, that we cannot understand the mere having of such judgments as evidence for the truth or correctness of what is asserted in the judgment (e.g. as evidence that the relationship in question is “really causal” (whatever that might mean) or as evidence about the “nature” of causation. So if the intuitive judgment is evidence for anything it must be evidence for something else. Goldman takes the judgments to be evidence about the structure of our concepts (even though the content of the judgments does not directly have to do with our concepts); I have suggested broadening this to take the judgments to be (in some cases) evidence for or as implying claims about shared practices. In other words, I suggest that insofar as appeals to intuition or judgments about cases has a *legitimate* evidentiary role, it is this[[19]](#footnote-19).

If we think of intuitions in this way, we have a defense of their role in philosophical and other sorts of argument—a defense that is thoroughly naturalistic in spirit. This is because information about shared practices can, as explained above, play a legitimate role in philosophical argument, especially when combined with other assumptions. However, this defense is limited in important respects—some obvious, some less so. One limitation, already noted, is that to the extent intuitions just report judgments about cases, they are unlikely to be a useful source of information about other aspects of causal cognition such as learning. In addition, such judgments often do not provide reliable information about the causal processes that underlie them. (Section 6.) Another limitation, which intuitions share with the use of verbal responses to verbal probes more generally (whether these are responses to surveys or judgments in experimental situations as is the case with a number of the experiments discussed below) is that the question the intuiter poses to herself may be unclear or ambiguous in ways that are not appreciated, either by the intutier or her audience. (Section 8).

It should be clear that this limited defense of appeals to intuition does not support a number of the uses to which such appeals have been put in the philosophical literature. First, note that on this conception, the mere having of an intuition, even if this is widely shared, does *not* by itself have implications for what the normatively best account of causation or causal reasoning is (or, as I have argued, what the nature of causation is). What appeals to intuition show (at most) is that people make certain judgments – as argued above, whether these judgments are normatively appropriate requires appeal to an independently grounded normative theory[[20]](#footnote-20). This stands in contrast to the tendency of many philosophers to think of intuitions about cases as prima facie self- warranting in the sense that the intuition itself provides evidence of the normative correctness (or “truth”) of what is reported in the intuition. Examples discussed below suggest that this picture of the role of intuition is naïve: for example, a number of subjects report judgments of causal strength regarding certain scenarios that are arguably normatively mistaken, although an approximately equal number also report inconsistent judgments that are normatively correct. If we think of these judgments as reports of intuitions, it is clear that both sets of intuitions cannot be correct and that some external standard for correctness is required. By the same token, however, we also see that the fact that an intuition or judgment is not universally held does not by itself show that it is “wrong” – another point to which I will return below.

Next, I turn to yet another under-appreciated limitation on appeals to intuition about cases. This is that although such appeals may in principle provide information about what people’s judgments *are*, they usually (or at least often) are *not* reliable sources of information about *why* people judge as they do or about the factors on which their judgments depend or about the representations and computations that underlie these judgments.

**6. Intuitions, Surveys and Causal Inference.**

To develop these points, I want to introduce some additional analytical distinctions. I begin, however, with some remarks about what I am *not* claiming. I distinguish below between bare surveys, which record what judgments people make and causal analysis, which attempts to uncover the factors that cause their judgments and the psychological processes that underlie these. A bare survey, however well conducted, does not by itself provide a causal analysis. This does not mean that surveys are unimportant or uninteresting—on the contrary[[21]](#footnote-21). I take some of the research conducted in X-phi to be survey-like—I give examples below. Moreover , as already intimated, many traditional armchair appeals to “intuition” can be thought of as like surveys in some relevant respects, but with a very small N. However, I do not claim—and it is not true—that all research in X-phi is survey-like; some significant portion of it aims at and succeeds in providing causal analysis—again I give examples below.

**6.1)** **Bare Surveys**. Suppose that a researcher exposes a group of subjects to a scenario or case description or a group of these and then asks for judgments about them. For example, subjects might be presented with a scenario in which a gardener fails to water plants and the plants die, with the subjects being asked whether the gardener’s omission caused these deaths. Or subjects might be presented with a Gettier type scenario or a set of these and asked whether person described in the scenario “knows” that such and such is the case, as in some of the research reported in Turri (2016) . If this is all that is done, I will call this procedure and the results it produces a *bare* *survey.* Of course a bare survey can be well or badly designed— examples of the latter occur when the questions employed in the survey are unclear or confusing, when subject responses are influenced by their expectations about what the experimenters are looking for, when there are order effects or when the subjects to whom the survey is given are unrepresentative of the population to which one wants to generalize. At least sometimes such methodological problems can be adequately addressed by improving the survey design— surveys are not intrinsically flawed just because they are surveys. However, although when methodological problems are adequately addressed, a properly designed bare survey can provide information about what people in the population of interest judge, there are many other questions it cannot answer—in particular a bare survey cannot tell us why subjects judge as they do (what causes them to judge as they do or what processing underlies their judgment—see 6.3 below) .

I suggested above that reports of intuitions are most charitably construed as implying claims about shared practices of judgment. In this respect, they resemble claims supported by bare surveys, although (of course) with the difference that, in the case of intuition, the response(s) may be those of a single person who is identical with the researcher or alternatively may be the responses of a small set of colleagues. This seems to fit many of the examples of reports of intuitive judgment in the causation literature. For example, Dowe (2001) considers a number of examples of alleged causation by absence and reports that these do not seem to him to be genuine cases of causation or at least that they seem “different” than more standard cases of causation which he thinks involves transfer of energy and momentum. By contrast, Schaffer (2000) describes a number of examples of double prevention involving mechanical devices and reports that his judgments are that these are straightforward cases of causation. Lewis (1986) considers cases of symmetric overdetermination (as when two riflemen shoot a victim, with either shot being sufficient for death) and claims that common sense delivers no clear verdict about whether either of the individual shots causes the death.

We may think of each of these authors as conducting surveys of their own responses to cases, which they all seem to expect their readers to share. To the extent that readers share their responses, this is information about what “people” or some substantial number of them think about the causal status of omissions, causation by disconnection and symmetric overdetermination. Of course, similar (and perhaps more reliable) information might be obtained by means of a conventional survey of judgments about hypothetical scenarios conducted over, e.g., Mechanical Turk. To the extent that we are interested in undermining or supporting claims about what most people would say or what the folk think, survey- style results can be very valuable. [[22]](#footnote-22).

**6.2)** **Surveys with Covariation**. In a somewhat more ambitious undertaking, the researcher might explore how the judgments of some group of subjects co-vary with some other variable. Unlike a bare survey (which requires recording only subject responses) , this *requires* that some additional variable besides those responses be measured and that there be variation in both the responses and the additional variable. Speaking generically, we may distinguish (at least) two possibilities. First, different subjects (or the same subject on different occasions) may be presented with different scenarios (scenarios whose content varies along some dimension) and differences in subject judgments recorded with the aim of determining whether there is covariation between differences in the content of the scenarios and the judgments. For example, oversimplifying somewhat, Walsh and Sloman (2011) presented subjects with two scenarios one of which involved a double prevention relation and the other of which involved a connecting process ( a marble knocks over a domino). Subjects were asked whether the relationships in the scenarios were causal. A higher percentage judged that causation was present in the second scenario than in the first. Here there is covariation between the content of the scenarios and subjects’ judgments.

