DRAFT VERSION ONLY
PLEASE CITE FROM FINAL VERSION

A Philosopher Looks at Tool Use and Causal Understanding

Jim Woodward
History and Philosophy of Science
University of Pittsburgh

1. Introduction

A philosopher reading the literature on tool use among non-human primates (and, for that matter, among human infants) encounters a bewildering range of competing claims. Researchers appear to disagree in fundamental ways about the abilities of particular species of animals and even more so about how those abilities are best explained. In these circumstances it would be foolhardy for the non-expert (at least for someone as non-expert as I) to attempt to adjudicate between competing experimental claims about e.g., what chimps or tamarind monkeys can or can't do. Instead what I propose to do is to step back from such claims and ask some more general questions about the sorts of abilities that are involved in tool use and "causal cognition". Roughly, what I will do is to suppose, for the sake of argument, that the results of various experiments reported in the literature concerning tool use and other aspects of causal cognition are correct as far as they go (that is, they are non –artifactual and are genuine measures of the abilities and limitations of the animals involved). I will then ask what follows from these results concerning the capacities of these animals for causal learning and cognition.

One general line I will be pushing is this: human causal cognition is not a unitary thing; instead it involves a number of distinct abilities, although these abilities are relatively well integrated in adult human causal thinking. The abilities in question are not just logically or conceptually distinguishable, rather, experimental results concerning non-human animals and human infants show us that these abilities dissociate as a matter of actual fact, in the sense that there appear to be animals which possess some of these abilities (or which possess these abilities to some degree) but do not possess other abilities. Moreover, there are stages in human development in which some of these abilities and not others are present. A similar point holds for tool use: this too involves a dissociable bundle of abilities that are present to different degrees in different animals[1].

[1] There appears to be no generally agreed upon understanding of "causal cognition" and related notions. One possibility is to simply stipulate that "causal cognition" , "possession of causal representations" and so on require the full panoply of abilities associated with causal learning and reasoning possessed by normal adult human beings. Adopting this stipulation allows one to argue that causal cognition and understanding are unitary, dichotomous notions—one either possesses the full article or does not. As will become apparent below, the disadvantage of such a stipulation is that the various

Several consequences follow from this idea. First, in my opinion it is probably a mistake (or at least is unfruitful) to ask whether non-human animals exhibit some unitary trait called "causal understanding" —it is a better research strategy to ask about whether they possess this or that more specific ability which is related to causal learning and understanding as these exist in human adults and to ask how these abilities connect (or support) or fail to connect with one another in various species. Relatedly, it is a mistake to suppose that because animals or infants possess some of the skills that go into the complete suite of abilities that make up adult human causal cognition, they must also automatically possess other abilities in that suite. Instead, we should always ask what the specific ability is whose presence or absence is suggested by this or that experimental result and we should be wary of inferring, absent specific supporting evidence, from the presence or absence of one ability to the presence or absence of others, even if we find these abilities associated in adult humans. When we do find that various abilities associated with human casual cognition co-occur in an integrated way, we should try to understand how and by what processes this integration is achieved, rather than regarding it as inevitable and automatic or not in need of explanation. For example, as discussed below, adult humans are able to use geometrical- mechanical cues to causal relationships (having to do with, e.g., spatio-temporal contact) that may be obtained from passive observation to guide actions aimed at manipulation and control— that is, humans integrate casual representations based on geometrical- mechanical cues with causal representations that are relevant to action. However, there is evidence that human infants as well as many non-human animals fail to do this or at least fail to do it as completely and effectively as adult humans do. Non-humans and human infants may exhibit

---

abilities that go into causal cognition, so understood, can dissociate and seem to be only gradually integrated in human development. Working with a dichotomous notion directs attention away from this and makes it harder to recognize the continuities as well as the discontinuities between adult human cognition and the abilities (whatever we decide to call them) possessed by human infants and non-human animals. Readers who prefer a dichotomous notion are invited to substitute other words to characterize various elements in causal cognition that I will be talking about—"proto-causal" , causal*" etc.

John Campbell (this volume) also argues for a notion of causal representation involved in what he calls "intelligent" tool use that is graded and multidimensional, rather than unitary or all-or-nothing. Campbell carves up the various possibilities and dimensions one might have in mind in talking about causal representation , understanding, and intelligent tool use in a way that is somewhat different from the alternatives I distinguish but his resulting landscape strikes me as complimentary to, rather than inconsistent with, the one I advocate. In particular, Campbell's focus on such factors as whether the tool user's awareness of the systematic covariation between the variable properties of the target and the variable properties of the tool is grounded in the "standing properties" of both resembles my emphasis on means/end decomposition, and the extent to which the subject is able to alter the means employed in the presence of changing circumstances to achieve some desired goal, and is able to generalize across different circumstances and integrate disparate pieces of causal information into a single map like representation.

sensitivity to geometrical-mechanical cues of a sort that would suggest some level of casual understanding in an adult human and they may also learn various routines for manipulation and control that again would suggest causal understanding in an adult human but they may fail to put the two together. Understanding how such integration develops or is acquired is thus crucial to understanding adult human causal competence[2].

Second, I am also skeptical that it is useful to look for single dividing line between non-human and human causal cognition—a sort of "mental rubicon" which only adult humans have crossed. Adult human causal cognition probably differs from causal cognition in non –human animals (and in human infants) along a number of different dimensions and in the way the abilities displayed along those dimensions are integrated. For similar reasons, it is also probably a mistake to suppose that different species of animals can be arrayed along a single dimension representing degree of causal understanding, with, say, primates possessing more of this than other mammals, humans possessing more than other primates and so on. Particularly when it comes to tool use, different species will have specialized skills that reflect the particular ecological niches in which they are located and these will vary along many different dimensions. A corvid, say, may be superior to many primates in some causal learning tasks and inferior in others.

Much of what follows is a kind of typology of different sorts of abilities that might be associated with the notion of causal understanding, the acquisition of causal beliefs, causally informed action patterns and so on. I will also ask how these various abilities relate to one another, whether some may play a role in the acquisition of others and so on. I have tried to draw where appropriate on relevant philosophical literature since if there is anything that philosophers are good at, it is drawing distinctions and noticing differences. As we shall see, different philosophical accounts of causation track, at least to some degree, different and dissociable competences that go together to make up adult causal understanding.

To preview in more detail, I will begin with a sketch of competing philosophical accounts of causation, emphasizing the difference between claims about causal relationships as these exist in the world and claims about the way in which we and other animals represent causal relationships (Section 2). I will then explore the contrast between two different families of approaches to (or ways of thinking about) causation, one of which I will call "difference-making" and the other "geometrical-mechanical". (Sections 3-4) Difference-making accounts in turn differ among themselves in the way that they explicate difference-making and role that they assign to various possible sources of covariational information that are relevant to the assessment of difference-making. For example, some but not all difference-making accounts assign a special significance to processes called interventions (the paradigm of which is an unconfounded manipulation) in the explication of difference-making. I then suggest (Sections 5-8) that the following elements seem relevant to whether there is adult human-like causal cognition.

[2] A corollary is that experiments that probe the extent of such integration and how it is achieved or develops (to the extent it does) can be particularly revealing in understanding the causal competences of different animals. For examples, see sections 4 and 5 below.

1).The extent to which the subject employs causal representations that integrate difference- making information (that is, covariational or contingency information) from various sources, including the subjects own interventions, observations of the interventions of other agents, and observations of covariation that is produced "naturally" rather than by the interventions of agents.

2). The extent to which the subject employs causal representations that integrate geometrical-mechanical aspects of causation (and the perceptual cues on which these are based) with difference-making aspects. This is related to the extent to which the subject exhibits perception/action integration or dissociation in causal understanding.

3). The extent to which the subject's causal representations decouple (rather than "fuse") the representation of means and ends and incorporate detailed information about how to alter means in face of changing circumstances to achieve the same goal. A related consideration is the extent to which subjects possess representations that allow them to generalize to new circumstances and situations.

4). The extent to which causal representation is explicit, rather than merely implicit. I take this to have to do at least in part with the extent to which causal information is not encapsulated or available only to specialized systems but is rather available more generally to other systems for reasoning, inference, action and planning. Arguably, explicitness of representation is also connected to the possibility of "insight" learning as opposed to reliance on extensive trial and error learning in the acquisition of causal information.

5). The extent to which causal representation is map-like or model-like in the sense of integrating representations of individual cause-effect relationships into a single, overall representation. This includes the representation of complex causal structures such as structures in which, e.g., two effects are represented as effects of the same common cause.

6). Finally, a negative claim: in contrast to the views of a number of other writers (Leslie, 1995, Povinelli, 2000, Wolff, 2007), I think that it is by no means obvious that conceiving of causal relations in terms of "force transmission", or unobservable or hidden mechanisms is required for "causal understanding" or successful tool use and manufacture. Or at least it is not clear that these features are required when they are understood as something over and above the features described in 1-5.

2. Causation as it is in the World and Causation as Represented Psychologically.

Philosophical "theories" of causation are typically intended primarily as accounts of what causation is, as it occurs in the world. Lest this sound alarmingly metaphysical, all that I mean is that such accounts are intended to describe whatever it is that is out there in nature to which causal beliefs, judgments, or representations are answerable and in virtue of which those judgments turn out to be (perhaps only roughly or approximately) correct or not.

This issue of what causation is should of course be distinguished from issues having to do with how we (and other animals) represent, think about, learn about causal relationships, use them to guide action and so on. The latter issues, unlike the former, are issues in the empirical psychology of humans and other animals. Nonetheless the two sets of issues—the worldly one about what causation is and the psychological ones—are closely interconnected. This is partly because, among the things, we would like to explain the patterns of success and failure in various tasks probing the nature and extent of various species' causal knowledge or competence. It is a natural assumption – made by many researchers and one that I shall accept – that successful performance means that the subject is in some way tracking or exhibiting a sensitivity to some features of causal relationships as they exist in the world (features that are relevant to success on the task) and that failure is an indication that those features are not being successfully tracked[3]. For example, Povinelli, 2000 (and following him Wolpert, 2003) contend that causal relationships (or at least casual relationships involving the kinds of mechanical interactions that are common in tool use) are mediated by the transmission of unobservable forces. Povinelli also holds that chimps and other non-human primates are incapable of representing or learning about such mediating forces and this explains certain patterns of failure that (according to Povinelli) such animals exhibit in tool use tasks. Povinelli's claims about chimps' mechanical abilities are controversial and I will return to them below. The primary point I want to make here is that Povinelli's proposed explanation would make little sense if his claims about the role of force transmission in mechanical causal interactions were not (at least roughly and approximately) correct and if it were not also correct both that chimps are incapable of representing or tracking this feature of the world and that human causal cognition is as successful as it is because it does accurately track this feature of the world. Here claims about what causation is and claims about the psychology of causation, although distinct, are intertwined in various ways.

