Reasons for Rejecting a Counterfactual Analysis of Conclusive Reasons

Abstract

In a recent article [1], Charles Cross applies Lewisian counterfactual logic to explicate and evaluate Fred Dretske's conclusive reasons account of knowledge, particularly the issues of transmission and closure. According to Cross, "Dretske's account of knowledge appeals in an essential way to counterfactual conditionals; but logics for counterfactuals, as they are currently understood, did not exist when Dretske first formulated his account". Whilst Cross makes good use of developments in counterfactual logic by applying Lewis's system VC to analyse a counterfactual interpretation of Dretske's idea of conclusive reasons, this discussion note provides some points to contend that such counterfactual logic is ultimately unsuitable to model Dretske's idea.

Keywords: conclusive reasons, Dretske, knowledge, counterfactuals, epistemic closure

In a recent article [1], Charles Cross applies Lewisian counterfactual logic to explicate and evaluate Fred Dretske's conclusive reasons account of knowledge, particularly the issues of transmission and closure. According to Cross, "Dretske's account of knowledge appeals in an essential way to counterfactual conditionals; but logics for counterfactuals, as they are currently understood, did not exist when Dretske first formulated his account". Whilst Cross makes good use of developments in counterfactual logic by applying Lewis's system VC to analyse a counterfactual interpretation of Dretske's idea of conclusive reasons, this discussion note provides some points to contend that such counterfactual logic is ultimately unsuitable to model Dretske's idea.

Cross presents Dretske's definition of the conclusive reason relation with the following:

(D1971) r is a conclusive reason for believing p iff r would not be the case if p were not the case.

Whilst Dretske's definition of conclusive reasons in his 1971 paper of the same name [4] employs terminology which suggests that it could be analysed in terms of counterfactual

logic, further investigation exposes problems with this approach. In fact, further analysis and attempts to logically explicate the Dretskean framework would do well to incorporate the details of his 1981 book *Knowledge and the Flow of Information* [5], which is a culmination of Dretske's thinking on his epistemological framework.

Writing the counterfactual conditional as $>$, (D1971) is represented as $\neg p > \neg r$. But as Cross points out, issues (semifactuals and Strong Centering) arise with this basic counterfactual formulation and "the truth of $\neg p > \neg r$ does not by itself make r a conclusive reason to believe p." Cross proposes a minor supplementation to get the following "as a friendly amendment to Dretske's theory":

(D1971+) r is a conclusive reason for believing p iff r is true but r would not be true if p were not true.

So for r to be a conclusive reason it must be actual and it follows by counterfactual *modus tollens* that p will also be the case.[1] Despite this amendment, we will now look at some further significant issues with (D1971+), some due to the intrinsic nature of counterfactuals and others due to their inadequacy in accommodating consequences of Dretske's notion of conclusive reasons.

Starting with the former, we can see that the logical nature of counterfactuals means that certain basic properties associated with conclusive reasons (or information flow) would be violated. Following are two examples.

- One basic principle is the following: if r is a conclusive reason for p, then r is a conclusive reason for $p \vee q$. Acceptance of this principle, which is endorsed by Dretske, should be straightforward enough.

  This principle would translate to the following argument in counterfactual logic using (D1971+)

  $$r, \neg p > \neg r \vdash r \wedge (\neg(p \vee q) > \neg r)$$

  which is equivalent to

[1] Whilst the requirement that the events be actual deals with said issues, it also means that the power to express a fuller range of conclusive reason relations for potential events is lost. For example, according to this definition, given that Sally is in Australia, her being in Germany cannot be said to be a conclusive reason for her being in Europe. This is not a deal breaker though and for now we can settle on the switch from (D1971+) to (D1971+).

$$r, \neg p > \neg r \vdash r \wedge ((\neg p \wedge \neg q) > \neg r)$$

However, this argument is not counterfactually valid due to the non-monotonicity of counterfactuals. So unless there is some good reason to reject this basic disjunction principle, this failure is one basic problem of the formal implementation of (D1971+).

- Demir [3] points out a highly problematic consequence of employing a counterfactual theory of information. Using his argument, it can be shown that 'r is a conclusive reason for p' entails that 'r is a conclusive reason for $p \wedge q$', for any q such that the closest not-q world is more remote than the closest not-p world.

  Suppose that Sally, a Francophile, is visiting France. Given that 'Sally is in France' (r) is a conclusive reason for 'Sally is in Europe' (p), it follows that 'Sally is in France' would be a conclusive reason for 'Sally is in Europe and Rome is the capital city of Italy' ($p \wedge q$). The reasoning is as follows:

  1. $\neg p > \neg r$ (r is a conclusive reason for p. The closest $\neg p$ world is a $\neg r$ world)
  2. $(\neg p \vee \neg q) > \neg r$ (since $\neg q$ [Rome not being Italy's capital] is much more unlikely than Sally not being in Europe, the closest $\neg p$ world is a $\neg p \vee \neg q$ world)
  3. $\neg(p \wedge q) > \neg r$ (r is a conclusive reason for $p \wedge q$)

  Whilst Sally being in France is a conclusive reason for her being in Europe, there is no reason to suppose that it is a conclusive reason for her being in Europe and Rome being the capital of Italy, as the capital of Italy has nothing to do with Sally's location.

  It also becomes apparent that this counterfactual approach does not support another basic principle termed conjunction distribution: if r is a conclusive reason for $p \wedge q$, then r is a conclusive reason for p (and r is a conclusive reason for q). It is translated to the following formal argument, which is invalid in counterfactual logic:

$$\neg(p \wedge q) > \neg r \vdash \neg p > \neg r$$

Furthermore, if conjunction distribution were imposed, a contradiction would result, as it would permit the inference of r being a conclusive reason for q, despite the fact

that Sally being in France (r) is not a conclusive reason for Rome being the capital of Italy (q) according to this account.

