A Makeover of the Pure Causal Theory of Knowledge

1. Introduction

Causal theories of knowledge are of two different kinds: pure and impure. A pure causal theory of knowledge (PCTK) makes knowledge of a proposition *p* dependent on the existence of some direct, or surrogate causal connection holding between the fact that *p* and the belief that *p*. Impure causal theories of knowledge (ICTK) are variations of what Goldman calls “a causal reliability approach”. [Goldman, 1986] According to this approach whether a true belief is knowledge depends on the reliability of the method/process that causes the belief. For some versions of ICTK (ex. Goldman’s process reliabilism) a process is reliable when it is generally truth-conducive. A process is truth-conducive if the ratio of true beliefs it generates is higher than a conventionally fixed threshold.

This paper takes a new look at the long abandoned PCTK. My attempt to re-investigate the explanatory power of PCTK is motivated by two reasons. One is the serious problems arising with different versions of ICTK. The second is the interesting results of the recent research on the nature of causation. Regarding the latter, I will make primary use of Christopher Hitchcock’s analysis of causation through the help of the techniques of causal modeling. I will argue that if we employ Hitchcock’s analysis to cash out the causal connection holding between a belief and the fact believed to obtain, PCTK tracks our intuitions about knowledge as well as, or better than some alternative accounts of knowledge that remain “trendy”.

In what follows I will say a few words about the classic version of PCTK. Then I will introduce two cases, which pose a serious challenge to different versions of ICTK. After giving a summary of Hitchcock’s analysis of causation, I will use his proposal to give a solution to these cases by way of PCTK.

2. The origin and motivation for PCTK

The most prominent contemporary articulation of PCTK is Alvin Goldman's "A Causal Theory of Knowing." The motivation behind Goldman's essay was to correct the inadequacy of the traditional account of "S knows that p" in view of Gettier-type counterexamples. Goldman's diagnosis of Gettier type counterexamples is the following: In each case there is no connection between the belief that (p) and the fact that makes the belief that (p) true. This diagnosis suggests the following base clause for knowledge:

(BC): "S knows that *p* only if the fact that *p* is the cause of S's belief that *p*"

BC is attractive for different reasons. Its obvious appeal has to do with the plausible explanation BC gives for cases of perceptual knowledge. Yet, as pointed out by Dretske, its most fundamental attraction consists in BC's ability "to capture the intuition that a belief to qualify as knowledge must have no admixture of accidentality in its correspondence with the facts." [Dretske, 1984] This latter characteristic explains BC's ability to make the account of knowledge immune to Gettier-type counterexamples similar to the one just described.

The theories of causation have changed quite a bit since Goldman's "A Causal Theory of Knowledge". Our understanding of BC must reflect these changes in a way that allows for a full appreciation of what PCTK offers in the current epistemological debate. I will try to do just that by using Hitchcock's recent analysis of causation as a way of cashing out the causal relation made reference to in BC.

3. On the knowledge of our relatives

Consider the following two cases:

> **Grandma case:** Grandma Betty lies sick in bed. Her niece Ann Marie comes to visit her. Betty sees Ann Marie and believes that
> (a): Ann Marie is alive. Yet if Ann Marie would have been dead Betty would have still believed that Ann Marie was alive based on the testimony of her lying relatives who motivated by compassion would spare her the news of her granddaughter's death.
>
> **Father-case**: A father is already firmly convinced that his son is innocent of committing a crime, via faith in his son. Then he hears a conclusive courtroom demonstration of his innocence and is convinced by that too. Yet, if the courtroom demonstration showed that his son

> was guilty, (and the son was guilty) the father would believe that he was innocent via the method of faith in his son.[1]

The pronouncement of the common sense differs in the two cases. We are inclined to say that Betty knows that Ann Marie is alive, but that the father doesn't know that his son is innocent. The question is why. My answer will be the following: In the grandma case the belief that "the niece is alive" is caused by the fact that she is alive. In the father-case (Option A) the belief that "the son is innocent" is not caused by the fact that he is innocent. Grandma Betty knows that her niece is alive because her belief satisfies the causal condition made reference to in BC. The father fails to know because his belief violates that condition.