One obvious way in which claims about causal relationships as they are in the world differ from claims about how a subject represents such relationships is that the latter (in addition to being psychological claims) may be more or less adequate representations of the former—and when they are not fully adequate, this may be because they leave out features or aspects that causal relationships, as they are in the world, possess (as Povinelli claims is the case for non-human primates) or because they introduce additional features or surplus structure that causal relationships do not in fact possess (as Hume claimed about human representation of causation as involving "necessary connections") . Either way, one hopes to use this mismatch between representation and how things really are that will explain patterns of success and failure on casual cognition tasks.

[3] This should be understood as including the possibility that the tracking in question works in an indirect way, via direct sensitivity to features that covary with but do not themselves comprise the features of causal relationships that are ultimately of interest. For example, in the so-called perception of causation in launching events or Michottean collisions the features that the subject directly tracks are (in the simplest cases) the spatio-temporal parameters governing the collision. This is consistent with its being the case that in some range of ecologically normal circumstances these parameters covary with whether the relationship between the moving objects is causal or not, thus allowing for the detection of causation on the basis of spatio-temporal cues. Of course this doesn't mean that the presence of causation in such cases reduces to or just consists in the presence of these spatio-temporal relations. Similarly for the detection of causal relationships on the basis of covariational information.

Another related point that is worth underscoring is that the notion of "causal representation" is ambiguous in an important way. On the one hand, a "causal representation" may be simply a representation of a relationship that is in fact causal. On the other hand, "causal representation" may refer to the representation of a relationship as causal, in the sense that the representer possesses a full adult human notion of causation and uses this in the representation of the relationship in question. To see the difference, consider an infant who learns that by kicking it can cause a mobile to which its foot is attached to move. The relationship between the movement of the foot and movement of the mobile is certainly causal and, assuming, that what is learned involves the acquisition of a representation of some kind, the infant will therefore have a representation of a relationship that is causal. On the other hand, there are many reasons for doubting that the infant represents this relationship as a causal relationship in the way that an adult human would—for example, the infant need not be representing (as an adult would) the relationship as one that might also hold between impulses communicated to the mobile by other events besides foot movements and the subsequent movement of the mobile. Instead the infant's representation may be more local and egocentric; its content may be simply something like: depending on whether I move my foot, the mobile moves.

It would thus be a mistake to simply infer, from the fact that a subject has a representation of a relationship that is causal, that the subject represents that relationship as causal, at least in the way that adult humans represent relationships as causal. On the other hand, it is also important to avoid the opposite mistake of inferring from the fact that a subject represents a relationship but not as causal to the conclusion that the subject's (non-causal) representation of the relationship plays no role in the acquisition of an adult concept of causation or capacity for causal representation. To anticipate an example discussed below, successful imitation of a tool using routine by a conspecific sometimes (perhaps often) occurs in the absence of detailed causal understanding. It does not follow from this, however, that in humans the capacity to imitate plays no role in the acquisition of the adult capacity to reason causally or to employ full-fledged adult causal representations. Instead, a great deal of empirical evidence suggests that the acquisition of representations of relationships that are causal but which are not represented as causal plays an important bootstrapping or scaffolding role in the achievement of modes of thinking in which relationships are represented as causal.

3. Difference Making versus Geometrical-Mechanical Conceptions of Causation.

Current philosophical approaches to causation fall into two broad categories that differ in some fundamental ways[4].

[4] The contrast that follows between difference-making and geometrical/mechanical accounts of causation seems to very roughly parallel the distinction that Peacocke (this volume) draws between what he calls "comparative explanation" and "causation" understood in "actualist" terms. But while Peacocke apparently regards only the latter and not the former as genuinely causal, I see both as components or elements in causal thinking and as corresponding to legitimate notions of causation or causal explanation. Arguing in detail that difference-making is genuinely causal would take us far beyond the scope of this essay, but for relevant discussion see Woodward, 2003. I will also add

Difference-making accounts focus on the idea that the distinctive feature of causes is that they make a difference for their effects. For example, exposure to tobacco smoke might be regarded as a cause of lung cancer in the sense that such exposure makes a difference to the occurrence of lung cancer (in this case, by increasing its incidence) in comparison with alternative situations in which there is no such exposure. By contrast, what I shall call geometrical-mechanical theories of causation focus on the idea that what is distinctive about causes is that they are "connected" in some appropriate way (often thought to be specifiable in geometrical or mechanical terms) to their effects. Often, in the philosophical literature (e.g., Salmon, 1984, Dowe, 2000), this notion of an appropriate connection is cashed out in terms of the cause being spatio-temporally contiguous with the effect or being linked to the effect via a spatio-temporally continuous process that transfers energy or momentum. (In the psychological literature, this is sometimes put in terms of the notion that there is transmission of "force" from the cause to the effect, as in Leslie, 1995.) A case in which a moving billiard ball collides with a stationary ball and causes the latter to move is a paradigm of such a geometrical mechanical interaction: the first ball comes into spatial contact with the second, the second begins to move immediately after the collision with no temporal gap, and energy and momentum is "transferred" (according to folk physics[5]) from the first to the second. There are many other possible forms of geometrical- mechanical interaction, some of which are described below

Difference-making Accounts of Causation. The philosophical literature contains a variety of different proposals about how to explicate the notion of difference-making. Probabilistic theories hold that a cause C must make a difference to its effect E in the sense that the presence or absence of (or a change in the value of) C changes the probability of E at least when various other factors (such as an appropriate set of other causes of E besides C) are controlled for (Eells, 1991)[6]. Counterfactual theories (e.g.,

---

that Peacocke's view that causation should in all cases be understood in actualist, non-difference making terms is in serious tension with a great deal of ordinary and scientific thinking about causation. For example, it seems to conflict with the widely accepted idea that randomized experiments are a particularly good way of identifying causal relationships since the relationships so identified are difference- making relations. It also fits badly with the observation from empirical psychology that people's explicit judgments of causation and causal strength closely track difference-making or contingency information.

[5] "According to folk physics" because it isn't clear how to make sense of this picture from the point of view of modern physics. This is because the laws governing the collision are invariant under any inertial transformation; hence it is equally valid to adopt a frame of reference according to which the first ball as stationary and the second moving prior to the collision in which case the transfer of momentum will look as though it is from the second to the first ball. Which is to say that the notion of transfer of momentum from one ball to another is not a frame-independent notion.

[6] A terminological convention that I will follow throughout this paper (unless explicitly indicated otherwise) is that upper case letters such as C, X etc. represent variables, rather than predicates or properties. A distinguishing feature of a variable is that it must be able to take two or more values. We can translate the usual philosophical talk of causal

Lewis, 1973) attempt to explicate difference-making in terms of counterfactuals – a very simple (naïve) version of such a theory (not Lewis') might hold that *C* causes *E* if and only if the following two counterfactuals hold. (1) If *C* were to occur, then *E* would occur, (2) If *C* were not to occur, then *E* would not occur[7].

Interventionist or manipulability accounts of causation of the sort defended in Woodward, 2003, 2007 may also be regarded as a version of a difference-making theory. Such theories have interesting implications for the psychology of causal learning and judgment and for this reason, I briefly elaborate on their structure. According to such theories, causes make a difference to their effects in the sense that if an intervention (roughly an exogenous unconfounded experimental manipulation which puts the variable intervened on entirely under the control of whatever causes the manipulation) on the cause variable were to occur (e.g., by introducing the cause into an appropriate situation in which it was previously absent or removing it from a situation in which it was previously present), the value of the effect variable would change. In the simplest formulation of such a theory *C* causes *E* if and only if *C* and *E* remain correlated under interventions on *C*[8]. Interventionist accounts attempt to capture in this way the common-sense idea that causes can be thought of as "handles" for manipulating or controlling their effects. Interventionist theories don't claim that the only way to learn abut causal relationships is to actually perform interventions -- obviously adult humans sometimes learn about causal relationships from passive observation of covariational information as well as from many other sources (e.g., testimony from others). However, according to interventionist accounts, the causal claim one learns in such a case has an interventionist interpretation – roughly, it is counterfactual information about what would happen if one were to perform an intervention. For future reference I will call such counterfactuals interventionist counterfactuals.

---

relations among events or properties by invoking binary variables corresponding to the presence or absence of the events or properties—these are the "values" of these variables. Representation of causal relations as relations among variables is standard in science and is the natural way of capturing the difference-making aspect of causation.

[7] This is the simplest possible version of a counterfactual theory and it is not news that it does not deal adequately with cases involving pre-emption, over-determination, and other complexities. Readers should appreciate that there are versions of counterfactual theories that do deal in a fairly satisfactory way with such cases, roughly by relying on counterfactuals with complex antecedents—detailed illustrations are provided in Hitchcock, 2001 and Woodward, 2003. For example, although from (1) if *C* causes *E*, it does not follow (in a situation in which some other cause of *E* would have been operative if C had not) that (2) if *C* had not occurred, *E* would not have occurred, a more complex counterfactual—roughly, (3) if *C* had not occurred and no other cause of *E* had occurred, then *E* would not have occurred, is naturally associated with (1). It is thus in my view a mistake to suppose that cases involving pre-emption show that no broadly counterfactual or difference-making account of causation is correct—rather what such cases show is the need for a more sophisticated difference-making theory

[8] More complex versions of the theory are needed to deal with examples involving pre-emption etc. For details, see Woodward, 2003.

Part of the appeal of interventionist accounts is that they provide a natural way of distinguishing between genuinely causal relationships and mere correlations. As an illustration, consider a common –cause structure in which C is the common cause of two joint effects E_1 and E_2 which are not themselves directly causally connected:

$$E_1 \leftarrow C \rightarrow E_2$$

In such a structure E_1 and E_2 will be correlated but any intervention on E_1 (which in this case will involve a manipulation of E_1 which is uncorrelated with C, unconfounded with any other cause of E_2 and so on) will disrupt the correlation between E_1 and E_2 telling us that E_1 does not cause E_2. This is because the intervention will "break" the previous causal connection between C and E_1 in the common cause structure, putting E_1 under the control of the exogenous source of variation provided by the intervention. By contrast some interventions on C itself will result in changes in E_1 and E_2, reflecting the fact that C causes E_1 and E_2. In this way, interventionist accounts have the resources to distinguish between correlations (or, for that matter, relations of counterfactual dependence) that do not reflect direct causal relationships[9]. (Translating this into psychological terms, one test for whether a subject understands or represents that she is dealing with a common cause structure like that described above would be to determine whether the subject appreciates that even though E_1 and E_2 are correlated, intervening on E_1 is not a way of bringing about a change in E_2 —the experiment due to Blaisdell et al. described below has this sort of structure)

While the difference-making accounts just described are theories of what causal relationships are, as they exist in the world, such theories have an obvious affinity with many theories of causal learning, representation, and judgment within empirical psychology that also assign a central role to difference-making information. In saying that such theories assign a "central role" to difference-making information, I mean to include possibilities like the following: the theory takes the representation of causal relationships to involve the representation of information about difference-making, or it takes causal relationships to be learned from difference-making information (although perhaps not only on the basis of such information), or it takes causal judgment to be guided by or sensitive to such information.