A second possibility is to present subjects with a single scenario about which different subjects judge differently. The researcher then determines whether there is co-variation between these judgments and other variables, which may include demographic factors. For example, Machery et al. (2004) presented “Western” and “Asian” subjects with scenarios such as Kripke’s Godel case, asking them to make judgments about the referents of names. These authors claim that the Asian subjects were more inclined to make “descriptivist” judgments about reference than Westerners.

Such surveys with covariation can be valuable for a number of reasons. Most obviously, they can show that judgments that were assumed to be universal or nearly so, vary considerably across subjects, including different demographic groups-- thus casting doubt on claims about “widely shared intuitions”. Moreover, to the extent that such judgments not only vary but covary with factors such as cultural background that we think are irrelevant to whether the judgments are “true” or “correct”, this seems to further undermine a number of the standard uses to which traditional philosophers have attempted to put them. On the other hand, the mere observation of a covariation between subject responses and some other variable (even if the covariation is genuine in the sense that it holds in some target population to which one wants to generalize) does not by itself establish that the second variable causes the variation in responses. This is the goal of causal analysis.

**6.3)** **Causal Analysis**. In causal analysis the goal is not just to describe subject responses or judgments or the variation that may occur in these but to discover the cause or causes of such variation and/or the processing or mediating variables that underlie it. I follow conventional ideas about causal inference in holding that in order to do this successfully one must rule out the possibility that other factors besides the candidate cause are responsible for the variation in question— as it is often put, one must rule out or control for possible “confounding” factors. Confounding is present when there is covariation between subject responses R and some other factor X but this covariation does not arise because X causes R but rather is due to the operation of other factors—e.g., a common cause of X and R. (Note that this is a different problem than the methodological problems that arise in connection with survey design such as non-representativeness and experimenter effects.)

In principle there are two different possible ways of controlling for confounders—experimentation and causal modeling. (These may also be used in combination.) The crucial feature of an *experiment*, as I shall use this term, is that there is active manipulation of the putative causal factor in a way that makes it independent of other possible causal factors that may influence the effect. When there is covariation between such an independently manipulated candidate cause and the effect, this is taken to show that the candidate cause is a genuine cause. In an experiment, such manipulation may be accomplished by randomization or by independent control of these alternative causal factors, assuming that they are known. As an illustration, Vasilveya et al. (Forthcoming) explored whether differences in the perceived background invariance or insensitivity of causal claims (the extent to which those claims continue to hold under variation in background circumstances) cause differences in judgments of the causal strength of those claims. Different causal claims can differ along many different dimensions and to establish that differences in the insensitivity of these claims were responsible for differences in subject’s judgments, Vasilveya et al. needed to rule out the possibility that such other differences were causally responsible for the difference in judgment. For example, they needed to rule out the possibility that differences in  *∆p* = (the probability of the effect conditional on the presence of the cause minus the probability of the effect conditional on the absence of the cause—see below) associated with the causal claims employed caused the difference in subject judgments. They accomplished this by using causal claims that were matched for *∆p* but differed in background invariance. Such careful thinking about what might be a confounding factor and taking steps to control for is essential for reliable causal inference.

An alternative strategy for causal analysis involves causal modeling of the factors that influence subject responses, where this may but need not make use of experimental manipulation. Such modeling requires measurement of possible confounders and other variables and the application of some causal inference procedure (there are many candidates for these—constraint-based procedures such as those described in Spirtes et al. (2000), Bayesian analyses (e.g. Tenebaum et al. 2006), and various structural equation methods, among others). Again, in all of these methods the goal is to show that subject responses R are caused by some factor X by ruling out competing explanations for the covariation between X and R. Causal modeling can of course be carried out on purely “observational” data in which there is no experimental manipulation, so experimentation in the sense described above is not required for reliable causal analysis, although in empirical psychology it is typical to rely at least in part on experiments[[23]](#footnote-23).

To illustrate the difference between successful causal analysis and a survey with covariation, return to the research conducted by Walsh and Sloman. These authors show (let us assume) that there is covariation between subject judgments of causation and whether the presented scenarios involved (i) a connecting causal process or (ii) dependence without such a process. Walsh and Sloman infer from this that this difference between (i) and (ii) is what caused this difference in causal judgment. In my view, they are not entitled to infer this on the basis of the data generated in their experiment. The problem is that they have not ruled out the possibility that some other difference between their two scenarios influences the difference in causal judgments[[24]](#footnote-24). For example, their disconnection scenario differed from the scenario involving a connecting process not just in terms of whether such a process was present but also in terms of the relative invariance of the dependence relations present in the two scenarios. If subject judgments are influenced by such differences in invariance (with less invariant relations judged as less causal or non-causal ) it could be this difference which is responsible for the difference in judgment. In fact, Vasilyeva et al, Forthcoming, as well as Lombrozo 2010 provide experimental evidence that this is the case. A similar analysis may apply to examples of causation by absence, as suggested in Woodward, 2006. Dowe, for example, notes that he and others judge that some cases in which there is a relation of dependence between an absence and an outcome are non-causal and infers that he and others make this judgment *because* there is an absence of a connecting process in the examples considers. Even supposing (as seems plausible) that he is correct about how most people would judge regarding his examples, it does not follow that he is correct about what features of his scenarios cause people to make these judgments. To show this one must control for other factors besides the absence of a connecting process that may drive judgments in causation by absence scenarios. Examples like this illustrate that even if one is a good judge of how others will judge regarding various scenarios, it is easy to be misled about why oneself and others judge as they do—introspection often is not a good guide to this and more rigorous causal analysis is required instead. It is thus important that the psychological research described in Section 8 involves genuine experiments that support causal analysis not just surveys or surveys with co-variation.

Despite this, it is fairly common to find philosophers not just reporting their intuitive responses (and claiming that others will have similar responses) but also either explicitly claiming that they can tell which factors are causally influencing those responses or at least writing as though they have reliable introspective access to information about this. Sometimes the argument goes like this: the philosopher finds two cases that (it is supposed) match exactly in all possibly relevant respects except for the presence or absence of a single feature X. The philosopher takes this to be a case in which she is controlling “in her mind” for all the other relevant differences between the two cases besides X—i. e., the philosopher thinks of herself as running a controlled experiment, albeit in her mind. The philosopher finds that her intuitive judgments about the two cases differ and attributes this to the difference made by X.. (See, for example, the discussion in Kamm, 1993, who explicitly endorses this method.) The obvious problem with this procedure is that it requires that the philosopher has introspective access to all of the other factors besides X that might influence her differential response and also that she can recognize when one of these is present and influencing her response and somehow remove or correct for the factor in question. I don’t mean to claim that people can *never* do this but I’m dubious that philosophers or anyone else can reliably execute such mentalistic analogues to an actual experiment in many cases of philosophical interest and that, moreover, they can be in a position to *know* that they have successfully done this. Indeed, both in the literature on causal cognition and elsewhere, there are many experiments that show that what is actually influencing peoples differential responses is not what they judge to be influencing them. To this we may add that if people really did have reliable introspective access to whatever influences their judgments, experimental psychological investigations into this would be unnecessary—everything could be done from the armchair.

Let me add that in distinguishing between surveys (with or without covariation) and causal analysis, I certainly do not mean to claim that experimental philosophers only do the former. In the field of causal cognition alone, examples of (in my judgment) convincing causal analysis carried out by experimental philosophers include Hitchcock and Knobe (2009), showing the influence of norms on causal judgment, Kominsky et al. (2015), showing factors influencing actual cause judgment and competition between causes, and Icard et al. (forthcoming) showing the influence of normality judgments on actual cause judgments.