4. Hitchcock's analysis of causation

Hitchcock's analysis is primarily motivated by problems arising with Lewis's original counterfactual theory of causation. According to this theory if *c* and *e* are distinct, occurrent events, *e* counterfactually depends on *c* if and only if, if *c* had not occurred then *e* would not have occurred. If *e* counterfactually depends on *c* than *e* causally depends on *c*, or in other words, *c* is a cause of *e*. In Lewis' theory counterfactual dependence is sufficient but not necessary for causation. Event *c* is a cause of event *e* if and only if they are connected by a chain of counterfactual dependence. (In the simple case, the chain has just one link.). [Lewis, 1986] As Hitchcock indicates this formula renders causation transitive by definition. [Hitchcock, 2001] So, given Lewis's formula for any three events *a, b, c*, if *a* counterfactually depends on *b* and *b* counterfactually depends on *c*, then *c* is a cause of *a* even if *a* does not counterfactually depend on *c*. This picture is very appealing in the representation of ordinary cases. Yet problems arise when dealing with extraordinary cases such as "the dog bite case" [McDermott 1995], and "the boulder case" [Hitchcock, 2001]. A quick run through "the boulder case' will serve to illustrate the point:

> "A boulder is dislodged and begins rolling ominously towards Hiker. Before it reaches him, Hiker sees the boulder and ducks. The

[1] Both cases are supplied by Nozick in his "Knowledge and Skepticism." Nozick credits the father-case to Armstrong who in turn credits it to Gregory O'Hare.

> boulder sails harmlessly over his head with nary a centimeter to spare. Hiker survives his ordeal."

It is clear that in this case the Hiker's ducking is counterfactually dependent upon the boulder's fall and the Hiker's survival is counterfactually dependent on the Hiker's ducking. Yet we are reluctant to say that the boulder's fall *caused* the Hiker's survival, even though there is a chain of counterfactual dependence running from one event to the other. Hitchcock's preliminary diagnosis of cases similar to "the boulder case" boils down to the claim "that causation *is not* transitive in general." Meanwhile, his alternative proposal has the burden of accounting for those cases in which a chain of counterfactual dependence is sufficient for causation in a way which explains *why* a similar counterfactual chain is not sufficient in cases similar to the boulder-case.

Hitchcock's analysis makes use of what he calls "systems of structural equations" which are construed as ordered pairs $\langle V, E\rangle$ where E is a sequence of equations relating the values of the variables (usually representing properties or events) belonging to the set of variables V. A quick run through the details of these causal models, as described by Hitchcock, is necessary for understanding his analysis.

The elements of V represent events, or properties. They can be, both, exogenous and endogenous variables. Equations with an exogenous variable in the left hand side have the simple form X=x, where x is the actual numerical value for X. (If X is a binary variable then x =1 or x = 0 depending on whether or not X occurred.) Equations with an endogenous variable on the left hand side express the value of this variable as a function of the value of other variables in V. This latter class of equations encodes counterfactual dependences. Hitchcock upholds Lewis's "no-backtracking-counterfactuals" restriction in his analysis of causation. So, the switching of the variables from one side of the equation to another is not permitted.

The structural equations use sentential symbols to represent relations between variables. So,

$\neg X \equiv 1 - X$

$X \vee Y \equiv \max [X, Y]$

$X \wedge Y \equiv \min [X, Y]$

If a variable Y appears on the right hand side of an equation with an endogenous variable X on the left hand side, then Y is ***a parent*** of X. The system of the structural equations is given a graphical representation. The ***nodes*** in the graph are elements of the set of variables V. An ***arrow*** is drawn from a variable-representing node X to a variable-representing node Y, if X is a parent of Y. “A ***route*** between two variables X and Z is a ordered sequence of variables ⟨X, Y1, Y2….Yn, Z⟩ such that each variable in the sequence is in V and is a parent of its successor in the sequence.” A route between X and Z is graphically represented by a ***directed path*** which is a sequence of arrows lined up “tip to tail” connecting X with Z. A variable Z, which is distinct from both X and Y is an ***intermediate*** between X and Y if it belongs to some route between X and Y.