By "difference-making information", I mean information about the contingency or covariation between cause and effect, including covariation conditional on other causal factors[10]. This information might be supplied by passive observation of contingencies as

---

[9] Roughly speaking, the notion of an intervention plays the same role as the "similarity metric" that characterizes closeness of worlds in Lewis' theory. The notion of an intervention requires that interventionist counterfactuals receive a non-backtracking interpretation, and allows us to distinguish these from counterfactuals with a backtracking interpretation, which do not receive a causal interpretation.

[10] It is important to understand that it is built into the notion of contingency or covariational information that it relates different values of two or more variables to one another or different states of some set of causal factors to one another. To say, for example, that property C is a sufficient condition for E does not in itself provide contingency information, since it tells us nothing about what would happen to E if C were

they occur in nature independently of the activities of any agent or it might instead involve the observation of the results of the agent's own manipulative activities, as when the agent learns the contingencies between her actions or interventions and the outcomes they produce. Alternatively, it might involve the observation of such action/outcome contingencies when another agent acts.

Specific examples of such psychological versions of difference-making theories include "associative" accounts of causal learning and judgment, accounts of judgments of causal strength according to which this depends on $\Delta p = P(O/A) - P(O/-A)$ (where A is some action the subject chooses and O an outcome[11]—see Dickinson and Shanks, 1995) and Patricia Cheng's causal power theory, according to which the causal power of a generating cause C in causing E is given by $\Delta p/1-P(E/-C)$ in the special case in which causes (-C) that are alternative to C both occur and influence E independently of C. (Cheng, 1997) Accounts according to which the representation of causal relationships has the structure of a Bayes' net (and/or according to which causal learning involves learning the structure of a Bayes' net) are also naturally understood as difference-making in spirit, since such structures involve the representation of claims about how the probabilities of the values of effect-variables will change depending upon changes of various sorts in the value of cause variables (Cf. Gopnik et al., 2004). Similarly for a psychologized version of an interventionist (or counterfactual) theory of causation, according to which subjects internally represent causal relationships as claims about what would happen if certain interventions were to be performed (or in terms of counterfactuals about what would happen if the cause were to be different in various ways.)

As remarked above, as I understand difference- making accounts of causation and causal representation, they need not be committed to the claim that the only way that subjects can learn about causal relationships is through extensive observation of patterns of covariation. It is consistent with such theories that, given the right circumstances and background information, subjects may learn about new causal relationships (interpreted along difference-making lines) on the basis of only a few or single observations or alternatively, from sources like the testimony of others. Such learning can still involve the representation of difference- making relationships as long as the content of what is learned has to do with the holding of such a relationship—that is, as long as it is true that in learning that *C* causes *E*, the subject learns (and represents) something like: *C* makes a difference for the probability of *E*, or intervening on *C* will change *E,* or some other difference-making relationship.

In what follows, I will, however, assume that difference-making accounts are committed to the claim that observations of (or information about) patterns of covariation or contingency are one (even if not the only) important source of

---

different or absent. Thus an agent who learns a contingency between some action she performs and an outcome does not just learn that if such and such an action A is performed, outcome O results. Instead she learns that whether O results depends on whether A is preformed—that is, that it is also true that if she does not perform A, O does not result. Thinking in terms of relationships among variables is the natural way of representing contingency information.

[11] Here I follow standard usage in psychology in using O and A to represent properties or events rather than variables. Similarly for E and C in Cheng's theory.

information about causal relationships; that is, I will take such accounts to assume that causal judgments will at least be sensitive to contingency information even if casual judgment may be sensitive to other factors as well and even if such information may not by itself be enough to fully fix which causal relationships or judgments are learned. As an empirical matter, the claim that causal judgment is sensitive to contingency information does not seem controversial. As Schlottman and Shanks (1992) remark, there is a huge body of experimental evidence for this claim. It is one of the virtues of difference-making accounts that they seem to provide a natural explanation of why contingency information is relevant to causal judgment: such information is relevant because causal judgment involves or amounts to the representation of information about difference-making (in some form or other).

Geometrical-Mechanical Accounts of Causation. I turn now, for purposes of comparison to some comments on geometrical-mechanical accounts of causation. I said above that a paradigmatic case of the kind of causal interaction such accounts are intended to capture is the collision of two billiard balls – and in fact, it is this sort of example which so-called causal process theories in the philosophical literature of the sort defended by Salmon, 1984 and Dowe, 2000 best fit. However, as I shall understand geometrical-mechanical accounts, they are meant to apply much to more broadly to phenomena involving mechanical interactions and contact forces, whether or not energy/ momentum transfer is present. These include (at least) interactions involving pushing and pulling (either directly or through the use of rigid objects as intermediaries), breaking, support of one object by another, and the role of solid objects in constituting impenetrable barriers to the movement of other objects. Obviously an appreciation of or sensitivity to such interactions and the mechanical properties that mediate them (properties like rigidity, impenetrability, weight and so on) play a central role in many kinds of tool construction and use. Although such interactions need not involve the episodes of energy/ momentum transfer on which causal process theories focus, they do often involve, from a physics perspective, interactions mediated by contact forces. For example, when one stationary object is supported by a second, the second exerts a contact force on the second, even though there is no energy/momentum transfer (However, not all phenomena involved in naïve physical reasoning and tool use are to be understood in terms of contact forces. The role played by gravity, which is not usually understood as a contact force, but which nonetheless plays an important role in the understanding of weight and in naïve physics more generally is the obvious exception.)

Just as difference-making accounts of causation within philosophy are paralleled by difference-making accounts of causal representation within psychology, geometrical-mechanical accounts of causation within philosophy are paralleled by accounts within psychology that emphasize the role of the representation of geometrical/ mechanical relationships in causal cognition and learning. Thus, within psychology, a subject's understanding or representation of the contact mechanical phenomena just described is often taken to involve notions like "force transmission" (Leslie, 1995) or "force dynamics" (Wolff, 2007) or to involve the representation of unobserved forces (Povinelli) or to involve the deployment of "theory of body" (Leslie, 1995 again) or various "core physical principles" (Spelke et al. 1995) The latter specify, for example, that moving solid objects follow spatio-temporally continuous paths and cannot pass through each other, that objects continue to exist when hidden from sight, that the parts of (many) objects move cohesively together and so on for other assumptions of "naïve" physics. In other words, a subject's abilities to recognize and perhaps reason about mechanical interactions involving collisions, pushing, support, containment, and so on are understood in terms of their possession of the concepts, theories, and principles just described. Often this is accompanied by the claim that the concepts and principles deployed in the recognition and understanding of mechanical interactions serve as the basis for more general notions of causation and causal mechanism, which are then taken to be central to causal cognition in other circumstances as well. One suggestion, to be considered below, is that non-human animals lack the full ability to represent or learn about such geometrical-mechanical relationships that is characteristic of adult humans and that this explains (at least in part) the limitations of such animals in tool use and causal understanding.

I noted above that in many cases, recognition of causal relationships involving geometrical mechanical interactions relies heavily on perceptual (and in particular visual) cues. For example, when presented with launching phenomena in which one moving object strikes another and the latter begins to move, adult humans will (when the spatio-temporal parameters governing this interaction are appropriate) have the impression that they visually perceive the impact of the first object to cause the movement of the second[12]. In other cases, subjects employ visual cues in recognizing that one object supports another, that an object is too large to fit through an opening in another and so on. (In saying that there is heavy reliance on visual cues in these cases, I do not mean that these are the only relevant considerations. For example, subjects may also rely on background assumptions that the objects in question are rigid or impenetrable and the presence of these latter properties is arguably not ascertainable just on the basis of "purely visual" processing. In this respect, a property like rigidity is different from properties and relationships having to do with shape or spatial contact. The notion of a "mechanical" property is not that of a purely visual property, even supposing that the latter has a clear meaning.)

4. The Relationship between Difference-Making and Geometrical-Mechanical Conceptions of Causation.

What is the logical or conceptual (or for that matter, psychological) relationship between causation (or causal representation) conceived in terms of difference-making and causation conceived of along geometrical mechanical lines? This is a complex question which has not received a settled answer in either the philosophical or psychological

---

[12] Peacocke (this volume) denies there can literally be perception of causation on the grounds that perception requires that " instantiation of the property (or relation) [that is perceived] explains the information from which perception of the property is computed". I see no reason to accept this requirement, but in any case, nothing will turn on this issue in what follows. Readers who do not think that there is such a thing as perception of causation can instead think of the above examples in terms of detection of causal relationships on the basis of perceptual cues.

literatures. Nonetheless it is an issue that is very much worth exploring because of its implications for the understanding of causal cognition.

One obvious point of difference between the two ways of conceptualizing causation concerns the range of application of the two approaches. The difference-making conception is domain general in the sense that a difference-making relationship can exist between virtually any two pairs of factors, as long as they covary together in the right way. Thus difference-making causal relationships can hold between physical or mechanical items like billiard balls (collision with the first ball makes a difference to whether the second ball moves), but also between mental items like beliefs, between mental and physical items (beliefs and behavior), and between social and economic variables (expansion of the money supply causes inflation.) In other words, the constraints on causation imposed by difference- making accounts are formal, rather than material—they don't restrict the content of causal claims in the sense of excluding certain kinds of items as candidates for causal relata on the apriori grounds that no items of those sorts could possibly stand in a causal relationship.

By contrast, the geometrical-mechanical conception is much more domain specific: it applies straightforwardly to events and properties that stand in certain well-defined spatio-temporal relationships or which can possess mechanical properties like energy and momentum, but it is far from clear how this conception might be extended to other items that seem to lack these features. For example, if thoughts lack a definite spatial location and/or if it makes no sense to ascribe energy or momentum to a thought, then it is hard to see how they can enter into causal relationships, conceived along geometrical mechanical lines, with other thoughts or behavior. And if we represent thoughts in such a way that they are not represented as having spatial location or energy/momentum, then our representations of the causal relations in which thoughts figure is presumably not such that these relations are represented in terms of spatial contiguity and energy/momentum transfer. To the extent that we (or other animals) are prepared to think in terms of mental or social causation or to learn or represents such relationships, this seems to involve a concept of causation that extends beyond the bounds of the geometrical- mechanical conception.

Conceptually, one of the most obvious differences between difference-making treatments of causation and geometrical-mechanical treatments is that (at least in standard philosophical formulations) the former conceive of casual claims as comparative while the latter do not. What I mean by this is that on difference-making accounts a causal claim always involves a comparison or contrast of some kind between what happens or would happen when the cause is present and what happens or would happen when the cause is absent or different. By contrast, on geometrical mechanical accounts, whether, say, *c* causes *e* depends just on whether c and e occur and on whether there is an appropriate connecting relationship between them. On the standard statement of such accounts in the philosophical literature, given that *c* and *e* occur, what happens or would happen in other situations in which *c* and *e* (or *c*-like and *e*-like events) don't occur (for example, whether it is true that if *c* does or would not occur, then *e* does or would not occur) is taken to be irrelevant (or at best only indirectly evidentially relevant) to whether there is a connecting relationship between *c* and *e* [13].