In discussions of X-phi, both pro and con, there is a tendency to focus on the difference between traditional armchair methods and surveys of what the folk think, with some experimental philosophes arguing for the superiority of the latter and more traditional philosophers rejecting such claims. However significant this difference may be, there is also an important divide that puts surveys and appeals to intuitions on one side and causal analysis, whether carried out by experimental philosophers or by psychologsts, on the other. In my view at least sometimes a philosopher in the armchair may have a good sense about how others will judge, although of course it will always be an empirical issue to what extent this is the case. However, even if armchair methods are sometimes reliable in this application, it is a further question whether they can be used reliably in causal analysis—and here the answer seems to be negative at least in many cases. If this is correct, the most innovative forms of X-phi are those that involve causal analysis—in contrast to surveys, they address questions for which armchair analysis seems particularly unsuited.

**7. Invariance: Normative Theory and Descriptive Results**

I turn now to a discussion of some more specific psychological hypotheses about causal cognition and associated experimental results. I will try to show how these illustrate the general methodological ideas described above. My focus will be on hypotheses and results having to do with the role of *invariance* in causal cognition.

The general idea that motivates this research has both a descriptive and a normative component. The descriptive component is that we tend to think and reason about causal relationships in terms of invariance and that, other things being equal, we prefer, when we can discover them, causal relationships that are more invariant rather than less. The normative element is that it is correct or appropriate to reason in this way since relatively invariant relationships better satisfy goals associated with causal reasoning. The general idea of invariance is that a relationship *C 🡪 E* is more invariant to the extent that it would continue to hold as various other factors change—“continue to hold” means that the relationship continues to apply or to correctly describe what is going on. These “other factors” come in a variety of different forms, corresponding to different aspects of invariance. For example, we can ask whether the *C*🡪E relationship continues to hold if various other factors distinct from C and *E*, (“background factors”) *,* change[[25]](#footnote-25). Another aspect of invariance has to do with whether the *C🡪 E* relationship would continue to hold under changes in the values taken by C or changes in the frequency with which those values occur. The research described below makes use of both of these aspects of invariance.

The normative appeal of invariance should be obvious: to the extent that a relationship is more invariant we can export or generalize or apply it to a range of different situations[[26]](#footnote-26). If causal cognition is well-adapted to the achievement of goals having to do with generalizability, we would expect, as a descriptive matter, that it reflects the influence of invariance-related considerations. This provides a motive for looking empirically at whether human causal cognition reflects such influences[[27]](#footnote-27).

**8. Cheng’s Causal Power Model**.

Cheng’s Causal Power model (Cheng, 1997) makes use of a number of invariance assumptions and is intended both as a descriptive account of how subjects make causal judgments and which judgments they make but it also has a normative motivation—it is also intended as an account of how people ought to reason, thus illustrating our general theme of the interrelation between the normative and descriptive in understanding causal reasoning. It attempts to capture the intuitive idea that causes have “causal powers” that they “carry around with them” in different contexts. Cheng’s model represents causes and effects as binary events, which can either be present or absent. Causes can be either “generative’—they can promote their effects or they can be “preventive”, interfering with the operation of generative causes. The “power” of a generative cause *i* to cause effect *e* is represented by *pi*, the probability with which *i* causes *e* if *i* is present. Note that this is *not* the same as *P(e/i*)— among other considerations, the latter quantity reflects the influence of other causes of *e* that are present when *i* is[[28]](#footnote-28). Let *a* represent all such other causes of *e*, and assume that when present they produce *e* with probability *pa* and that these are all generative rather than preventive causes of *e.* Assume also that *e* does not occur when it is not caused by either *i* or *a*. In a typical experiment, subjects have access to data about the frequencies of occurrence of *i* and *e* (e.g., in the form of a contingency table) but do not directly observe either the occurrence of *a* or *pi* and *pa* —these have to be inferred, to the extent that they can be. Cheng makes the following two additional assumptions about *i* and *a*:

8.1. *i* and *a* influence the occurrence of *e* independently

8.2. The causal powers with which *i* and *a* influence the occurrence of *e* are independent of the probability with which *i* and *a* occur so that, e.g. the probability that *i* occurs and causes *e* is just *P (i*). *pi*.

Both (8.1) and (8.2) are invariance assumptions. (8.1) says that *pi* is invariant under changes in *pa* and similarly *pa* is invariant under changes in *pi* . (8.2) says that *pi* and *pa* are invariant across changes in the probability with which *i* and *a* occur. Cheng thinks of (8.1) and (8.2) as “default” assumptions people bring to situations involving causal learning and judgment. I will return to the status of these below but the basic idea is that although nothing guarantees that such assumptions will be true in the situation of interest, they are nonetheless useful points of departure for reasoning which can be relaxed as the empirical evidence warrants[[29]](#footnote-29).

Given these assumptions, causal power can be represented and (in the appropriate circumstances) estimated in the following way [[30]](#footnote-30): First, *P(e)* is given by the union of the probability that *i* occurs and causes *e* and that *a* occurs and causes *e*:

*P(e)* = *P(i*). *pi* + *P(a). pa* - *P(i).* *pi*. *P(a)*. *pa*

Conditionalizing on the presence of *i* we obtain

*P(e/i*)= *pi* + *P(a/i).* *pa* - *pi*. *P(a/i). pa*

Conditionalizing on the absence of *i* we obtain

*P(e/not i)= P(a/not i).* *pa*

Defining (8.3) *∆p(i)= P(e/i)- P(e/not i),* it follows that

*∆p (i*)= *pi* + *P(a/i). pa* – *pi*.*P(a/i).* *pa*

- *P(a|not i). pa*

= *[1 - P(a/i). pa]. pi* + *[(P(a/i) - P(a/ not i)]. pa*

Thus

*∆p(i)- [P(a/i)- P(a/not i)] pa*

*pi* = ------------------------------------------

*1-P(a) pa*

Now consider the special case in which the probabilities with which *a* and  *i* occur are independent so that *(P(a/i)* = *P(a/ not i).* Then

*pi* = *∆P(i)/ 1-P(a)* *pa*.

*pa* cannot be estimated from the frequency data available but since *e* is caused either by *i* or *a*, we can replace *P(a). pa* with *P(e/not i*) yielding

(8.4) *pi* = *∆ p (i) / [1- P(e/not i)]*

Thus under the specified assumptions (8.4)is a normatively correct estimate for causal power, *pi.* [[31]](#footnote-31)

In many empirical studies (including Cheng’s), subjects are presented with frequency information in some format about the patterns of co-occurrence between a candidate cause, *c* and an effect *e* and are then asked to estimate (what is called) the “causal strength” of the relationship between *c* and *e*. Cheng’s model claims that such causal strength judgments track *pi*, causal power. As we shall see, there is disagreement about the verbal probe that is most appropriate for eliciting such judgments, but a commonly used question is some variant of “ On a seven point scale, how appropriate would it be to describe the relationship between *c* and *e* as one in which *c* causes *e?*”. Cheng’s model, as well as a number of other competing models, aim at (among other explananda) describing patterns of causal strength judgments and the representations and computations that underlie these. In many cases, what researchers aim to fit is something like average judgments across subjects—a practice that I comment on below.

