Coherence and truth conducive justification

CHARLES B. CROSS

Does epistemic justification turn out to be truth conducive under a coherence theory of epistemic justification? Peter Klein and Ted Warfield argue that it does not:

> ... coherence, *per se*, is not truth conducive; that is, ... by increasing the coherence of a set of beliefs, the new, more coherent set of beliefs is often less likely to be true than the original, less coherent set. (Klein and Warfield 1994: 129)

What they show is that a more coherent set of beliefs may have a smaller unconditional probability of joint truth than some of its less coherent subsets. Responding to objections raised by Trenton Merricks (1995), Klein and Warfield (1996) affirm an even stronger conclusion:

> ... if justification is both truth conducive *and* as coherentists characterize it (namely as coherence, a property of sets of beliefs) then systems of beliefs cannot possibly become more justified as they grow in logically independent beliefs. (Klein and Warfield 1996: 120)

Since Klein and Warfield (1996) insist that it is coherentists that they are addressing, and since the coherence theory of justification advanced by Laurence BonJour (1985) seems to be the theory that most influences their own discussion, one would expect their argument against the truth conduciveness of coherentist justification to apply, if at all, to BonJour's theory. In what follows I will show that this is not the case. I will also show that an important premiss is omitted from Klein's and Warfield's argument against the truth conduciveness of justification as understood in the version of coherentism they explicitly consider, which I will call the Naive Coherence Theory. According to the Naive Coherence Theory, an agent's belief set is justified iff that set is epistemically coherent, the degree of justification of the belief set being defined as the degree of its coherence. Fortunately, the premiss that Klein and Warfield omit can be defended without begging the question against the Naive Coherence Theory. The same cannot be said of an analogous premiss that would make Klein's and Warfield's argument effective against BonJour's theory.

In BonJour's version of coherentism, as BonJour takes pains to point out, the coherence of an agent's beliefs at a particular moment is not a sufficient condition for their being justified. BonJour writes:

> ... the force of a coherentist justification depends ultimately on the fact that the system of beliefs in question is not only coherent at a

ANALYSIS 59.3, July 1999, pp. 186–93. © Charles B. Cross

> moment (a result which could be achieved by arbitrary fiat), but remains coherent in the long run. It is only such long-run coherence which provides any compelling reason for thinking that the beliefs of the system are likely to be true. But the idea of long-run coherence is only genuinely applicable to a system of beliefs which is actually held by someone. (BonJour 1985: 153)

Long-run coherence is *not* a matter of how well a single set of propositions hang together: it is a matter of whether a sufficiently high degree of hanging-together is preserved across times in the belief history of an actual agent, which history consists of a temporally indexed *sequence* of sets of propositions. How does the requirement of long-run coherence fall out from BonJour's theory? First, BonJour (1985: 153) stipulates that in order to count as empirically justified an agent's contingent beliefs must belong to a system of beliefs that satisfies his Observation Requirement. In order to satisfy the Observation Requirement a system of beliefs

> ... must contain laws attributing a high degree of reliability to a reasonable variety of cognitively spontaneous beliefs (including in particular those kinds of introspective beliefs which are required for the recognition of other cognitively spontaneous beliefs). (BonJour 1985: 141)

BonJour later observes:

> ... a coherence theory which incorporates the indicated conception of observation bases justification not on the *static* coherence of a system of beliefs considered in the abstract but rather on the *dynamic* coherence of an ongoing system of beliefs which someone actually accepts. (BonJour 1985: 144)

BonJour goes on to argue for the truth conduciveness of justification under his version of the coherence theory by defending the following conclusion, in which by 'stability' he means the degree to which the belief sets in a given belief history tend to converge to a single view of the world, which view changes little except to reflect changes in the world over time:

> A system of beliefs which (a) remains coherent (and stable) over the long run and (b) continues to satisfy the Observation Requirement is likely, to a degree which is proportional to the degree of coherence (and stability) and the longness of the run, to correspond closely to independent reality. (BonJour 1985: 171)

If we are to use the resources of probability theory to describe this claim, then we must interpret it as a claim about *conditional probability*. Suppose that B is the conjunction of the members of some set of propositions.[1] The claim of the passage just quoted is not a claim about the unconditional probability of B; if it were, it would be a claim about 'a system of beliefs considered in the abstract'. BonJour's truth conduciveness thesis is instead a claim about the probability of B *given* that B is the conjunction of the current beliefs of an actual agent whose belief history consists of belief sets that have remained coherent and stable while continuing to satisfy the Observation Requirement. The truth conduciveness claim is that this *conditional* probability is positively correlated with the degree of stability of the agent's belief history, its length of run, and the degree to which the agent's beliefs have remained coherent. Let us take a moment to formulate this claim more carefully.