Presumably, this is not unrelated to the role that perceptual cues play in establishing geometrical mechanical relationships. Suppose that you see a moving ball strike a stationary ball and the latter begin to move. In this case it is a very natural thought that all the information that you need to establish that the impact of the first ball caused the second to move is contained in the perceptually accessible features just described—the movements of the balls, their spatio-temporal contact, and so on. In particular, while you may form a judgment about the how the second ball would have behaved in the absence of a collision, it appears that you don't need to form such a judgment in order to reach a causal conclusion about this situation. In accounts such as Salmon, 1984 and Dowe, 2000, this intuition is reflected in the idea that the notion of a connecting causal process can be explicated without any appeal to counterfactuals or other sorts of difference-making information. Put in terms of empirical psychology, what this seems to suggest is that it might be possible for a creature (e.g., a human infant or a non-human primate) to possess some version of a geometrical- mechanical conception (or representation) of causation or at least be sensitive to the presence or absence of connecting causal processes and other sorts of geometrical-mechanical relationships without being able to represent counterfactual or other sorts of difference-making information or at least without being able to integrate difference-making information with information about connecting processes. I will return to this possibility below.

Another contrast between difference-making and geometrical mechanical approaches to causation, implicitly assumed above, is this: geometrical mechanical accounts seem to apply most naturally or directly to individual causal sequences involving particular events (so-called "token" causation, in contrast to "type" causation, in philosophical parlance) — e.g., this particular ball causing another to move on a specific occasion. After all, spatio-temporal contact, connecting causal processes, and so on hold (or not) between such particular events, rather than between types of events. Of course, one may form generalizations ("billiard balls must come into contact if one is to cause the other to move") incorporating information obtained from particular causal interactions but it is the particular interactions that seem primary and seem to serve as the basis for the generalizations. This is in contrast to the situation with respect to difference-making accounts. Difference-making accounts seem to apply very naturally and straightforwardly to causal relationships that are general or repeatable or to representations of such relationships as general and repeatable. This is so even though one can also construct difference-making accounts that are meant to apply to causal relationships between particular events—David Lewis' counterfactual theory (Lewis, 1973) of causation is a prominent example.

Further evidence for deep conceptual differences between difference-making and geometrical mechanical conceptions of causation is provided by the observation that the presence of the features emphasized in each account seem neither necessary nor sufficient for the presence of the features emphasized in the other. For example, difference- making can apparently be present without the presence of spatio-temporal contact or connecting

---

[13] The geometrical/mechanical conception is thus an actualist conception of causation in the sense characterized by Peacocke (this volume).

processes emphasized in geometrical mechanical accounts. Causes that "act at a distance" such as Newtonian gravity are a case in point—the gravitational force exerted by the sun certainly makes a difference to the trajectory of the earth but a least within the Newtonian framework there is no connecting process between these two bodies. Similarly, psychological or mental causation (as ordinarily conceived) can be present in the form of difference-making without causal connectedness understood in geometrical mechanical terms.

Conversely, geometrical/mechanical connectedness is apparently not sufficient difference-making or at least the presence of such connectedness may tell us very little about what difference-making relationships are present. Hitchcock, 1995 asks us to consider a cue stick whose tip has been rubbed with blue chalk. The cue stick is used to hit the cue ball which then hits the eight ball, sending it into the pocket of a pool table. Particles of blue chalk are transmitted from the cue stick to the cue ball to the eight ball—this is a connecting causal process in the Salmon/Dowe sense which transmits energy and momentum in a spatio-temporally continuous way. Nonetheless, in ordinary circumstances the transmission of the blue chalk does not make a difference to whether the eight ball falls in the pocket; the outcome would have been the same even if no blue chalk had been present. Similarly, when a tennis ball is thrown against a wall and rebounds from it, there is a connecting causal process from the thrower's hand to the wall, but this is not what makes a difference for whether the wall stands up or not.

These last two examples show that whether or not there is a connecting causal process (or a geometrical -mechanical interaction) between *c* and *e* may provide little detailed or useful information about the factors that make a difference for the occurrence of *e* – indeed the presence of such a process is compatible with the occurrence of *c* making no difference at all to the occurrence of *e*. (And even when some feature or property of *c* does make a difference to *e* and there is a connecting process, this last fact may tell us nothing about what the difference-making feature is.) For example, the information that there is a connecting causal process from the cue stick to the eight ball by itself tells us nothing about the detailed features of this interaction (the momentum communicated to the cue ball) that "make a difference" for whether the eight ball goes into the pocket. Given a creature who wishes to control or manipulate ("make a difference for") whether the eight ball goes into the pocket, the advice to establish a connecting causal process between the cue stick and the eight ball is of very limited usefulness.

To see the relevance of this to issues about tool use, consider some experiments conducted by Povinelli and others (cf. Povinelli, 2000) A primate is given a choice of tools with which to retrieve a food source that is some distance away. Some of the tools are appropriate for this task; they are extended sticks with rigid hooks at the end (or alternatively they are rake-like devices with tines) which can be used to snare the food; other tools are sticks with nothing or non-rigid, inappropriately shaped devices at the end. Roughly speaking, the primates put the tool they choose in spatial contact with the food, but they do not, without extensive trial and error, preferentially choose tools that are otherwise appropriate for retrieval In other words, they behave as though they appreciate that the tool must be in spatial contact with food source in order to retrieve it, but they do not choose tools in a way that suggests that they further appreciate the significance of whether the tool is appropriately shaped, or rigid. Moreover, even when they select tools with a hook at the end, they often do not appropriately orient these to the food. It is though the primates grasp the idea that retrieval of the food requires that there be a causal process connecting their hands to the food (putting the stick in contact with the food constitutes such a process), but don't get the idea that using the tool in a way that makes a difference for food retrieval requires something more.

It is not much of a stretch to see Povinelli's primates as providing a real-life illustration of the conceptual distinction described above between information about connecting causal processes and the more fine-grained difference making information required for successful manipulation. Roughly speaking, the primates seem to be in the position of a person who, in Hitchcock's example, realizes that to use the cue stick and cue ball to manipulate whether the eight ball goes into the pocket, the cue stick must be brought into contact with the cue ball and the latter with the eight ball, but who does not appreciate that it also matters further exactly how (with what force and direction) the cue strikes the cue ball. Or, to put the matter slightly differently, the subjects in Povinelli's experiment seem to be sensitive to the presence of some simple spatial/mechanical properties that are causally relevant in a very general way to the outcomes they wish to produce (e.g., they are sensitive to whether there is spatio-temporal contact) but they are apparently much less sensitive to the presence or absence of other (perhaps more complex or abstract) properties that also seem "mechanical" (e.g., properties having to do with shape and rigidity) that are very important in successful tool use.

I noted earlier that a number of researchers (in philosophy, psychology, primatology etc.) claim that human beings (or at least adult humans) conceptualize or represent causal relationships in terms the operation of mechanisms involving physical force transmission, communication of energy/ momentum and the like. It is also commonly suggested that non-human animals, including other primates, fail to represent causal relationships in this way and that this provides an underlying explanation for deficits in causal understanding like those described in the tool use experiments above.

Although this is a seductive idea, my discussion above suggests that it is far from obvious that it fully accounts for the difference between human and non-human performance. At the very least additional evidence and argument seem to be required before we accept this contention. Consider first an adult human who is successful in the tasks described above, choosing a correctly shaped implement to retrieve a food source and so on. We may think of this person as exhibiting in her behavior (and, in a sense to be discussed in more detail below, being guided by her knowledge of) various interventionist counterfactuals: if I use a tool with a hook at the end and orient it correctly and put it in contact with the food source, then I can use this to move the food toward me. If the tool is not in spatial contact with food, pulling on it will not move the food. And so on. These are all counterfactuals of the form, "If I do *X*, then *Y* will result", where *X* and *Y* are variables - that is, we think of *X* and *Y* as taking different values such as "present" and "absent" so we that have a representation of what happens when contact is present, when it is absent and so on . As we have seen, the general ideas that causal relationships involve force transmission or energy transfer via spatio-temporal contact and so on does not by itself tell us which of these more detailed interventionist counterfactuals are true. It is logically or conceptually possible to possess a concept of causation in terms of force transmission (or geometrical-mechanical connectedness) and yet to be completely clueless about the detailed dependency relationships that are relevant to particular task involving object manipulation. Conversely, it also appears logically possible to possess knowledge of such detailed dependency relations and reflect them in one's behavior without possessing an abstract representation of causation in terms of force transmission. Indeed, for what it is worth, there is evidence that, as an empirical matter, adult humans have surprisingly shallow and inaccurate declarative knowledge about many familiar physical mechanisms and interactions: they are certainly unable to provide correct, coherent explanations of the behavior of these in terms of notions like force and energy transmission. (Keil, 2003)

I take these considerations to raise the following question. Even if it is true that adult humans often conceptualize causal relationships in terms of some very general idea of causal transmission through geometrical-mechanical connectedness (causes are thought of as transferring "force", "energy", "umph", "biff" or some such to their effects), it is far from clear that this conceptualization is what explains patterns of success or failure in specific tool using or object manipulation tasks. On the one hand, the generic idea of force transmission does not by itself give one the specific interventionist counterfactuals tool manipulation requires. On the other, if a subject has acquired these interventionist counterfactuals, why isn't this by itself enough to explain tool use— why is having an abstract representation of causation in terms of forces or energy flow is required as well? I don't claim that these questions are unanswerable, but I do think that they require more attention than they have hitherto received[14].

Two Concepts of Causation? So far I have been emphasizing the differences between geometrical mechanical and difference making conceptions of causation. Awareness of these differences have led at least one prominent philosopher (Hall, 2004) to propose that adult humans operate in ordinary life with two distinct "concepts" of

---

[14] In her talk at the Warwick workshop and in several recent papers (e.g. Visalberghi et al., 2009), Elisabetta Visalberghi presented information about effective tool use by wild bearded capuchin monkeys in Brazil. These animals selected stones of appropriate size, weight and friability to crack open nuts—they even selected stones of the appropriate weight when presented with artificial, non-naturally occurring stones for which size was not correlated with weight. If we wish to explain (or even just describe) the contrast between the successful performance of these capuchins and the unsuccessful performance of Povinelli's apes at food retrieval, does it really help to appeal to the idea that the former must have some abstract way of representing causal relationships (in terms of force transmission or anything else) that the latter lack? Of course, the capuchins are sensitive in some sense to the relevance of weight for the particular purpose for which they use the stones, and the apes are apparently not sensitive to the relevance of shape and rigidity for food retrieval. One might go on to ask why this is so—one might conjecture that the answer to this question has to do with such considerations as the prior experience of the capuchins with their task and perhaps its relative ecological naturalness (in comparison with the task faced by the apes). Perhaps also there is something about shape and rigidity that make their causal relevance particularly hard to learn for non-human primates. But this is still a matter of learning how particular factors do or do not make a difference for manipulation tasks, not a matter of having a representation of causation in terms of forces or not.

causation: one of which, “dependence” corresponds roughly to what I have called difference- making and the other of which, “production” has affinities with (although it does not coincide exactly with) what I have called the geometrical-mechanical conception. I put to one side the issue of whether this two distinct concepts idea is the best way to describe the differences to which I have drawn attention or whether instead it might be better to speak of two different accounts of a single concept of cause (perhaps focusing on distinct strands or elements within that concept); the point that I want to stress is that even if Hall’s view is accepted, his two concepts usually seem to be well integrated in ordinary adult human causal thinking. For example, adults typically move readily and smoothly from geometrical-mechanical cues to causal connectedness to difference-making judgments: seeing one billiard ball strike another and the latter change direction, we perceive or infer on the basis of geometrical-mechanical considerations that the impact of the first caused the change in motion of the second), but also immediately judge (in the absence of complicated scenarios involving the presence of other, over-determining causes) that the impact of the first ball is what made the difference for the change in motion of the second. Seeing the apple resting on the table (a geometrical-mechanical relationship), we immediately judge that the presence of this support is what makes a difference for whether the apple falls, that we can make the apple fall my removing it from contact with the table and so on. Moreover, at least in many cases, when we find difference- making, we expect to also find causal connectedness or geometrical/mechanical relatedness: if flipping the switch makes a difference for whether the light is on, we expect that there will be a (perhaps hidden or non-apparent) connecting process between the two. In other words, even if there are two distinct concepts, we expect systematic connections between them.