One of the main alternatives to Cheng’s model is the so-called *∆p* model, according to which subject’s causal strength judgments will track the quantity (8.3) *∆p(i)= P(e/i)- P(e/not i).* (This model has roots in associative models of animal learning such as the Rescorla-Wagner model.)  As is apparent from (8.4), on the assumption that subjects’ causal strength judgment track *pi*, the predictions of Cheng’s model and the *∆p* model diverge. Although there is non-trivial disagreement about the empirical facts (see below), there is significant evidence favoring some features of the power pc model over *∆p* and other competitors.

To illustrate these diverging predictions, first consider situations in which *P(e/not i) =*1. In this situation *∆ p =* 0, since *P(e/i)=*1, assuming *i* is a generative cause, the presence of which promotes *e*. Thus subjects guided by *∆p* in their causal judgments should report that in this situation *i* does not cause *e*, assuming that a causal strength of zero corresponds to the absence of causation. By contrast the denominator of (8.4) is zero when *P (e/not i) =* 1 so that (8.4) is undefined in this circumstance. Thus subjects guided by *pi* in their strength judgments should report that they are unable to reach any conclusion about the causal strength of *i* in this situation. Note that the latter judgment rather than the judgment based on *∆p* is the normatively correct one: When *P(e/not i) =* 1 a “ceiling effect” is present—since *e* is always present the power (if any) of *i* to cause *e* cannot reveal itself in any differential probability of occurrence of *e* in the presence versus the absence of *i*. As an empirical matter, when ordinary subjects are given this option, a substantial number (but by no means all) chose this “unable to reach conclusion” alternative.

A second prediction, if subjects are guided by causal power *pi* in their strength judgments, is this: As the probability of the effect *p(e)* increases, *p(e/not i)* will increase (assuming *pa. P(a)* is not zero). Thus, for a constant *∆p,* *pi*will increase—in other words *i* will be judged a stronger cause of *e* (again given the above assumptions) the more frequent the occurrence of *e*, for a fixed *∆ p*. As an empirical matter, many subject judgments do tend to exhibit this feature but this has often been treated as an “irrational bias” of some kind—it is certainly normatively inappropriate if the correct normative theory for judgments of casual strength is given by *∆ p*. By contrast, this feature is both predicted when *pi*is manipulated for constant *∆p*  and shown to be normatively reasonable by the causal power theory[[32]](#footnote-32).

It is worth reflecting briefly on the normative differences between the causal power model and *∆p.* Intuitively speaking, *∆p* is normatively deficient as a measure of the causal strength of *i* because it does not correct for confounding—both *P(e/i)* and *P(e/not i*) reflect not just the relationship between *i* and *e*, but also the extent to which other causes of *e*, captured by *a*, are operative, even if those other causes operate completely independently of *e*. Among other limitations, this measure will not generalize appropriately to new situations in which the distribution of other causes of *e* is different than in the original situation. The invariance assumptions built into the causal power model correct for this (in effect by normalizing *∆p* to correct for other causes of *e* besides *i)*, assuming the applicability of the assumptions that go into its derivation. In fact, in other experiments, Cheng and co-authors have shown that, as an empirical matter, causal power does a much better job of predicting which causal judgments (and associated measures of strength) generalize to new situations with different distributions of new causes than alternative measures like *∆p—*again, a pattern of judgment that is normatively reasonable.

I said above that although Cheng and others have obtained results supporting the empirical predictions of the model, the overall empirical adequacy of the model remains controversial. I turn now to a description of some discordant empirical results and the response of Cheng and her collaborators to these[[33]](#footnote-33). As we shall see, this discussion has a number of interesting philosophical and methodological implications. These include problems that can arise when verbal reports are used as evidence, both in experimental and armchair contexts as well as issues having to do with subject heterogeneity. Lober and Shanks (2000) agree that the causal power theory is the correct normative theory of causal judgment in the situations that satisfy the background assumptions of Cheng’s theory. They draw attention, however, to two patterns present in human causal strength judgments that are prima –facie inconsistent with the causal power model The first consists in the fact that some subjects provide positive causal strength ratings in the presence of “non-contingency”—that is, when *∆ p*= 0, with the magnitude of these ratings being influenced by *P (e/not i).* This is inconsistent with both the causal power and *∆p* models, which predict strength ratings of zero in such cases. Second, recall the experiments of Cheng discussed above in which *∆p* is held constant, and causal power varied and causal judgments are shown to track causal power. Lober and Shanks were able to replicate this result but they also did the “opposite” experiment in which causal power is held constant across different experimental conditions and *∆ p* varied. Of course the causal power theory predicts no difference in judgment across these conditions but averaging over the experimental population, such judgments *are* found to vary, appearing to support *∆ p* over causal power. In an extremely interesting analysis of their data, Lober and Shanks show that their subjects can be separated into two groups one of which (the power group) seems to be guided by the normative considerations that led to the construction of the power pc theory (e.g., this group is aware of ceiling effects and tries to take them into account in their causal strength judgments) and the other of which (the “contingency participants”) seems not to take these considerations into account. When causal power is held constant across different values of *∆ p* the power participants behave pretty much as the power pc theory predicts— their ratings are fairly constant across different values of *∆ p*. By contrast, the ratings of the contingency group increase with increasing values of *∆ p*., which is what the *∆ p* but not the causal power model predicts[[34]](#footnote-34).

In an attempt to account for these and other results that appear to be inconsistent with the causal power model, Cheng and colleagues (Buehner et al., 2003) appeal to several considerations (as well as further experiments). One has to do with what they call “ambiguity of the causal question”. First, they note (following Tenenbaum and Griffiths, 2001) that standard verbal probes for causal strength (e.g. “how appropriate would it be to describe this as a situation in which *c* causes *e*?”) may conflate a subject’s degree of confidence that a causal relationship exists with the question of how “strong” that relationship is, given that it does exist. Second, they note a further potential ambiguity: when asked about the causal strength of *c* with respect to *e*, subjects might (i) interpret this as asking “what difference does the candidate cause make in the current learning context, in which alternative causes already produce *e* in a certain proportion of the entities” (1126). Here the question is understood as asking, “what additional difference *c* makes to *e*, given that alternative causes are already causing some instances of *e*?”. As should be obvious, *∆ p* *is* the normatively correct answer to this question. A second, alternative way of interpreting the causal strength question is (ii) “what difference does the candidate cause *c* make in a context in which alternative causes never produce *e*”, where the normatively correct answer is given by the causal power model. Buehner et al. suggest that the results for experiments with positive contingencies that appear to be inconsistent with the causal power model might be explained by the fact that approximately half of the subjects are interpreting the causal strength question along the lines of (i) and the other half along the lines of (ii) and provide an analysis of their data that supports this interpretation. They also performed a second experiment in which the causal question was altered along the lines of (ii) since (as they see it) this corresponds to the notion of causal strength that the power pc model is intended to capture. This revised causal question asked subjects to estimate “how many entities out of a group of 100 which did not show an outcome would now have the outcome in the counterfactual situation in which the candidate cause was introduced” (p. 1128). As they note, this is in effect an “intervention” question, which asks what the effect of the cause would be if it were introduced by an intervention into a situation in which it was previously absent. Asking for an estimate in terms of proportion of entities also makes it clearer that the question is not about reliability or degree of confidence that a causal relation exists. Employing this revised verbal probe for causal strength and certain other modifications in their experimental design, Buehner et al., 2003 obtain results which seem to show that the great majority of their subjects judge in accord with the causal power model.

**9. Some Philosophical Morals**

Let me now try to extract some general philosophical morals for projects having to do with the empirical study of causal reasoning from this discussion and relate them to the themes discussed above.