How do structural equation systems help Hitchcock deal with problem cases such as “the boulder case” within the framework of the counterfactual theory of causation? To illustrate Hitchcock’s proposal let us represent the causal structure of the “boulder case” using the structural equations apparatus developed so far. The causal graph for “the boulder case” is depicted in Fig. 1 below.

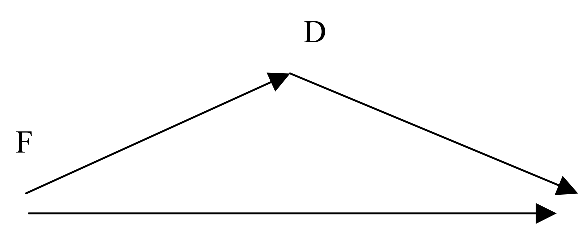

Fig. 1

F=1,0 depending on whether or not the boulder falls. D=1,0 depending on whether or not the Hiker ducks. S=1,0 depending on whether or not the Hiker survives. The set of structural equations E is the following:

F=1; D=F; S= $\neg$ F $\vee$ D

Given this set of structural equations, we can determine the value for S.

D=1

S= max [$\neg$ F $\vee$ D]

S= max [0, 1]

S=1

Hitchcock's central proposal is the following:

> "Let c and e be distinct occurrent events, and let X and Z be variables such that the values of X and Z represent alterations of c and e. respectively. Then c is a cause of e if an only if there is an ***active causal route*** from X to Z in an appropriate causal model ⟨V, E⟩." [Hitchcock, 2001]

A route from X to Z is **active** in the causal model in question if and only if Z **depends counterfactually** upon X within a new system of equations E1 constructed from E as follows: for all variables Y such that they are ***intermediate*** between X and Z but do not belong to this route, we replace the equation for Y with one that sets Y equal to its actual value in E. (If there are no intermediate variables that belong to this route, then E1 is just E)

The direct route from F to S is not active, because holding D fixed at its actual value, and changing the value of F from 1 to 0, the value of S remains the same as before, namely 1. Intuitively, if the Hiker had ducked while boulder had not fallen the Hiker would have survived. So, S doesn't counterfactually depend on F along this route. If we consider route ⟨F, D, S⟩ we see that there are no intermediate variables to hold fixed along other routes. If the value of F changes from 1 to 0, the value of D changes from 1 to 0, but the value of S= max [1-0, 0] remains the same, namely 1. Thus, S doesn't counterfactually depend on F along this route either. Intuitively, if the boulder didn't fall, then the Hiker's life wouldn't have even been put at risk, i.e. he would have survived. According to Hitchcock's theory the fact that there are no active routes from F to S, is sufficient to conclude that the boulder's fall didn't cause the Hiker's survival. The advantage of the theory consists in its ability to better track the pronouncements of the common sense in these problematic cases. In the next section I will use Hitchcock's proposal to examine the problematic cases introduced in Section 3.

5. Grandmother, Father and the Hitchcock graphs

Let us first draw the graph with the appropriate structural equations for the grandmother case. What we know is that the following counterfactuals are true:

Had the niece been dead she would not have visited. Had she been dead her relatives would have told Betty that she is alive. Betty would have believed that she was alive either by seeing her niece if she was alive, or by believing her relatives if she was dead. A=1,0 depending on whether niece is alive or dead. V =1,0 depending on whether or not niece visits. S=1, 0 depending on whether or not Betty visually perceives niece. R =1,0 depending on whether or not relatives lie to Betty about her niece. B=1,0 depending on whether or not Betty believes niece is alive. The set of structural equations will be E: $A = 1$, $V = A$, $R = \neg A$, $S = V$, $B = R \vee S$. The graph for the case is depicted in Fig 2 below:

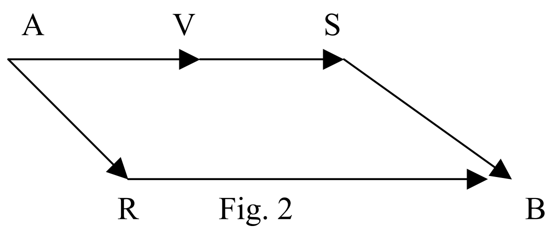

Fig. 2

The question is did the fact that Ann Marie was alive cause Betty's belief that she was alive? Given Hitchcock's proposal, we need to check whether there is an active route from A to B? First, we obtain the value of B by solving the set of equations in E. A=1, V =1, R=0, S =1 and B=1. Next, we need to check whether or not there is an active route from A to B. We start by selecting one of the routes from A to B and see whether it is active. Let's first try route $\langle A, V, S, B\rangle$. We change the value of A from 1 to 0, holding fixed the value of the intermediate variables between A and B along other routes, in this case the value of $R = 0$. Then we write a new set of equations with $A = 0$ and see whether the value of B changes. We get, $A = 0$, $R = 0$, V=0, S= 0, $B = 0$. Since the value of B changes, when we change the value of A, holding the value of R fixed, then B counterfactually depends on A on this new set of equations, which means that there is an active route, (namely, $\langle A, V, S, B\rangle$) from A to B. So, the fact that the niece is alive causes Betty to believe that she is alive. Using the same method, my intention is to show that the son's innocence is not the cause of the father's belief that he is innocent in the "father case".

There is an ambiguity in the father example that we need to clarify. The ambiguity results from the fact that we are given no information on how the father's deliberations work. Therefore we must consider the following two possible options:

Option A: The belief forming processes of faith in court-room demonstration and faith in one's son operate independently, and each of them operates in the father's case. So the father is convinced that his son is innocent both because of faith in his son and faith in courtroom demonstration.

Option B: The belief forming processes of faith in court-room demonstration and faith in one's son do not operate independently. The father is convinced by courtroom demonstration and not by faith in his son, though faith in his son would have been operating in the absence of courtroom demonstration. In this case the father's belief is formed via faith in courtroom demonstration and not via faith in his son.

Option B describes a case which is identical in structure to the "grandma case." So, given Option B we are inclined to say that the father knows that his son is innocent. Yet, our intuitions tell us that the father doesn't know if his deliberations match the story told in Option A. As I understand it, it is the story told in Option A that better matches the father's deliberations in the "father-case". We are told that the father already firmly believes that his son is innocent via faith in his son. Then, he hears the courtroom demonstration of his innocence and is convinced by that too. The case was originally introduced by Nozick, as one in which a person is convinced via two independently operating methods (one truth-tracking one not) of a particular fact p. So, it is very clear by the way in which Nozick describes the case that the father believes his son to be innocent *for both reasons*: faith in his son and conclusive courtroom demonstration. In what follows I will argue that our intuitions regarding the father's knowledge in Option A and B vary in accordance with the causal facts behind each option.

Let's first consider the system of structural equations for Option A.

S=1, 0 depending on whether or not the son is innocent. C=1, 0 depending on whether or not it is demonstrated in court that the son is innocent. Let P stand for the process/method/way by which father believes his son is innocent. P is not a binary variable. The father can believe his son is innocent by way of faith in his son, a combination of faith in his son and faith in courtroom demonstration, or he might not believe that his son is innocent by any method/process whatever, i.e. not believe that his son is innocent. The numerical values of P corresponding to these three alternatives are respectively 1, 2 and 0. B = 1, 0 depending on whether or not father believes son is innocent. The set of structural equations for this system is E: $S = 1$, $C = S$, $P = 1 + C$, $B = P \wedge 1$.

The father case is depicted graphically in Fig. 3 below:

Fig. 3

Solving the equations in E, we get, S=1, C=1, P=2, B=1. So, if the son is innocent the father believes he is innocent. In this case, however there is no active route from S to B. Route ⟨S, C, P, B⟩ is inactive, because if we change the value of S from 1 to 0, we get the following set of equations E*: S = 0, C = 0, P = 1, B = 1.

So even if the son was not innocent, the father will still believe he was innocent. This means that B doesn't counterfactually depend on S within E*. Given that ⟨S, C, P, B⟩ is the only route from S to B, the event represented by S (son's innocence) is not the cause of the event represented by B (father's belief he is innocent). According to BC the lack of this causal connection explains our inclination to claim that the father doesn't know that his son is innocent.