However, both the differences to which we have drawn attention and experimental results like Povinelli’s suggest the possibility that there is nothing inevitable about the integration of the two concepts (or elements or strands) in causal cognition. That is, it seems entirely possible that a creature might have (or appear to show) sensitivity to some of the geometrical mechanical cues that adult humans take to be relevant to the presence of a causal relationship and yet not automatically move from these to an appreciation of their difference- making significance. Or alternatively, the creature might be sensitive to some simple spatial or geometrical cues to causal relationships but not other such cues and might not appreciate the difference-making significance of the latter. For example, the creature might be sensitive to whether there is spatial contact but not to more complex spatial properties or to properties that seem “mechanical” but not purely spatial or geometrical (e.g., properties having to do with rigidity, impenetrability, weight etc. whose detection and representation may depend on haptic or kinesthetic as well as visual experience). Or the creature may not appreciate how these matter for the kind of difference-making associated with successful manipulation.

Put the other way around, one capacity that may be important for successful tool use, at least when carried out by adult humans, is the ability to appreciate the relevance of geometrical-mechanical information for difference-making and manipulation. Human tool users can often “read off” from geometrical-mechanical cues having to do with spatial contact, shape, rigidity and so on—cues that can be recognized on the basis of vision or kinesthetic/ proprioceptive experience—what difference-making relationships are likely to be present. Inability to see the relevance of the former to the latter or to integrate the two is one possible source of a deficit in causal cognition and tool making ability[15]. Perhaps this is one aspect of or element in the limitations in causal understanding exhibited by Povinelli's primates. Very roughly, Povinelli's subjects may have elements of each of Hall's two concepts, but may not fully put these together or integrate them in the way human adults do.

Does a creature with these sorts of limitations have a "concept" of cause or causal representations or causal understanding? Rather than trying to provide a "yes" or "no" answer to this question, I think it is better to simply say that we have in this case a fragmentation or dissociation of competences that typically go together in human adults. Mapping these competences and how they connect or dissociate in the case of different species seems more worthwhile than arguing about which animals have genuine causal representations.

Two additional remarks may help to clarify these suggestions. First, note that the claim that some creature does not appreciate relevance of a geometrical-mechanical property for difference-making does not necessarily mean that the creature is completely insensitive to the presence or absence of the property or that the property is "unobservable" for the creature. A primate might be able to distinguish sticks with hooks from sticks lacking hooks, but may not recognize that this distinction makes a difference for the ways in which the stick can be used in manipulation[16]. In other words, it may visually recognize the difference in shape, but may not be able to use this information to guide its actions.

A related point is that it may also be the case that the creature is able to recognize when an object is behaving anomalously in ways that violate constraints associated with a geometrical-mechanical conception of causation, but again may fail to use recognition of these constraints to successfully guide action. Indeed, this pattern arguably is illustrated by a number of perception/action dissociations that are found among both human infants and non -human primates For example, human infants and non-human primates will look longer at an object that appears to fall through a solid surface behind an occluder than at an object whose fall appears to be blocked by the surface. This is often taken to show that it registers with these subjects that the former object exhibits causally anomalous behavior or violates a causal constraint. However, the same subjects will also search below a solid surface when an object is dropped above it, thus apparently failing to incorporate into their actions the "knowledge" of the causal constraint that they are taken to exhibit in the looking time task[17]. (Hauser, 2001) In these cases, it is presumably not sensible to think of the non-optimal search behavior as a result of failure

---

[15] Goldberg (this volume) discusses apparent cases of such a deficit or dissociation: human subjects with left parietal damage who are unable to read off difference-making features of tools from their geometrical/mechanical properties—cf. footnote 21.

[17] For additional examples and discussion of perception/action dissociations in human toddlers and non-human primates, see Santos, Seelig, and Hauser, 2006.

to observe or detect some property; it rather seems to reflect a failure of integration (of perception and action) of some kind [18].

5. Aspects of Difference-Making and their Interrelations.

So far my argument has been that by distinguishing between, on the one hand, geometrical -mechanical features of causation/causal representation and, on the other, difference-making features and by exploring their interrelations, we may get some insight into primate and infant causal competence. I want now to pursue the same line of inquiry with respect to difference-making itself, exploring some of the interrelationships among different aspects of this notion.

It is useful to begin by distinguishing three sources of difference-making information.

First, difference-making information that a subject learns from its own actions or interventions. For example, S learns that if it performs action A, outcome O follows, and that if S refrains from A, O does not (assuming no other cause of O is operative[19]). In other words, S learns that A makes a difference for whether O occurs

Second, difference-making information that is learned from observing the actions of others and the action-outcome contingencies that result. S* learns that A makes a difference for O by observing some other subject S perform O.

Third, difference-making information that is learned from observing covariational information that occurs in nature that does not involve the actions of any agent. For example, a subject learns that naturally occurring rainfall makes a difference for plant growth by observing the covariation between these variables.

Associated with these three different sources of difference-making information we may distinguish three different possibilities for causal learning.

1) Let us say that S is an egocentric causal learner if S is capable of learning contingencies between S's own actions and outcomes caused by those actions, as when a baby learns that kicking its foot will move a mobile to which the foot is attached. Ordinary operant or instrumental conditioning (or learning) falls into this category, but I leave it as an open question whether there are egocentric forms of causal learning that do not involve operant conditioning[20].

---

[18] Again, it is very tempting to ask whether the subjects in these experiments have "causal knowledge" of the fact that (or a concept of cause according to which) falling objects can't pass through solid surfaces? Some researcher say "yes", attributing the search failures to "performance" error. However, this seems ad hoc. Again, I recommend the view that the subjects have some aspects of adult human causal competence but not others. Appreciation of the relevance of considerations having to do with impenetrability for planning and action is part of adult causal understanding of impenetrability.

[19] Here A and O are of course types of events or properties, rather than variables.

[20] Peacocke (this volume) considers an animal that has "grasped" a "goal action pair" of the form "to get G, do A". Taken literally (and this seems to be how Peacocke understands the idea) this requires only that A is a sufficient condition (or perhaps a conjunct in a sufficient condition) for G. The egocentric causal learner that I am envisioning is different from this—it grasps or represents the *contingency* between

2) S is an agent causal learner if S can learn about causal relationships both from action/ outcome contingencies involving its own interventions and those of other agents and if S is able to integrate this information in the sense that S is able to appreciate that the outcomes of other's interventions have implications for what would result from S's own interventions and vice-versa. In other words, S is able to learn, by observing that some other subject S*'s action A produces outcome O, that S itself could produce O by doing A and also able to learn that if its own actions A produce O, other subjects can also produce O by doing A.

3) S is an OA (observation/action) causal learner if S learns action/ outcome contingencies involving its own actions, action/outcome contingencies involving the actions of other agents and contingencies deriving from patterns of covariation in nature that may occur "naturally", rather than being produced by any agent and suitably integrates these, regarding each as a source of information about the other. In other words, S is both an agent causal learner and regards the results of observational learning not involving agents as relevant to its own action-outcome learning.

I assume that adult humans are, at least in principle, OA learners. Indeed, whatever else may be true of adult human conceptions of cause and adult human causal representations, it seems uncontroversial that adult humans think of causal relationships in terms of the assumption that the very same kind of causal relationship can be present between their own actions and outcomes, between other's actions and outcomes, and between naturally occurring events not involving the actions of any agents. Thus when I put water on a plant and it grows, when I observe you put water on a plant and it grows and when I observe rain falling on a plant and it grows, I assume that the very same sort of causal relationship between water and plant growth is present in all three cases and that observation of any one of these cases can furnish information about the others.

---

whether it does A and whether G is achieved—in other words, it grasps that if A is done, G follows and if A is not done, G does not follow. It seems unlikely that it would useful in most circumstances for animals to have representations merely to the effect that A is sufficient for G, since this is consistent with A's being irrelevant to G, and G's occurring whether or not A occurs. In fact empirical studies show that what is learned and represented even in instrumental conditioning is contingency information rather than mere sufficiency information, so that the latter possibility seems irrelevant to causal learning on empirical as well as conceptual grounds . Peacocke also notes that the obtaining of relationships of form "If A, then G, and if not A, then not G" is consistent with the relationship between A and G not being causal, as in the case in which A and G are joint effects of a common cause. Of course this is correct, but, as explained above, in this case it will not be true that intervening on whether A is performed will be associated with whether G is realized. My suggestion is that we think of the egocentric causal learner as grasping or representing or being guided by claims like "If I intervene to do A, then G" and "If do not intervene to do A, then not G". The truth or falsity of these claims does track whether A causes G. In other words, the egocentric causal learner is sensitive to the difference between A's causing G and a mere correlation between A and G, although the egocentric causal learner does not possess a full-fledged adult human notion of causation.

However, it seems logically or conceptually possible for a creature to be an egocentric causal learner only and not an agent causal or OA learner. Such a creature would be able to learn contingencies linking its own actions to the outcomes they produce, but either (i) would not be able to learn about causal relationships from the interventions of others or from covariational information not involving other agents or (ii) would not be able to put together what it learns from its own interventions with what it learns from covariational information from other sources. Thus such a creature would not, for example, infer from covariational information from other sources to how such information might be relevant to producing desired outcomes from its own interventions.

One might think of such an egocentric learner as learning that (or as having representations to the effect that) if I do X, goal G results and if I don't do X, G does not result, but not as capable of learning or representing that the very same relationship which is present between X and G when it does X can also be present when another agent does X or when X occurs "naturally".

It also seems conceptually or logically possible that a creature might be an agent causal learner only in the sense that while it can learn from its own interventions and those of other agents and can represent that the very same relationship which is present between X and G when it does X can also be present when another agent does X, but does not learn from observations not involving other agents and does not represent that the same relationship between X and G which is present when another agent does G can also be present in nature, independently of the activities of any agent[21].