**9.1) The central role played by normative ideas**. Although Cheng’s model is intended to account for empirical features of human causal judgment, it is motivated by normative considerations linking causal judgment to certain assumptions about the invariance properties of causal relationships. The link between the normative and the descriptive is provided by the claim that human causal judgment and learning are to some considerable degree normatively reasonable, so that normative models predict to some significant degree how people in fact learn and judge. This also illustrates the idea of “explaining success”— reasoning about causal relationships in terms of invariance (and adopting learning strategies that lead to invariant relationships, including estimating causal power from the relationship (8.4) rather than from *∆p*), contributes to success when this is understood as the discovery of relationships that are exportable to new situations.

Normative theorizing also enters in more subtle ways—for example, it motivates various experiments and the interpretation of experimental results. One would probably not think to do Cheng’s experiments in which causal power is varied for constant *∆ p* and the effect on judgment observed in the absence of a theory motivating the causal power model[[35]](#footnote-35). Normative theorizing also plays a role in interpreting the verbal probes used in eliciting strength judgments—e.g., it is normative analysis that tells us that there is a an important distinction between, asking (i) what additional difference *c* makes to *e* in circumstances in which other causes of *e* are assumed to be present and asking (ii) what difference *c* would make to *e* if *c* were introduced in circumstances in which all other causes of *e* were absent. However, this role for normative theorizing in motivating the choice of verbal probe can introduce worries about a kind of circularity or lack of robustness which I discuss below.

I will add that this moral seems to me to generalize well beyond causal reasoning. In many cases, the most successful empirical or descriptive theories in the human sciences are those that are tied to normative theorizing—decision theory, and theories of learning and belief change in response to evidence furnish additional examples. Normative theories can structure empirical investigation, motivate experiments and help to interpret results. In philosophy of science, use of empirical data of any kind, whether it comes from experiments, surveys, case studies or other sources is likely to be most fruitful when connected to a normative theory. Moreover, the normative theory is not going to emerge just from the empirical data alone, instead some independent rationale (typically connected to means/ends patterns of justification) is required.

**9.2. The significance of intuitive judgments**. In the experiments described in Section 8, people express judgments of causal strength when presented with various stimuli. Following a traditional armchair methodology, one might be tempted to interpret these as reports of intuitions that give us some sort of veridical insight into the nature of or our concept of causation, which the philosopher should then try to systematize. One obvious problem is that, as we have seen, different subjects have different and indeed inconsistent judgments about causal strength regarding the same cases. If these subject’s judgments reflect intuitions, then at least in this case, the mere having of an intuitive judgment, no matter how firmly held, does not establish that such a judgment is correct or veridical. Nor is there any obvious reason to think that correctness of these intuitions can be established merely by systematizing them or bringing them into reflective equilibrium with one another. Instead, in the case under discussion, the normative correctness (or not) of intuitive judgments is established by an independently justified normative theory—the causal power model which in turn is supported by various invariance assumptions.

A better way of thinking about the evidential significance of the judgments obtained in Cheng’s experiments is simply that they tell us (or may tell us) something about what certain groups judge—this is certainly how Cheng understands their significance. We can then ask which models, normative or non-normative, best account for such judgments. As noted, finding certain patterns in people’s judgments may also prompt us to ask whether there may be some previously unconsidered normative rationale for those judgments but it does not by itself show that there is such a rationale. As suggested above, I think that philosopher’s appeals to intuitions should be treated in a similar way — such intuitive judgments can sometimes provide information about shared practices of judgment by others, just as Cheng’s experimental results do and we can then go on to ask whether there is some independent normative rationale for these judgments. When understood in this way, it is hard to see what grounds there are for a wholesale dismissal of appeals to intuition that would not in also be grounds for dismissal of the use of verbal judgments as a source of information in psychological experiments[[36]](#footnote-36). But it also follows that the mere having of an intuition is not normatively probative.

**9.3, What people say versus why they say it**. The research described above illustrates the important difference between reports of what people judge and the factors and processes that cause these judgments. Even if subjects accurately report their causal strength judgments, it seems clear that the representations, computations, and learning strategies that underlie those judgments (whether these are explained by the causal power model, some more associationist model incorporating *∆p* or something else) are not themselves accessible (or capable of being established) via intuition. These are instead predicted on the basis of normative analysis and mathematical modeling and then require investigations, either experimental or observational that control for confounders, for their confirmation.

**9. 4. Causal Cognition and Causal Concepts**. I noted above the tendency of philosophers following traditional intuition-based methodologies to frame their conclusions as claims about concepts (or their application), so that in the causation literature the primary conclusions such methods are taken to establish is that various scenarios do or do not fall under “our” concept of causation. As we see from the research described there are many important features of causal cognition (or if you like, causation) that are not well captured in this way. To begin with, even if one holds that whether a causal relationship is present or not in some situation always requires a binary, yes- no judgment, it is clear that people make further more graded discriminations, distinguishing among causal claims with respect to how strong they are. Moreover, although I lack space for discussion, it is clear from other experimental and analytical work that “strength” has several distinct dimensions, some of which may be captured by the causal power model and others of which are not[[37]](#footnote-37). A complete theory of causation, whether normative or descriptive, should reflect this. In addition, there are a number of other features of causal reasoning that are not naturally viewed as constitutive of causal concepts but which are nonetheless important. These include the role played by various defaults, as illustrated above. As we saw, Cheng’s model incorporates various invariance assumptions such as the assumption that the tendency of *i* to cause *e* operates independently of the tendency of alternative causes *a* to cause *e.* Obviously such assumptions can be violated—causes can interact with one another to produce effects. (Indeed, Cheng devotes several papers to modeling how people reason about such interactive causes.) Cheng’s claim is that people tend to treat such invariance assumptions as defaults—in the absence of evidence to the contrary they tend to first assume that causes operate non-interactively, only modifying their judgments and reasoning when they get evidence contradicting the original default. Such assumptions play the role of structuring inquiry (or search) in certain ways, leading inquirers to consider certain possibilities before others so that search among alternative hypotheses can proceed in an organized, systematic way. Because default assumptions can be violated and causal relations still can be present, it does not seem right to think of them as built into our concept of causation, but they still are important in causal reasoning. A similar remark applies to many other features of causal cognition. Ranging further afield, my guess is that philosophers of science interested in doing empirical work on “confirmation”, “evidence” and related notions would also do well to focus less on what belongs to these concepts and more on the role of strategies of search, default assumptions and the like. Again, there is no reason why empirical work relevant to scientific methodology should be organized around studies of concepts.

**9. 5. Potential Ambiguity of the verbal probe**. Issues surrounding the interpretation of verbal probes have important implications not just for the interpretation of experimental results but also for the role assigned to intuitive judgment in philosophical discussion. In an experimental context in which a question is posed (e.g., about causal strength or whether a casual relationship is present) and subjects are asked for a verbal response, it is obvious that one needs to worry that different subjects may interpret the question differently from one another or differently from what the experimenter intends. If so, the verbal probe may not measure what the experimenter thinks it is measuring or may not measure the same variable for all subjects.