A set of propositions is justified, on BonJour's theory, if it is the current belief set of an agent whose belief history has exhibited stability and remained coherent while continuing to satisfy the Observation Requirement. But justification is a matter of degree, and the passage quoted above indicates that degrees of justification are to be measured along three dimensions. One of these dimensions is *level of coherence preserved across times*: other things being equal, the more coherent an agent's beliefs have remained throughout the agent's belief history, the more justified her current beliefs. A second dimension is *stability*: other things being equal, the more nearly the belief sets in an agent's belief history tend to converge upon a stable view of the world, the more justified her current beliefs. Finally we have *length of run*: other things being equal, the longer an agent's belief history, the more justified her current beliefs.[2] The three-dimensional notion of degree of justification assumed in the passage quoted above is captured if degrees of justification are represented as vectors of the form $\langle c, s, r\rangle$ and if overall comparisons of degree of justification are measured according to the following definition:

(DEF1) $\langle c_2, s_2, r_2\rangle$ represents a greater degree of justification than $\langle c_1, s_1, r_1\rangle$ iff $c_2 \geq c_1$ and $s_2 \geq s_1$ and $r_2 \geq r_1$ and either $c_2 > c_1$ or $s_2 > s_1$ or $r_2 > r_1$.

[1] We will assume for the sake of this discussion that belief sets are finite, but the probability of a set of propositions can be defined even if the set is infinite and the algebra of propositions does not allow infinite conjunctions. Elsewhere (see Cross 1995) I have endorsed a way of doing this that defines the probability of an infinite set $\{A_1, \ldots, A_n, \ldots\}$ of propositions as $\lim_{n\to\infty} \Pr(A_1 \& \ldots \& A_n)$. I no longer favour this way of handling probabilities of infinite sets but now prefer an approach endorsed by Field (1977) that takes the probabilities of infinite sets as primitive.

[2] It might be objected that a correlation between length of run and degree of justification is plausible only if agents are continuously subject to observational input, which they are not. BonJour could answer this objection by adjusting how length of run is measured: Periods of time when an agent is temporarily shut off from observational input (e.g. periods of unconsciousness) but during which her beliefs do not change should simply be deducted from a belief history's length of run.

Since degree of justification is measured along three dimensions, the justification condition in terms of which truth conduciveness is evaluated must also reflect this:

(DEF2) B is justified to a degree defined by $\langle c, s, r\rangle$ [symbolically $J(B, c, s, r)$] iff B is the conjunction of the members of the current belief set of an actual agent whose belief history has length r and consists of belief sets that have remained coherent to degree c and stable to degree s while satisfying the Observation Requirement.

Then, where Pr is a measure of objective probability, BonJour's truth conduciveness claim can be formulated as follows, assuming that $0 < \Pr(B_1) < 1$ and $0 < \Pr(B_2) < 1$ and $\Pr(J(B_1, c_1, s_1, r_1)) \neq 0 \neq \Pr(J(B_2, c_2, s_2, r_2))$:

(TC1) If $\langle c_2, s_2, r_2\rangle$ represents a greater degree of justification than $\langle c_1, s_1, r_1\rangle$, then $\Pr(B_2 \mid J(B_2, c_2, s_2, r_2)) > \Pr(B_1 \mid J(B_1, c_1, s_1, r_1))$.

That is, for BonJour's theory (as for any theory of justification) truth conduciveness means that the probability of B_2 given that B_2 is justified to a greater degree exceeds the probability of B_1 given that B_1 is justified to a lesser degree.