Indeed, these are not just logical possibilities. Tomasello and Call (1997) suggest that, as a matter of empirical fact, apes are not OA causal learners in the sense described above, even though they are presumably (at least) egocentric causal learners -- that is, although they learn from the results of their own interventions, and also track naturally occurring covariation, they do not move back and forth between these, applying the results of observations of naturally occurring covariation to the design of their own interventions:

> we are not convinced that apes need to be using a concept of causality in the experimental tasks purporting to illustrate its use, at least not in the humanlike sense of one independent event forcing another to occur. More convincing would be a situation in which an individual observes a contiguity of two events, infers a cause as intermediary, and then finds a novel way to manipulate that cause. For

[21] To avoid confusion let me emphasize that what characterizes a agent causal learner is the sources of information from which such a learner can learn. The notion of an agent causal learner is not meant to suggest that the adult human notion of causation is somehow reducible to or acquired just from the experience of agency, as is advocated by agency theories of causation. There are many reasons, discussed in Woodward, 2003 for rejecting such a view. Indeed, the point of introducing the notion of an agent causal learner is to make it clear that such a learner possesses something less than the adult human notion of causation. However, I take it to be fully consistent with this that learning from one's own interventions and by observing the results of the interventions of others plays an important role in the acquisition of the full adult capacity for causal reasoning—as I note below, there is a great deal of empirical evidence that this the case.

> example, suppose that an individual ape, who has never before observed such an event, for the first time observes the wind blowing a tree such that the fruit falls to the ground. If it understands the causal relations involved, that the movement of the limb is what caused the fruit to fall, it should be able to devise other ways to make the limb move and so make the fruit fall. … we believe that most primatologists would be astounded to see the ape, just on the bases of having observed the wind make fruit fall, proceed to shake a limp, or pull an attached vine, to create the same movement of the limb[22]. (1997, p. 389)

It is interesting to compare the prediction in this thought experiment with the results from a real experiment. In a study conducted by Blaisdell et al. (2006) rats were first exposed to an observational learning phase (that is learning that did not involve interventions) in which, it was claimed, the rats acquired a common cause model in which a light (*L)* , was represented as a common cause of both a tone (*T)* and whether or not food was present (*F*): $T \leftarrow L \rightarrow F$. In subsequent tests, when the rats were presented with the tone, they behaved as though they believed that food was present (they increased their search for food, as measured by nose poking), which is of course consistent with their adoption of the common cause model. In the next, crucial "intervention" phase of the experiment, a lever was introduced, the pressing of which by the rats caused the tone to be presented. In this case, the rats were less inclined to search for food after tones caused by the lever press, despite the fact that tone and food were associated in the observational phase of the experiment. This of course is the normatively appropriate behavior if the rats grasped the causal structure of the situation they were dealing with: intervening on the tone "breaks" the connection between the tone and the light and renders the tone statistically independent of the food presentation[23]:

$$I \rightarrow T \qquad L \rightarrow F$$

If these experimental results are taken at face value, they do show, as the authors claim, that rats can, in some respects "distinguish between causal and spurious correlations" and "that they are capable of deriving predictions for novel actions after purely observational learning" (Waldmann, Cheng, Hagmayar, and Blaisdell, 2008, p. 469), although the "prediction" in this case concerns the absence of a correlation between their lever presses and the presentation of food. However, this experiment does not

---

[22] The reader may note the assumption in the first sentence of this passage that the "humanlike " concept of causality involves the idea of one event "forcing" another to occur and that apes lack this. However, the particular deficit imagined by Tomasello and Call is, on the face of things, a failure on the part of the apes to move from observation to the design of their own interventions. As argued earlier, it is unclear how Tomasello's and Call's claim about the ape's lack of a representation of causes in terms of forces relates to this failure.

[23] This relies on the "arrow breaking" interpretation of interventions, according to which intervening on a variable "breaks" all of the arrows directed into the variable (in a direct graph representation of the causal relationships into which the variable enters) – in other words, the variable is rendered independent of its previous causes. See Spirtes, Glymour and Scheines, 2000 and Woodward, 2003 for additional discussion.

provide evidence that rats can design interventions to achieve novel goal objects on the basis of purely observational information--- that is, that they can do what Tomasello and Call describe in their thought experiment, inferring, e.g., on the basis of observations of the wind shaking the tree branches and the fruit fall that if they were to shake the branches, that would make the fruit fall. It is this latter sort of integration that is required for full OA learning[24].

6. Explicitness of Representation.

I turn now to an additional distinction which has to do not with which causal or contingency relationships a subject represents but rather with how explicit (rather than merely implicit) these representations are. The general contrast between explicit and implicit representation is (to say the least) not terribly clear, but the intuition I mean to invoke is this: Consider a baby who acquires the ability to make a rattle sound by kicking her foot, to which the rattle is attached. What is learned in this case is an action/ outcome contingency but it is a further question whether the baby is consciously aware of what is learned or explicitly represents or conceptualizes it in a form which allows it to combined with other sorts of information. It seems possible or even plausible that the baby instead may have acquired only "implicit" or "procedural" knowledge of the action/outcome contingency, rather than an "explicit" representation of it. A similar point probably holds for at least some behavioral routines acquired as a result of instrumental conditioning – these will also involve implicit rather than explicit representations of dependency relationships. This contrasts with a case in which an adult human learns on the basis of her own interventions that flipping a light switch will turn a certain light on and off—in this case the adult presumably will explicitly conceptualize the relationship between switch and light as causal, will be able to communicate this information to others, use it in a variety of different sorts of reasoning and planning, and so on.

But what does this "implicit" versus "explicit" contrast amount to? One possibility would be to tie a subject's possession of explicit causal representations to the ability to offer explicit verbal reports about the representation or to use it in explicit verbal reasoning. This seems plausible as a sufficient condition for possession of an

---

[24] Blaisdell et al also trained rats on a "chain" model in which (according to the authors) the rats acquire a representation in which *T* causes *L* which in turn causes *F*. They do not report that after observing the association between *T, L* and *F*, the rats spontaneously intervene on *T* in order to get *F* , which they presumably want. It is this sort of observation that would provide evidence that the rats produce interventions to achieve goal objects on the basis of purely observational information.

Also relevant to this question are the results in Fawcett et al. (2002) Starlings who were presented with a trained conspecific which performed one of two actions in manipulating a plug on a bottle in order to obtain a food reward were more likely to reproduce that action themselves in order to obtain the reward. However, when presented with a "ghost" condition in which the plug moved in similar ways spontaneously, the starlings did not act so as to produce the plug motion they had seen. The starlings thus behaved like agent causal but not OA learners. Thanks to Alison Gopnik for drawing my attention to this paper.

explicit representation but is very restrictive when taken as a necessary condition, since it would mean that pre-verbal children and non-verbal animals do not, as a matter of definition, possess any explicit representations. An alternative would be to say that a representation is more or less explicit according to whether it is more or less informationally incapsulated, more or less available to other systems for inference, reasoning, and planning, and more or less integrated with or connected to other knowledge and representation the subject possesses. One might also tie the explicitness of a representation to the extent to which its possession facilitates insight or rapid as opposed to slow, trial and error learning[25].

With this second understanding of "explicit", one might conjecture that causal representations are likely to be more explicit to the extent that they enable or are associated with OA learning, and likely to be more implicit to the extent that they are associated only with egocentric causal learning. (Thus to the extent that the baby in the example above has learned only an egocentric representation, one might conjecture that this is likely to be only implicit) To some extent the correctness of this conjecture is guaranteed simply by the fact that OA representations are by definition more integrated in some respects than those associated with causal learning that is only egocentric, but the conjecture arguably has additional empirical content in suggesting that a certain set of abilities cluster together -- OA learning, availability of causal information for a variety of planning and reasoning tasks, and use of such information to facilitate rapid "insight" learning.

One possible reaction to the suggestion that subjects may have "implicit" causal representations is that this is an oxymoron: whatever else may be true of such representations, it is not appropriate to think of them as "causal" in any full-blooded sense. Instead, we should think of them as mere behavioral routines or patterns of conditioned responses that subjects acquire. However, as a number of writers have emphasized (see especially Dickinson and Shanks, 1995), instrumental conditioning in particular has a number of features that are also characteristic of full fledged human casual learning and judgment with explicit representations—most fundamentally,

---

[25] Goldberg (this volume) draws attention to the fact that for some simple tools, including tools that are unfamiliar, there is (for normal adults) a "transparent" relationship between the structure of the tool and "mechanical reasoning" about its appropriate use—one can read off whether the tool may be effectively used in certain way just from its structure, without the need for trial and error learning. For example, a normal adult who is unfamiliar with forks should nonetheless be able to "see" immediately that a fork cannot be used to eat soup. Patients with left parietal damage are compromised in their ability to engage in such reasoning. The ability to select an appropriate tool just on the basis of its structure prior to actually experimenting with it on a trial and error basis is a good example of insight learning. It also illustrates the interconnections in normal adult causal thinking between awareness of geometrical/mechanical properties (structural properties) and their possible role in "difference-making" and manipulation—a normal adult can infer immediately from the perceived shape of a fork that it will not be usable in transferring liquid to a new location.

sensitivity to difference-making information and temporal delay, but in addition, e.g., sensitivity to discounting and blocking effects of various kinds[26].

Following the methodology described earlier, rather than dismissing such learning as irrelevant to causal learning and cognition because if does not have all of the features of adult human causal cognition, I instead recommend seeing such learning as embodying some but not all of the elements of full fledged casual thinking. (It is possible, I suppose, that adult human causal thinking does not in anyway build on or make use of the capacities that are involved in instrumental conditioning but in view of the similarities between the two this seems like a conclusion that requires a strong supporting argument rather than something that should be assumed by default.)

7. Agent Causal Learning Again.

Recall that an agent causal learner integrates information about difference-making or contingency relationships between its own actions and the outcomes they produce and information obtained from observing contingencies between the actions of other agents (usually conspecifics) and the outcomes these actions produce. However, an agent causal learner need not be able to move from observations of causal relationships as they occur in nature without the involvement of any agent to use such information to guide its own interventions.

Meltzoff (2007) suggests that the notion of an agent causal learner may provide a " reasonable description of the pre-linguistic toddler". Meltzoff draws attention to the fact that humans are much better at imitation and at learning from imitation than non-human animals, including other primates[27]. Moreover such abilities emerge very early among

---

[26] See Woodward, 2007 for additional discussion.