It is natural to wonder whether the same thing sometimes may happen when philosophers elicit intuitions, either from themselves or others. Suppose Philo describes a case and reports having such and such an intuition about it. Cleanthes, reports the same or a different intuition about the same case. Each is in effect asking themselves a question: (“What is my intuition or judgment about whether this is a case of X?”) When (or how) can we be confident that they are asking themselves the same question? If Philo and Cleanthes both report their intuitions about the strength of the causal relationship present in a certain scenario (or whether a causal relationship is present at all) don’t we need to consider the possibility that they are interpreting causal strength (or “causal relationship”) differently (as the subjects in experiments described above may be.)? Indeed, going further, shouldn’t we also be concerned about the possibility that Philo himself may be unaware of possible ambiguities in the question he is asking himself—he may report his intuition about causal strength (or whatever), without recognizing that the question he asks himself may be unclear or that when he asks himself what he thinks are versions of the same question, expressed slightly differently, he is actually asking himself different questions to which different answers are appropriate[[38]](#footnote-38).

As noted above, normative analysis of verbal probes can help to make us aware of possible ambiguities and unclarities in verbal probes, including those used in eliciting intuitions[[39]](#footnote-39). In addition, as illustrated above, additional empirical work can either confirm or disconfirm the possibility that different subjects are interpreting the same probe in different ways or differently from what the researcher intends. Still potential problems remain. When a normative model of causal judgment is invoked to support the use of a particular verbal probe, and the results of that probe are then used to support the descriptive adequacy of the normative model, there is an obvious worry about question-begging. For example, the revised verbal probe employed by Buehner et al. above appears attuned to direct subjects to formulate strength judgments just on the basis of the features that the causal power model claims do drive strength judgments. This does not make the fit between the model and the elicited judgments automatic or uninteresting (in fact critics have claimed that the judgments elicited by this probe still do not fully track the predictions of the causal power model) but it does raise questions about how to assess the descriptive adequacy of a model which appears to be sensitive to the exact wording of the probe employed.

Should we be bothered by the possibility that if we were to employ a different verbal probe we would arrive at a different assessment of the adequacy of the model? At the very least, it seems that we should try to understand the relationship between different possible verbal probes and when, as an empirical matter, they elicit the same or different results and why this is the case. Going further, it would be highly desirable to combine verbal measures with measures of non-verbal behavior which can be less susceptible to concerns about misinterpretation. For example, in a causal learning task, the dependent variable might be whether the subject succeeds in activating a certain machine on the basis of presented information, with its being completely unambiguous whether this has been accomplished[[40]](#footnote-40). Many of the best designed and most persuasive experiments in causal cognition make use of non –verbal measures in part for this reason. I suggest that the same lessons can be applied in philosophy—to the extent we employ reports of intuitions about cases (or surveys) we should consider whether different ways of eliciting the intuition generate the same or different results, whether the judgments generated are consistent with what is suggested by non-verbal measures and so on [[41]](#footnote-41).

**10. Implications for X-Phi**.

I believe that much of what I have said about the role of intuitions (construed in the deflationary manner described above) transfers to survey–like X-phi investigations as well. Such surveys can be valuable in virtue of producing direct evidence bearing on how widely intuitions/judgments are shared, thus contributing to both negative, debunking programs and to more positive programs directed at the description of shared folk thinking. On the other hand, a number of the limitations of an intuition-based methodology are also potential problems for survey-style X –phi. We still have the problem that the verbal probes used in such surveys may be interpreted differently by different subjects or may contain unnoticed ambiguities. This is not in principle an insurmountable problem (just as it is not for reports of intuitions) and it is receiving more attention recently from experimental philosophers[[42]](#footnote-42). Nonetheless it seems uncontroversial that experimental philosophers doing survey-style work should employ different verbal probes and try to understand the relationships between the results they produce. They should also try to understand what verbal responses tell us about non-verbal behavior and practices. In addition, as is the case with appeals to intuition, surveys by themselves will not tell us about the underlying causes of survey responses.

Finally there is another feature of survey-style research (and for that matter, traditional appeals to intuition) that deserves brief mention. There is some tendency in this research to focus on the question of whether responses are very widely or nearly universally shared. On the one hand, this is a very natural and appropriate question when a philosopher claims, in an unqualified way, that “people judge that so and so..”. On the other hand, to the extent that our interest is in normative theory and explaining success, discoveries that judgments and other practices are non-universal may be less consequential than is sometimes supposed. The fact (if it is a fact) that not all subjects judge in accord with the causal power model does not in itself undermine the normative status of that model or the normative appeal of invariance- based ideas. It is also consistent with the model providing an adequate explanation of the judgments of those subjects who do conform to the model and explaining why their causal cognition succeeds in certain respects. (Of course some different explanation will be needed for those subjects who do not conform to the model.) To the extent that X-phi results are employed in a negative or debunking role (showing that many folk don’t judge as philosophers claim), such results may not matter so much for projects of the sort outlined above.

**References**

Angrist, J. and Pischke, S. (2009) *Mostly Harmless Econometrics*. Princeton: Princeton University Press**.**

Buehner, M., Cheng, P., Clifford, D. (2003.) [From covariation to causation: A test of the assumption of causal power.](http://reasoninglab.psych.ucla.edu/CHENG%20pdfs/Buehner_et_al.2003.pdf) *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 29, 1119-1140.

Cheng, P. (1997) “From Covariation to Causation: A Causal Power Theory” *Psychological Review* 104: 367-405.

Dowe, P. *Physical Causation*. Cambridge: Cambridge University Press.

Goldman, A. (2007) “Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status”. *Grazer Philosophische Studien* 74: 1-26.

Gopnik, A., Glymour, C, Sobel, D., Schulz, L., Kusnir, T. and Danks, D. (2004) “A Theory of Causal Learning in Children” *Psychological Review* 111: 1-30.

Gopnik, A. (2009) *The Philosophical Baby*. New York: Farrar, Straus and Giroux.

Hesslow, G. (1976), "Two Notes on the Probabilistic Approach to Causality." *Philosophy of Science* 43: 290 – 92.

Hitchcock, C and Knobe, J. (2009) “Cause and Norm” *Journal of Philosophy* 106: 587- 612.

Holyoak, K. & Cheng, P. (2011). “[Causal learning and inference as a rational process: The new synthesis”.](http://reasoninglab.psych.ucla.edu/wp-content/uploads/2010/09/Holyoak_Cheng.AR_.2011.pdf) *Annual Review of Psychology*, 62, 135-163.

Icard, T., Kominsky, J. and Knobe, J. (Forthcoming) “Normality and Actual Causal Strength” *Cognition.*

Lewis, D. (1986) *Philosophical Papers, Volume 2*. Oxford: Oxford University Press

Kamm, F. (1993) *Morality, Mortality*, Volume 1. New York : Oxford University Press.

Knobe, J. (2016) “Experimental Philosophy is Cognitive Science” in Systema, J. and Buckwalter, W. (eds.) *A Companion to Experimental Philosophy.*  Wiley Blackwell, pp 37-52.

Kominsky, J., Phillips, J., Gerstenberg, T., Lagnado, D. and Knobe, J. (2015) “Causal Superseding” *Cognition* 137: 196-209.

Lober, K., & Shanks, D. (2000). “Is causal induction based on causal

power? Critique of Cheng (1997)”. *Psychological Review*, 107, 195–212.

Lombrozo, T. (2010). “[Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions”.](http://www.concepts.dreamhosters.com/wp-content/uploads/2013/11/13.pdf) *Cognitive Psychology, 61*, 303-332.

Machery, E., Mallon, R., Nichols, S. and Stich, S. (2004) “Semantics: Cross-Cultural Style” *Cognition* 92:B1-B12.

Paul, L. and Hall, N. (2013) *Causation: A User’s Guide*. Oxford: Oxford University Press.

Tenenbaum, J., Griffiths, T. and Kemp, C. “Theory-based Models of Inductive Learning and Reasoning” *Trends in Cognitive Science* 10: 309-318.