The behaviour of the counterfactual analysis in these examples does not accord with the idea of conclusive reasons or information flow that Dretske developed. Nor arguably with what one would generally expect from such a framework.

Aside from such peculiar results, there are other aspects of Dretske's framework that make a counterfactual logic unsuitable. Whilst much of the focus regarding Dretske's framework has been on the failure of closure under implication, it actually turns out that conclusive reasons are not closed under conjunction either. That is, the following does not always hold: if r is a conclusive reason for p and r is a conclusive reason for q, then r is a conclusive reason for $p \wedge q$.

Take Dretske's classic zebra at the zoo scenario. Let v stand for visual zebra signal, z stand for zebra and m stand for mule painted to look like a zebra. To facilitate the following general reasoning (without recourse to any particular logic), we denote A being a conclusive reason for B as $A \Rightarrow B$.

We start off with the fact that v is a conclusive reason for z:

$$\text{(CR1) } v \Rightarrow z$$

Whilst v is a conclusive reason for z, it is also a conclusive reason for $z \vee m$:

$$\text{(CR2) } v \Rightarrow z \vee m$$

Given the following necessary relations

- $\Box(z \supset \neg m)$
- $\Box(m \supset \neg z)$

it stands to reason that we can derive the following from (CR2)

$$\text{(CR3) } v \Rightarrow ((z \wedge \neg m) \vee (m \wedge \neg z))$$

Whilst $v \Rightarrow z \wedge \neg m$ would be problematic and something that Dretske does not want as it entails $v \Rightarrow \neg m$ by standard logical reasoning, (CR3) does not entail $v \Rightarrow \neg m$ and there seems to be no reason not to accept it.

Now, take (CR1) and (CR3). Given the requirements of Dretske's system, can these two conclusive reason relations be joined as per a conjunction principle to get:

- $v \Rightarrow z \wedge ((z \wedge \neg m) \vee (m \wedge \neg z))$

Well, since $z \wedge ((z \wedge \neg m) \vee (m \wedge \neg z))$ is logically equivalent to $z \wedge \neg m$, this would lead to $v \Rightarrow (z \wedge \neg m)$, which given conjunction distribution leads to $v \Rightarrow \neg m$, a result that Dretske does not want.

So in general, for any account of $\Rightarrow$ which has the following, such a consequence will follow:

1. $A \Rightarrow B, A \Rightarrow C \vdash A \Rightarrow (B \wedge C)$ (conjunction principle for conclusive reasons)
2. $B \equiv C, A \Rightarrow B \vdash A \Rightarrow C$ (substitution of logically equivalent propositions)
3. $A \Rightarrow (B \wedge C) \vdash (A \Rightarrow B) \wedge (A \Rightarrow C)$ (conjunction distribution)

Rather than abandon either of (2) or (3), it is (1), the conjunction principle for conclusive reasons, which is problematic and should be rejected. The basic diagnosis is that relations such as (C1) and (C3) are made relative to different sets of relevant alternatives in such a way that precludes their combination. In this case, $v \Rightarrow z$ requires a set of relevant alternatives which rule out m whereas $v \Rightarrow ((z \wedge \neg m) \vee (m \wedge \neg z))$ does not. See [2] for more discussion on this matter.

Thus the logic for conclusive reasons (or information flow) will be one that accommodates the failure of closure under conjunction. Counterfactual logic however validates closure under conjunction and thus is not suitable in this regard. That is, according to even minimal counterfactual logics, the following argument is valid:

$$\neg p > \neg r, \neg q > \neg r \vdash \neg(p \wedge q) > \neg r$$

This discussion note has covered some critical problems with using counterfactual logic to analyse the Dretskean notion of conclusive reasons. Rather than simply taking Dretske's original conclusive reasons definition and translating it to a counterfactual logic, it seems that a logic for Dretske's framework requires something more. Attempts to explicate and evaluate Dretske's framework would do well to incorporate his work on the concepts of information flow, relevant alternatives and channel conditions [5]. A logic for Dretske's framework would require the development of some tailored machinery to capture these notions and perhaps might need to incorporate some extra-logical properties. Suppose that $\Box(p \supset q)$ and that r is a conclusive reason for (carries the information that) p. Whether r carries the information that q depends on what exactly r and q are and whether q is a channel condition required

for r to carry the information that p. These factors also determine what are the relevant alternatives.

For example, a visual zebra signal carries the information that there is a zebra given that not-mule is a channel condition (and thus disguised mule scenarios are not amongst the relevant alternatives). As a consequence, the visual zebra signal does not carry the information that the creature is not a disguised mule. However, if the creature was determined to be a zebra by a DNA test, such information does not require not-mule as a channel condition and thus the DNA test also carries the information that the creature is not a disguised mule. Perhaps a Dretskean logic should be able to capture such distinctions.

References

[1] Charles B. Cross. A logical transmission principle for conclusive reasons. *Australasian Journal of Philosophy*, pages 1–18, forthcoming.

[2] Simon D'Alfonso. The logic of knowledge and the flow of information. *Minds and Machines*, 24(3):307–325, 2014.

[3] Hilmi Demir. The counterfactual theory of information revisited. *Australasian Journal of Philosophy*, 90(1):183–185, 2012.

[4] Fred Dretske. Conclusive reasons. *Australasian Journal of Philosophy*, 49:1–22, 1971.

[5] Fred Dretske. *Knowledge and the Flow of Information*. MIT Press, Cambridge, MA, 1981.