[27] Tomasello (1999) and a number of other writers distinguish between imitation in the sense of high-fidelity copying of both ends and means (including causally superfluous elements) and "emulation" , where a subject tries to reproduce an observed end but not necessarily through the observed means. In my remarks above, I use "imitation" in a broader sense to cover both high fidelity copying of ends and means and attempts to reproduce observed ends that may employ appropriately modified means. Tomasello suggests that humans, including human infants, are more likely than non –human primates to imitate, while non-human primates, to the extent that they reproduce observed agent –outcome sequences, are more likely to emulate. I won't try to comment on this claim here except to observe that the infants in the experiments of Meltzoff's described below seem to engage in both imitation in Tomasello's sense and to employ modified means when this is appropriate. However, a few remarks on imitation and emulation as strategies for learning causal relationships may be helpful by way of orientation to what follows. Notice first that highly fidelity imitation in Tomasello's sense of a means/end sequence can be achieved without any real grasp of the causal relationships between means and ends-- indeed, one may think of this is an advantage of high fidelity imitation over forms of learning that do require such understanding. One may conjecture that causal relationships are sometimes learned by first reproducing them through high fidelity imitation without much understanding and then only gradually coming to a more detailed understanding of them. One obvious limitation of Tomasello –type imitation,

humans—according to Meltzoff some form of these abilities may well be present at birth. To the extent that a subject is able to imitate the manipulative activities of another agent by observing that agent, this provides a route to the learning of causal relationships through the observation of the interventions of others, which is just what agent causal learning involves. Meltzoff suggests that the ability to imitate and to learn from the interventions of others is based on the fact that infants as well as adults represent both the perception of the actions of others and production of the same actions by themselves in terms of a common cross-modal or amodal code—i.e., an abstract, non-modality-specific code that captures what is common to the visual perception of other's actions and to the proprioceptive/ kinesthetic experiences that subject have of their own actions. This allows subjects to move readily from observations of other's actions to their own performance of such actions. Thus, according to Meltzoff, humans are never in the predicament of purely egocentric causal learners although other animals may well be.

Experimental results presented by Meltzoff show that even very young children readily learn how to perform novel manipulations by observing adults perform those manipulations —they learn this from one or very few observations rather than requiring extensive trials. For example, fourteen month olds learned from watching an adult model how to activate a blicket detector by touching it with their forehead. Eighteen month olds learn to pull apart a dumbbell from watching an adult do this. When this was made difficult for them to do (the ends of the dumbbell were glued together) they tried a variety of alternative means in attempting to separate the ends. When given dumbbells which were too large for them to separate in the way demonstrated by an adult model, they adopted alternative hand grips which were easier for them to execute. Interestingly, the children also pulled the dumbbell apart successfully when they observed adults trying to do this and failing—they inferred the adult's intentions or goals and reproduced a successful version of the action the adult attempted to perform, rather than the unsuccessful version[28]. Strikingly, however, the children did not do this when presented

---

though, is that high fidelity copying may not lead to successful achievement of goals if imitator is of different size, strength, in somewhat different circumstances etc. Emulation, if achieved, can avoid these difficulties but presumably it is most likely to be successful when whatever means the subject employs to achieve end E is already within its behavioral repertoire. Attempts at emulation are presumably less likely to be successful when a subject does not know any means to produce E and must learn this by observing the behavior of an other. As the examples discussed below illustrate, successful human causal learning, including learning by infants, often seems to be both sensitive to the means other agents employ and not just to the ends they achieve but, unlike Tomasello-style imitation, also shows awareness of how means might be varied and still achieve the same end.

[28] Note that the children's ability to vary their behavior in normatively appropriate ways so as to achieve the goal states in these experiments, using alternative means as circumstances change suggests that they have learned something more than single isolated action/ goal sequences of the form, "If I do A in C, G will result". Instead what they have learned is something more like "If C_1, then if I do A_1, G will result", "If C_2

with a mechanical device that unsuccessfully attempted to pull the ends of the dumbbell apart. Similarly, the children did not produce the action of pulling the dumbbell apart if they were merely presented with snapshots showing the object assembled and then disassembled or if they observed the object appear "spontaneously" disassemble and then assemble again, without human intervention.

When presented with an adult model, the children thus modified the means they employed in appropriate ways to achieve the goal of the adult model, which suggests some causal understanding of how different means contribute (or not) to the end state they were trying to achieve, as well as recognition of that end state even when the adult model failed to achieve it. However, the presence of an agent to serve as this model seemed to be crucial to the children's causal learning[29].

Taken together, these results suggest the possibility that there is a stage in human development in which observation of the actions of other agents plays a crucial role in learning about causal relationships and manipulative possibilities -- a stage in causal learning in which infants are agent causal but not yet full OA learners. In this stage, children's skills in parsing the actions of others, inferring intentions, and imitation are an important source of access for learning about causal relationships, where this information may be difficult for them to acquire through observation of events not involving other agents. (One might think of this as a matter of skills that are important for social cognition more generally being co-opted for learning about non-social causal relationships). Obviously, a creature with the ability to learn about causal relationships from observing the manipulative activities of others has a great advantage in causal leaning over a creature who can learn only from its own interventions, but not those of others. Such learning from others of course plays an important role in human tool use, as both anthropological and experimental evidence suggests. It also suggests the possibility that, at least to some extent, limitations in the tool use abilities of non-human animals (in comparison with humans) reflect the limitations of the former as agent causal learners – the route to learning about causal relationships by observing the manipulative activities of others that humans are able to exploit so effectively is less available to non-human animals[30]

---

then if I do A_2, G will result" etc. I see this sort of flexibility in selecting the appropriate means to a goal under changing circumstances as one element involved in possession of a causal representation.

[29] More recent experimental results reported in Bonawitz, Ferranti, Saxe, Gopnik, Meltzoff, Schulz and Woodward (2010) provide additional support for these claims, showing that children learn to produce desired outcomes from their own interventions much more readily when they observe the interventions of an adult model than when they observe relevant contingencies that do not involve interventions.

[30] Peacocke (this volume) argues that successful imitation does not require genuinely causal representation of what is imitated. If this means simply that a creature can have the ability to imitate without having the full adult human notion of causation or an adult human understanding of the causal relationships mediating the process that is imitated, then I agree. I would emphasize, however, that it doesn't follow that imitation plays no role in learning causal relationships or in the eventual acquisition of full adult human

8. Means/ Ends Decoupling versus Fused Action/Outcome Representations.

Another relevant consideration which may be related to explicitness of representation and which is naturally highlighted by an interventionist approach is the extent to which there a subject's causal representations "decouple " means and ends or instead fuse them in a single representation and the extent to which representations of means themselves are decoupled into representations of more proximate and distal causes. To the extent that there is such a separate representation of means and ends and of intermediate links in causal chains involving means, this is likely to be associated with greater flexibility in causal learning and behavior. The thought experiment above described by Tomasello and Call again provides one illustration of some aspects of this. Here the desired goal state is the falling of the fruit. As reflected in the diagram below (taken from Tomasello and Call, 1997), the proximate cause of this goal is the vigorous movement of the limb ("Limb Shakes" in the diagram-- this is the means by which the goal state is produced) and this in turn may be caused in three different ways or through three different means—through the manipulations of self, others, or through the activity of the wind. The movement of the limb is a common intermediate step in all three of these causal chains. Recognition that there is a such a common intermediate step that can be produced in any one of three different ways may help to make it easier to regard the observation of one such chain as a source of information about the others. Also, once it is recognized that there is such an intermediate step that immediately precedes the goal state, this opens up the possibility that may be still other novel ways of producing that intermediate step and thus the goal state.

---

ways of thinking about causal relationships. Indeed imitation can facilitate such learning exactly because successful imitation does not require that one already possess full causal understanding of what is imitated. Moreover, while imitation does not always require fully causal representation, I think that Meltzoff is correct to think that certain kinds of imitation—for example, those that involve flexible and appropriate modification of means in the face of changing circumstances -- are suggestive of the acquisition of at least one important element in causal representation and understanding. As I see it, there is a range of cases between "blind" imitation with little or no causal understanding and imitation that is mediated by a sophisticated causal understanding of what is imitated, with the former sometimes (at least in humans) playing a role in the acquisition of the latter.

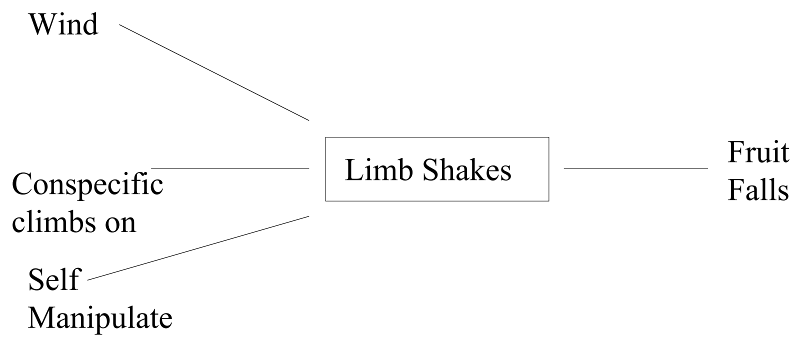

Figure 1

Consider, by contrast, an animal that only has a representation connecting its own action on the limb to the production of the desired goal state but does not decompose this causal sequence into intermediate steps in the manner described above. In other words, the animal has only a representation of the form, "If I do *X*, desired outcome *G* results" with no representation of intermediate causal links or means leading from *X* to *G*[31]. In Tomasello's and Call's example, this animal's representation would look something like this

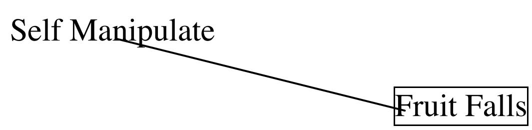

Figure 2

or perhaps like this

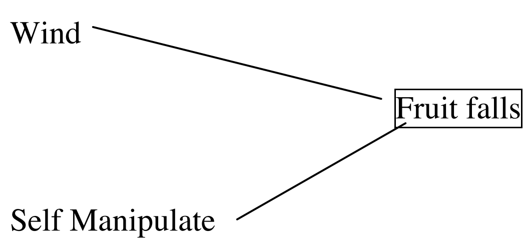

Figure 3

[31] *X* and *G* here are variables.

if the connection between wind movement and fruit falling is also represented.

To the extent that the animal is guided just by a representations like that in Figure 2, it presumably would have no reason to expect that if another animal or the wind were to shake the limb, this would lead to fruit on the ground. Even if the animal has a representation like that in Figure 3, there is nothing in the representation that suggests that the way in which its own manipulations produce the goal state and the way in which the movement of the wind produces the goal state have something in common. Thus, (in the absence of some representation of the common intermediate step, "Limb Shakes") there is nothing to suggest how, e,g., observation of the "wind moves /fruit falls" sequence bears on the issue of how the animal might obtain fruit through its own actions. In other words, to the extent that the animal learns only particular action/ outcome sequences (and representations fusing these together, with means or intermediate steps not represented), these are likely to be isolated, with their interconnections and interrelationships unrecognized, rather than integrated.

There are other related cases in which an animal's behavior may also suggest a failure to adequately represent means/ends relationships or to understand how a particular set of means contributes to an end (when it does). Consider an animal that behaves as if it recognizes that the production of action A through the introduction of means M in certain circumstances C is sufficient for some desired outcome O, as when a pigeon pecks and obtains a desired food reward or a primate learns that pushing a food item with a stick will dislodge it from a transparent tube. In effect, the animal has in A a specific behavioral recipe or routine that is sufficient for the production of O in circumstances C. Intuitively, however, this sort of capacity is consistent not only with the animal's failing to recognize that there may be alternative means besides M for the achievement of O, but also with its failing to employ those alternative means when circumstances change and M is no longer appropriate for achieving O. It is also consistent with the animal's behavioral routine containing elements that are not necessary for or superfluous for the achievement of O. More adequate means/ends understanding would allow the animal to recognize when these superfluous elements are present and also to recognize when changed circumstances have the consequence that the original means are no longer appropriate.