Pearl, J. (2001) *Causality*. Cambridge: Cambridge University Press.

Schaffer, J. (2000) “Causation by Disconnection” *Philosophy of Science* 67: 285-300.

Sobel, D., Tenenbaum, J. and Gopnik, A. (2004) “Children’s causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers” *Cognitive Science* 28: 303-333.

Stich, S. and Tobia, K. (2016) “Experimental Philosophy and the Philosophical Tradition”. In Systma, J. and Buckwalter, W. (eds.) *A Companion to Experimental Philosophy*. Wiley Blackwell, pp 5-21.

Turri, J. (2016) “Knowledge Judgments in ‘Gettier’ Cases” In Systma, J. and Buckwalter, W. (eds.) *A Companion to Experimental Philosophy*. Wiley Blackwell, pp 337-348.

Tenenbaum, J. and Griffiths, T. (2001) “Structure learning in human

causal induction”. In Leen, T., Dietterich, T. & V. Tresp (Eds.),

*Advances in neural processing systems* (Vol. 13, pp. 59–65). Cambridge,

MA: MIT Press.

Tenenbaum, J., Griffiths, T. and Kemp, C. (2006) “Theory-based Models of Inductive Learning and Reasoning” *Trends in Cognitive Science* 10: 309-318.

Turri, J. (2016) “Knowledge Judgments in ‘Gettier’ Cases” In Systma, J. and Buckwalter, W. (eds.) *A Companion to Experimental Philosophy*. Wiley Blackwell, pp 337-348.

.

Walsh, C. and Sloman, S. (2011). “The Meaning of Cause and Prevent: The Role of Causal Mechanism.” *Mind and Language,* 26: 21–52.

Woodward, J. (2003) *Making Things Happen*. New York: Oxford University Press.

Woodward, J. (2007) “Interventionist Theories of Causation in Psychological Perspective” In A. Gopnik and L. Schulz (eds.) *Causal Learning: Psychology, Philosophy and Computation.* New York: Oxford University Press, pp. 19-36.

Woodward, J. (2010) “Causation in Biology: Stability, Specificity and the Choice of Levels of Explanation” *Biology and Philosophy* 25: 287-318.

Woodward, J. (2018) “Laws: An Invariance- Based Account”. In Ott, W. and Patton, L. (eds*.) Laws of Nature*. Oxford: Oxford University Press, pp 158- 180.

Woodward, J. (Forthcoming) *Causation With a Human Face*.

.

1. Thanks to Richard Samuels and Daniel Wilkenfeld for helpful comments on an earlier version. [↑](#footnote-ref-1)
2. Of course I don’t intend this as a criticism of experimental work on actual causation and related subjects. [↑](#footnote-ref-2)
3. It is of possible to trivialize this sort of approach: for example, one might argue that the goal of causal thinking is to correctly describe what the causal facts are, thus rendering the approach completely unhelpful. What we want is a characterization of the goals associated with causal reasoning that gives us some independent purchase on when causal reasoning is correct or incorrect in virtue of contributing or not to these goals. “Describing the causal facts” is not a goal that can play this sort of role. [↑](#footnote-ref-3)
4. Of course one might adopt a more expansive conception of what is involved in our concept of (or the nature of) causation according to which this incorporates normative considerations, includes features of causal reasoning that are not analytically or constitutively part of the concept, and allows for the possibility that our concept may need clarification and re-engineering in various ways. Such conceptions will not be distinct from the normative project I describe. [↑](#footnote-ref-4)
5. See the discussion of “controlled” versus “natural” direct effects in Pearl, 2001. [↑](#footnote-ref-5)
6. Here I follow Knobe, 2016 who argues that a substantial amount of recent research in X-phi as well as in cognitive science is not organized around investigations into concepts and that, moreover, there are good reasons why it should not be. [↑](#footnote-ref-6)
7. I write “in a negative way” to accommodate the possible argument that intuitive judgments about cases as well as cognitive psychology show that the ordinary concept of causation is deeply confused, thus preparing the way for some very revisionary alternative account appealing to physics. Since I don’t think there is any evidence for such confusion I will ignore this possibility in what follows. [↑](#footnote-ref-7)
8. I will add that there are ways of construing the “what causation is” project that do not seem to me objectionable. For example, one might construe Cheng’s causal power model, discussed below, as in some sense an account of what causation or at least what causal power is—an account according to which causal power is to be understood in terms of certain invariance assumptions. (One might also argue that such assumptions are part of the way we think about causation.) But this account is not the sort of thing metaphysicians are looking for when they ask what causation is—there is no accompanying story about special metaphysical entities or relationships that serve as truth makers for causal claims, no “reduction” of causation to other sorts of claims or anything along similar lines. [↑](#footnote-ref-8)
9. In other words, the idea is not just that causal reasoning allows us to successfully get around in the world but that it enables a particular kind of success which is associated with manipulation and control. [↑](#footnote-ref-9)
10. Although the commonly accepted distinction between system 1 and system 2 reasoning is problematic in many respects, in light of these observations it is particularly problematic to assume system 2 reasoning is always normatively superior. [↑](#footnote-ref-10)
11. This too is a theme in much of the psychological research on causal cognition. See, e.g., Gopnik, 2009. [↑](#footnote-ref-11)
12. It should be obvious that this sort of ends/means justification also depends on such considerations as whether the goals are coherent and achievable, whether the means proposed are such that they can actually be carried out by human beings and so on. I take all of this to be built into the idea that such justification requires showing that the means *successfully* conduce to the goals. [↑](#footnote-ref-12)
13. Normative analysis is thus one point at which *apriori* or conceptual considerations legitimately enter into our story. Philosophers are not wrong to think that *apriori* reasoning has a role to play in thinking about causation but they tend to mislocate that role, thinking that it has to do with the role of intuitions in delivering truths about the concept of causation. [↑](#footnote-ref-13)
14. This strategy is extremely common in the literature on causal cognition. In addition to the examples discussed below, see Gopnik et al. 2004, and Sobel et al. 2004. [↑](#footnote-ref-14)
15. For example, the important normative distinction between intervening and conditioning (Pearl, 2000) prompted experimental work showing that human causal cognition respects this feature. Notice, though, that both this and some of the possibilities discussed above are *not* a matter of empirical results providing “evidence” for the correctness of the normative theory. The interaction between normative and descriptive can take many other forms besides this evidential connection. [↑](#footnote-ref-15)
16. This process may also involve deciding that certain judgments are correct on the basis of systematic considerations even when people have no clear intuitions, as in Lewis’ “spoils to the victor” arguments. [↑](#footnote-ref-16)
17. As an additional illustration of this difference, consider the contrast between attempting to justify some feature of our inductive practice in terms of reflective equilibrium versus justifying it in terms of means/ ends considerations. The reflective equilibrium strategy proceeds by collecting various inductive judgments we are inclined to make and then systematizing them, perhaps with further constraints. By contrast, in means/ends justification one proceeds by specifying certain goals—e.g., minimizing the probability of accepting false hypotheses and rejecting true ones—and then shows that certain testing procedures will achieve these goals. Intuitive judgments that cannot be justified in this way are rejected as mistaken, however deeply entrenched they may be and however much they cohere with other judgments. This, rather than appeals to reflective equilibrium, is the sort of justification procedure that is adopted in classical statistics. [↑](#footnote-ref-17)
18. The extent to which philosophers who report judgments about cases or intuitions are good judges of how others will judge is debated in the literature on intuition. In some cases there is some evidence that philosophical and lay judgment diverge. In other cases, philosopher’s judgments seem to accurately mirror lay judgment—see, for example, Nagel, 2016. Nagel’s defense of appeals to intuition has important similarities with my own, since it rests on a “good judge of other’s judgments” premise. [↑](#footnote-ref-18)
19. In other words, like Goldman, I don’t take the content of the judgment to be what it is evidence for. In particular, I don’t claim the overt content of the judgment necessarily has to do with shared practices. [↑](#footnote-ref-19)
20. The following example may help to clarify how this works. Consider the debates among philosophers of science in the 1950s and 60s over the role of the directional or asymmetric features in explanation. Critics of the DN model such as Scriven argued that the model was defective because it did not capture these features, mistakenly allowing the length of the shadow cast by a flagpole to explain the length of the pole. In a recent discussion, Stich and Tobia, 2016 treat this an example of an appeal to intuition, with the DN model being rejected because it is contrary to our intuitions about the flagpole case. I agree with Stich and Tobia that it is not justifiable to reject the DN model merely because it is contrary to intuition in this way. On the other hand, the intuition we have about the example brings to our attention that our practices of explanatory judgment are such that directional considerations play an important role. This can motivate us to ask whether there is some normative basis for distinguishing between flagpole to shadow explanations and the reverse. Although I won’t argue in their support here, several recent accounts claim to provide such a normative basis. In appealing to such accounts, we don’t conclude that the original intuition is correct merely because people have it—rather its correctness derives from a normative theory that supports it. [↑](#footnote-ref-20)
21. I also make no claims about whether a lot of X-phi research that is survey-like suffers from methodological flaws qua surveys. Again, in distinguishing between bare surveys and causal analysis, I don’t mean to claim that the surveys employed in X-phi and elsewhere are generally bad or flawed surveys—my point is rather that there are important limitations on what even a well-designed survey can accomplish. [↑](#footnote-ref-21)
22. One standard typology distinguishes between the negative and positive program in X-phi. The negative programs appeals to empirical results about how ordinary subjects judge to undermine armchair philosophical claims about how most people judge and claims about the nature of “our” concepts based on such claims. The positive program attempts to use claims about how most people to judge to provide evidence for how the folk think about such and such or for claims about the structure of folk concepts. Note that in both cases survey-like results seem to be all that is required to accomplish these ends. To the extent that some significant portion of X-phi research falls into either of these two categories, this helps to explain why it is (legitimately) survey-like.