BonJour defends the truth conduciveness claim quoted earlier and formalized above as (TC1) on the basis of an argument that correspondence to reality would be the best explanation of why an ongoing belief system that continued to satisfy the Observation Requirement would remain coherent and stable over the long run (see Bonjour 1985: 171–79). It is not my aim here to assess BonJour's argument, but regardless of the merits of that argument, it is clear, as I will presently show, that Klein's and Warfield's critique of coherentism does not undermine the truth conduciveness claim that BonJour defends.

Consider Klein's and Warfield's Dunnit case, in which an agent's initial belief set (the conjunction of whose members is B) is expanded, with no initial belief being dropped, to become a new and more coherent belief set (the conjunction of whose members is B^*) by the mere addition of a new belief b that improves the degree to which the agent's beliefs 'hang together'.[3] Klein and Warfield correctly point out that if b is logically independent of B and $B^* = B$ & b and $\Pr(B) > 0$ and $\Pr(b) < 1$ (as in the Dunnit

[3] Merricks (1995: 308) provides the following succinct formulation of the example:

> Now consider two belief sets. The first, B, includes the following beliefs:
>
> b_1 Dunnit had a motive for the murder.
>
> b_2 Witnesses claim to have seen Dunnit do it.
>
> b_3 A credible witness claims to have seen Dunnit two hundred miles from the scene of the crime at the time of the murder.
>
> b_4 Dunnit committed the murder. *(continued)*

case), then, even though the belief set represented by B^* may be more coherent than the belief set represented by B, still it must turn out that $\Pr(B^*) < \Pr(B)$. But what matters about this example if we are evaluating the truth conduciveness of justification on BonJour's theory is not the relation between the unconditional probability of B^* and the unconditional probability of B. Instead, what matters is the relation between the probability of B^* *given* that B^* is justified to a greater degree (represented, let us say, by the triple $\langle c^*, s^*, r^* \rangle$) and the probability of B *given* that B is justified to some lesser degree (represented, let us say, by the triple $\langle c, s, r \rangle$). The Dunnit example fails as a counterexample to (TC1) because from the fact that $\Pr(B^*) < \Pr(B)$ we cannot infer that $\Pr(B^* \mid J(B^*, c^*, s^*, r^*)) \leq \Pr(B \mid J(B, c, s, r))$. That is, $\Pr(B^* \mid J(B^*, c^*, s^*, r^*))$ may well be *greater* than $\Pr(B \mid J(B, c, s, r))$, just as (TC1) requires, despite the fact that the belief set represented by B^* is a superset of the belief set represented by B. This possibility exists because the distinctness of B^* and B (as well as the distinctness of $\langle c^*, s^*, r^* \rangle$ and $\langle c, s, r \rangle$) ensures that $J(B^*, c^*, s^*, r^*)$ and $J(B, c, s, r)$ are distinct propositions, so that the conditional probability of B^* and the conditional probability of B are computed using different conditions.

The argument of the preceding paragraph turns on its being possible for $\Pr(B^* \mid J(B^*, c^*, s^*, r^*))$ to be greater than $\Pr(B \mid J(B, c, s, r))$ even when $\Pr(B^*) < \Pr(B)$, but we would be able to rule out this possibility if we were entitled to assume not only that $\Pr(B^*) < \Pr(B)$ but also that B^* is probabilistically independent of $J(B^*, c^*, s^*, r^*)$ and that B is probabilistically independent of $J(B, c, s, r)$, for in that case it would follow that $\Pr(B^* \mid J(B^*, c^*, s^*, r^*)) = \Pr(B^*) < \Pr(B) = \Pr(B \mid J(B, c, s, r))$, and the Dunnit example would be a counterexample to (TC1) after all. But to make these independence assumptions would beg the question against BonJour: part of what his truth conduciveness claim means is precisely that the joint truth of the propositions in an agent's belief set is *not* probabilistically independent of the justification status of the agent's belief set, assuming that the justification status of an agent's belief set is determined by how the agent's belief history stands with respect to certain requirements of coherence, stability, and observational input. Thus, Klein's and Warfield's analysis notwithstanding, the Dunnit example presents no obstacle to its turning out that justification as defined on BonJour's coherence theory is truth conducive.