Now let's consider the system of structural equations for Option B:

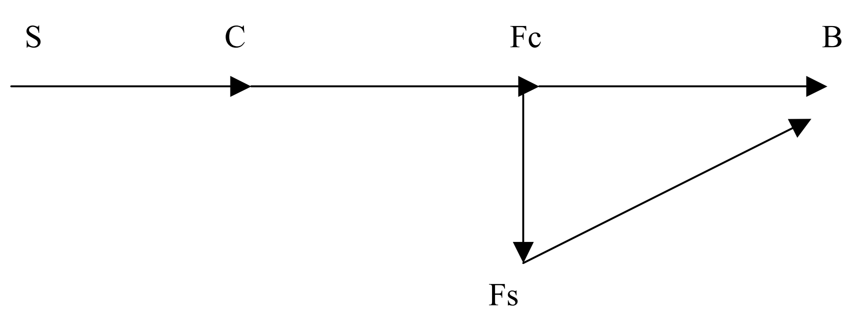

Fig. 4

Fs stands for “father believes son is innocent via faith in his son”, Fc stands for “believes son is innocent via faith in courtroom demonstration” The rest of the variables stand for the same events as in the earlier presentation. The set of equations for this case would be E: $S=1$, $C=S$, $Fc=C$, $Fs=\neg Fc$, $B=Fc \vee Fs$. In this case there would be an active route from S to B (route $\langle S, C, Fc, B\rangle$. In E we have $S=1$, $C=1$, $Fc=1$, $Fs=0$, and $B=1$. If we change the value of S from 1 to 0 keeping the value of Fs fixed ($Fs=0$) get the new set of equations E*: $S=0$, $C=0$, $Fc=0$, $Fs=0$, and $B=0$. So, B counterfactually depends on S along $\langle S, C, Fc, B\rangle$. This means that $\langle S, C, Fc, B\rangle$ is active which given Hitchcock’s proposal means that the son’s innocence causes the father’s belief in Option B. In view of BC the existence of this causal connection explains our intuition that the father knows in this case.

One might claim that there are other ways of graphically representing the father-case. A possible candidate can be the following;

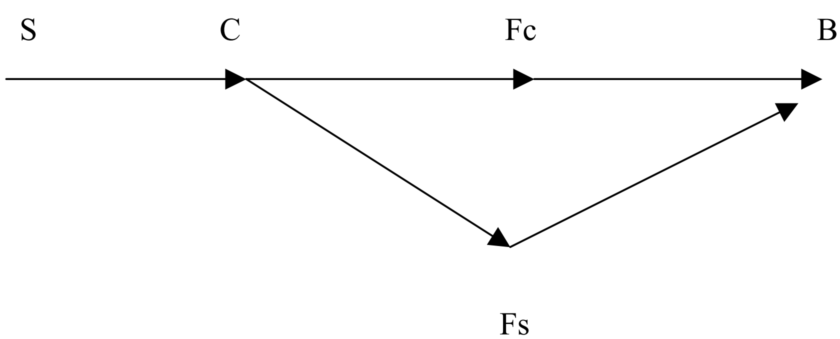

Fig. 4

Again, Fs stands for “father believes son is innocent via faith in his son”, Fc stands for “believes son is innocent via faith in courtroom demonstration” The rest of the variables stand for the same events as in the earlier presentations. The set of equations for this case would be E: $S=1$, $C=S$, $Fs=\neg C$, $Fc=C$, $B=Fc \vee Fs$. In this case there would be an active route from S to B (route ⟨S, C, Fc, B⟩. The problem with this presentation is, however, the fact that the equation $Fs=\neg C$ encodes a false counterfactual. Whether the father believes via faith in his son that his son is innocent doesn't counterfactually depend on whether the courtroom demonstration shows that he is innocent.

According to PCTK, then, what explains our different epistemic intuitions in the grandma, and father cases has to do with our different judgments regarding the causal connection between the belief and the fact believed to obtain. In the grandma case the belief that "the niece is alive" is caused by the fact that she is alive. In the father-case (Option A) the belief that "the son is innocent" is not caused by the fact that he is innocent. Grandma Betty knows that her niece is alive because she satisfies the causal condition made reference to in BC. The father (in Option A) fails to know because he violates that condition.