As an illustration, consider a primate who has learned to move a food item in a tube in a direction away from a trap in the tube. If this behavior persists even though the tube is rotated so that the trap is on top and can no longer trap food, then the animal is employing a behavioral routine that contains elements that are in the circumstances causally superfluous or unnecessary for the outcome it desires, although they are sufficient for that outcome. Similarly consider an animal which has learned to insert a stick into the end of a tube which is closest to the food trap in order to remove a food item—a strategy which, as it happens, works because the food item is always between the trap and the end of tube farthest from the trap. If the animal persists with this strategy even when the food source is between the trap and the end in which the stick is inserted, the animal again is employing a behavioral strategy which is sufficient for the goal in some circumstances, but not others. In both cases, it seems natural to describe the animal as lacking (what in human terms would be described as) any deep "causal understanding" of the relationship between the behavioral routine A and the outcome O it produces. Put differently, there is a lack of means/ends understanding in the sense that the animal does not grasp what it is about the means it employs that makes these effective or does not understand why the means are effective (or not) when they are.

The results of these experiments may be contrasted with the experiments conducted by Meltzoff described above. Recall that when it was difficult for the children to pull the dumbbell apart in the way demonstrated by the adult model, they tried a variety of alternative means to separate the parts. They also correctly inferred the adult's goal even when the adult was not successful in pulling the dumbbell apart. In addition to illustrating the role that other agents play in children's causal learning, these experiments also suggest that the children have some capacity to decouple means from ends and some sensitivity to when different means are effective or not in achieving those ends. They exhibit, in short, some "causal understanding" of means/ends relationships.

But what exactly is involved in such "causal understanding" or in grasping or failing to grasp "what it is about the means employed that makes these effective or ineffective"? I noted above that in attempting to unpack what the quoted phrases mean a popular move is to appeal to notions like "unobservable mechanism" or "generative force" – the idea being that human beings think in terms of these concepts in representing the relationship between means and ends, but the non-human animals in the experiments described above do not and this accounts for the difference in their performance.

For reasons already discussed, I am not convinced that this invocation of representations of mechanisms and forces (or their absence) really helps to illuminate the patterns of success and failure in human tool use. A simpler and less tendentious description of the limitations in the animals' behavior is that they fail to incorporate appropriately detailed information about difference-making relationships in their behavior (e.g., they fail to recognize that certain elements in the behavioral routines do not make a difference for their goals and are hence superfluous or that means that are difference-makers in some circumstances are not in others.) Mere possession of an abstract representation of causal relationships in terms of forces or hidden mechanisms does not by itself provide such detailed difference-making information and there is no obvious reason why an animal cannot behave as though it is sensitive to such information and incorporate it in its actions, even if it lacks a forced-based representation of causal relationships. It seems more straightforward to suppose that "greater causal understanding" and "a better grasp of means/ends relationships" in this sort of context just amounts to, e.g., the possession of routines and representations which guide behavior in such a way that there is sensitivity to what is necessary and not just sufficient for a desired outcome. Similarly, for representations that guide behavior so that it changes appropriately in such a way that it continues to produce desired effects in the face of changing circumstances. This allows us to capture the idea that such an animal has capabilities that incorporate something more than just a routine sufficient for a goal in a specific set of circumstances, without committing ourselves to the idea that this "something more" consists in representations of forces and hidden mechanisms.

9. Conclusion

Although this is a longish paper, I've still left a great deal out. One additional consideration, hinted at in portions of my discussion above but also not discussed in the detail it deserves, is that adult human causal representation involves the ability to put together or integrate information about individual causal relationships into an overall model, to export such information into new contexts, and (at least to some extent) to anticipate what will happen when the relationships represented in such models are modified. Thus a human cognizer who recognizes that smoking causes yellow fingers and that smoking causes lung cancer is also able to represent that the same factor, smoking, can act as a common cause of both, that steps taken to prevent finger-yellowing while still smoking will not affect lung cancer and so on. Similarly, as we saw in connection with the Tomasello and Call thought experiment, human cognizers can represent causal chain structures and what will happen under modifications of such structures in which intermediate variables have different causes.

Finally, adult human causal representation often has a complex hierarchical structure (cf. Tenenbaum et al., 2007). Particular causal relationships are grouped together into more abstract categories in a way that greatly facilitates causal learning and inference. For example, in addition to whatever knowledge we humans may possess about the causal relationships involved in particular diseases –colds produce runny noses, parasites produce problems with the digestive system and so on-- we also have the more general abstract category of disease itself, the idea that this maybe produced by invisible pathogens, the idea that particular diseases have characteristic causes and effects and so on. This allows us to organize our causal knowledge and also suggests, when we encounter a new disease, that we should look for its characteristic symptoms and causes. Presumably these sorts of highly structured representation are heavily dependent on language and capacities for abstraction that may be uniquely human.

References:

Blaisdell, A., Sawa K., Leising, K., Waldman M. (2006) “Causal Reasoning in Rats”, *Science* 311: 1020- 1022.

Bonawitz, E., Ferranti, D., Gopnik, A., Meltzoff, A. Woodward, J. and Schulz, L. (2010) “Just do it? Toddlers’ Ability to Integrate Prediction, Action and Expectation about Contact Relations” *Cognition*. 115(1):104-17.

Call, J. and Tomasello, M. (1997) *Primate Cognition*. New York: Oxford University Press.

Campbell, J. “Intelligent Tool Use” This volume.

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

Dickinson, A. and Shanks, D. (1995) “Instrumental Action and Causal Representation”. In Sperber, D., Premack, D, and Premack, A. (eds.) *Causal Cognition*. Oxford: Oxford University Press.

Dickinson, A. and Balleine, B. (2000) "Causal Cognition and Goal Directed Action". In C. Heyes and L. Huber, eds. *The Evolution of Cognition*. Cambridge, MA: MIT Press.

Dowe, P. (2000) *Physical Causality*. Cambridge: Cambridge University Press.

Eells, E. (1991) *Probabilistic Causality* (Cambridge: Cambridge University Press.

Fawcett, T., Skinner, A. and Goldsmith A. (2002) "A Test of Imitative Learning in Starlings Using a Two-action Method with an Enhanced Ghost Control" *Animal Behaviour* 64: 5 47 – 556.

Gopnik, A. and Schulz, L. 2004 "Mechanisms of Theory Formation in Young Children" *Trends in Cognitive Science* 8: 371- 377.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L. Kushir, T. and Danks, D. (2004) "A Theory of Causal Learning in Children: Causal Maps and Bayes' Nets" *Psychological Review* 111: 3-22.

Hall, N. 2004, "Two Concepts of Causation" in Collins, J., Hall, N., and Paul, L. eds. *Causation and Counterfactuals* Cambridge, MA: MIT Press.

Hauser, M. (2001) "Searching for Food in the Wild: a Nonhuman Primate's Expectations about Invisible Displacement", *Developmental Science* 4: 84-93.

Hitchcock , C. (1995). "Discussion: Salmon on Explanatory Relevance" *Philosophy of Science* 62: 304-320.

Hitchcock, C. (2001) "The Intransitivity of Causation Revealed in Equations and Graphs" *Journal of Philosophy* 98:273-299.

Keil, F. (2003) "Folk Science: Coarse Interpretations of a Complex Reality", *Trends in Cognitive Science* 18, 663-692.

Leslie, A. (1995) "A Theory of Agency". In Sperber, D., Premack, D. and Premack, A. (eds.) *Causal Cognition*. Oxford. Oxford University Press. pp 121-141.

Lewis. D. 1973 "Causation" *Journal of Philosophy* 70: 556-67

Meltzoff, A. (2007) "Infants' Causal Learning" (2007) In A. Gopnik and L. Schulz, *Causal Learning: Psychology, Philosophy, and Computation*. New York: Oxford University Press.

Peacocke, C. "Representing Causality" (This volume)

Pearl, J. 2000. *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University. Press.

Povinelli, D. 2000. *Folk Physics for Apes*. Oxford: Oxford University Press.

Salmon, W. (1984) *Scientific Explanation and the Casual Structure of the World*. Princeton: Princeton University Press.

Santos, L., ‘‘Core Knowledges: a Dissociation between Spatiotemporal Knowledge and Contact-Mechanics in a Non-Human Primate?”, *Developmental Science* 7 (2004), 167–74.

Santos, L. and Hauser, M., “A Non-Human Primate’s Understanding of Solidity: Dissociations Between Seeing and Acting”, *Developmental Science* 5 (2002), F1-F7.

Santos, L., Seelig, D. and Hauser, M. (2006) “Cotton-Top Tamarins' (Saguinus oedipus) Expectations About Occluded Objects: A Dissociation Between Looking and Reaching Tasks” *Infancy* 9(2), 141–165

Schottmann, A. and Shanks, D. 1992, “Evidence for a Distinction Between Judged and Perceived Causality” *Quarterly Journal of Experimental Psychology* 44A: 321-42.

Spelke, E. S., Phillips, A., and Woodward, A., “Infants’ knowledge of object motion and human action” (1995). In Sperber, D., Premack, D. and Premack A. (eds.), *Causal Cognition: A Multidisciplinary Debate*. Oxford: Clarendon Press,

Spirtes, P. Glymour, C. and Scheines, R. (2000) *Causation, Prediction and Search*. Second Edition. Cambridge: MIT Press.

Tenenbaum, J., Griffiths, T., and Niyogi, S. (2007) “Intuitive Theories as Grammars for Causal Inference” In Gopnik, A., & Schulz, L. (eds.*), Causal learning: Psychology, Philosophy, and Computation*. New York: Oxford University Press.

Tomasello, M. (1999) *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Visalberghi, E. and Trinca, L. (1989) “Tool Use in Capuchin Monkeys: Distinguishing Between Performance and Understanding” *Primates* 30: 511-21.

Visalberghi, E., Addessi, E., Truppa, V., Spagnoletti, N., Ottoni, E., Izar, P. & Fragaszy, D. (2009). “Selection of Effective Stone Tools by Wild Capuchin Monkeys”. *Current Biology* 19, 213-217.

Waldmann, M., Cheng, P., Hagmayer, Y., & Blaisdell, A. (2008). “Causal Learning in Rats and Humans: A Minimal Rational Model”. In N. Chater & M. Oaksford (Eds.), *The Probabilistic Mind: Prospects for Rational Models of Cognition*, Oxford, Oxford University Press, pp. 453-484

Wolff, P. (2007) “Representing Causation”, *Journal of Experimental Psychology: General* 136: 82-111.

Wolpert, L. (2003) “Causal Belief and the Origins of Technology” *Philosophical Tranactions of the Royal Society* London 361: 1709–1719.

Woodward, J. 2003. *Making things Happen: A Theory of Causal Explanation*. 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