    On the other hand, although I have not tried to gather systematic evidence for this claim, it is a plausible conjecture that although a substantial amount of X-phi research was once organized around investigations, negative or positive, into claims about concepts, this is much less true of more recent X-phi research, some significant portion of which looks much like cognitive science (as Knobe, 2016) argues, and shares its goals, which include causal analysis. In any case, I make no claims about how much of X-phi falls into the categories I distinguish; what matters for my purposes is the distinctions themselves. [↑](#footnote-ref-22)
23. Thus the difference between surveys and causal analysis is not that the former is observational and the latter is experimental. What is distinctive of a survey is that the design and the observations made are not such that they support causal analysis. [↑](#footnote-ref-23)
24. That something else besides the fact that no connecting process is present influences causal judgment in disconnection cases is strongly suggested by the fact that typically subjects distinguish *among* dependence claims in which there is no connecting process, assigning some greater causal strength than others. [↑](#footnote-ref-24)
25. In Woodward, 2006, following David Lewis, I used “insensitive” to describe this particular aspect of invariance—that is, invariance under changes in background conditions. [↑](#footnote-ref-25)
26. I have discussed invariance related notions in a number of different places (2003, 2006, 2010, 2018) and draw on this discussion for motivation in what follows. [↑](#footnote-ref-26)
27. In addition to the research by Cheng described below, other recent empirical research that highlights the role of invariance in causal reasoning includes Lombrozo, 2010, and Vasilveya, forthcoming. Kominsky et al. (2015) and Icard (forthcoming) highlight a role for invariance in actual cause judgment. [↑](#footnote-ref-27)
28. Thus within this framework “cause” is not defined in terms of probabilities although it is assumed that there are systematic relationships between these notions. [↑](#footnote-ref-28)
29. I see this notion of defaults as a useful way of thinking about the status of methodological maxims in science more generally. [↑](#footnote-ref-29)
30. Of course Cheng does not assume that subjects consciously reason in accord with the algebra that follows— what she presents is a computational level, rational reconstruction, with subjects judging as if they computed and represented causal power in the manner described below. Subjects have access to their judgments but typically not to the processes that produce those judgments. [↑](#footnote-ref-30)
31. Note that causal power is not *defined* as (8.4). (8.4) provides a formula for computing causal power when certain specific additional assumptions are satisfied. When these assumptions are not satisfied causal power may still be well defined but it is not identifiable from the data. [↑](#footnote-ref-31)
32. The fact that judgments of causal strength of *i* with respect to *e* increase as *P(e)* increases is a striking example of an observation that will initially seem normatively unreasonable to many but which has a non-obvious normative rationale. This is thus an illustration of the claim in Section 4 that observing that certain judgments occur, as a descriptive matter, may prompt a search for a normative theory that makes sense of them. [↑](#footnote-ref-32)
33. What follows is, for reasons of space, a very partial and incomplete description of a complicated empirical situation. [↑](#footnote-ref-33)
34. To the extent that such subject heterogeneity is real, merely reporting average subject behavior across an entire experimental population omits important information and can be quite misleading. Of course this is a general methodological problem in psychology. [↑](#footnote-ref-34)
35. A similar point holds for the experiments described in Vasilyeva et al. (Forthcoming) that test whether subject causal judgments are influenced by other sorts of invariance features of a sort described in Woodward, 2010. [↑](#footnote-ref-35)
36. To (so to speak) turn this point around, if (as I assume is likely) you are not tempted to think that the causal strength judgments made by the subjects in Cheng’s experiments are evidence for the truth of those judgments or sources of rational insight into the nature of causation, you should adopt a similar stance toward the intuitive judgments of philosophers. [↑](#footnote-ref-36)
37. See Woodward, Forthcoming for additional discussion. [↑](#footnote-ref-37)
38. For example, in considering whether *C* causes *E*, one needs to distinguish the question of whether *C* has a non-zero net effect on *E*, from whether *C* causally contributes to *E* along some route. When there is cancellation along different routes these questions should receive different answers. A look at the philosophical discussion around, e.g. Hesslow’s birth control pills example (Hesslow, 1976) will show that many philosopher’s judgments about this example failed to note this distinction. [↑](#footnote-ref-38)
39. It is worth noting explicitly that this is the kind of task (noting ambiguities, making distinctions) which traditional philosophers are good at. Properly conducted empirical analysis will thus be dependent on this sort of work as well as on normative analysis more generally. [↑](#footnote-ref-39)
40. As in Gopnik et al. 2004. [↑](#footnote-ref-40)
41. I don’t have worked out ideas about how to do this, but there are suggestive examples from the psychology literature: researchers have looked at the relationship between verbally expressed judgments of causal strength or power and measures having to do with subject’s willingness to select one cause rather than another to bring about some desired goal, willingness to generalize to new situations as evidenced in non-verbal behavior and so on. [↑](#footnote-ref-41)
42. Such ambiguity can often be detected by a combination of analysis and additional experimentation, as we see illustrated by the research by Cheng and Shanks described above. [↑](#footnote-ref-42)