Now consider B^* which has the members of B, plus the following:

> *b* Dunnit has an identical twin which was seen by the credible witness two hundred miles from the scene of the crime during the murder.

B^* is more coherent than B.

But if Klein's and Warfield's Dunnit example does not show that BonJour's conception of justification is not truth conducive, does it at least show that justification is not truth conducive on the Naive Coherence Theory, the explicit target of their argument? The example does show this, but our discussion has exposed a gap in Klein's and Warfield's argument on this point. Fortunately, the gap can be filled.

Recall that on the Naive Coherence Theory, what makes a belief set justified is its being epistemically coherent, the degree of justification of the belief set being defined as its degree of epistemic coherence. Accordingly, the appropriate justification condition to use in evaluating the truth conduciveness of justification for the Naive Coherence Theory is the following:

(DEF3) B is justified to degree c [symbolically $j(B,c)$] iff B is the conjunction of the members of a set of propositions whose degree of coherence is c.

Since truth conduciveness is *for any theory of justification* a matter of how likely a bearer of justification is to be true *given* how justified it is, the truth conduciveness claim for the Naive Coherence Theory can be formulated as follows:

(TC2) If $c_2 > c_1$, then $\Pr(B_2 \mid j(B_2, c_2)) > \Pr(B_1 \mid j(B_1, c_1))$.

That is, the probability of B_2 given that B_2 is justified to a greater degree exceeds the probability of B_1 given that B_1 is justified to a lesser degree. Now, recall that in the Dunnit example, B^* is more coherent than B, yet $\Pr(B^*) < \Pr(B)$. Letting c^* represent the degree of coherence of B^* and letting c represent the degree of coherence of B, does it follow that the Dunnit example is a counterexample to (TC2)? This depends on whether it follows from $\Pr(B^*) < \Pr(B)$ that $\Pr(B^* \mid j(B^*, c^*)) \leq \Pr(B \mid j(B, c))$, but the latter does *not* follow given only that $\Pr(B^*) < \Pr(B)$. Once again we are computing conditional probabilities using different conditions. The difference is that in the case of the Naive Coherence Theory, a move from $\Pr(B^*) < \Pr(B)$ to $\Pr(B^* \mid j(B^*, c^*)) \leq \Pr(B \mid j(B, c))$ can be justified in a non-question-begging way given an additional premiss. Let us consider how this can be done.

We noted in our discussion of (TC1) that the Dunnit example is a counterexample to (TC1) if certain probabilistic independence assumptions are true. The Dunnit example is a counterexample to (TC2) if analogous probabilistic independence assumptions hold concerning truth and justification as analysed by the Naive Coherence Theory:

(PrInd) In the Dunnit example, B^* and $j(B^*, c^*)$ are probabilistically independent, and B and $j(B, c)$ are probabilistically independent.

The Dunnit example does refute (TC2) if (PrInd) holds, for in that case (assuming that $\Pr(j(B,c)) > 0$ and $\Pr(j(B^*,c^*)) > 0$) $\Pr(B^* \mid j(B^*,c^*)) = \Pr(B^*) < \Pr(B) = \Pr(B \mid j(B,c))$ even though $c^* > c$. But, in giving a critique of the Naive Coherence Theory, is it *fair* to assume (PrInd)?

In our discussion of (TC1) and (TC2) we noted that it would beg the question against BonJour to assume that the joint truth of a given set of propositions is probabilistically independent of its being the current belief set of an actual agent whose belief history meets certain requirements of coherence, stability, and observational input. By contrast, it does not beg the question against the Naive Coherence Theory to assume (PrInd), or indeed to assume that in general, the joint truth of a set of propositions is probabilistically independent of its degree of coherence, taken in the abstract. Indeed, the following (non-question-begging) argument is sufficient to establish (PrInd): if c^* and c are the degrees of coherence of B^* and B, respectively, considered in the abstract, then it is not a contingent matter that B^* and B have these respective degrees of coherence. Thus, if true at all, $j(B^*,c^*)$ and $j(B,c)$ are necessarily true, and if $j(B^*,c^*)$ and $j(B,c)$ are necessarily true, then each has unit unconditional objective probability, i.e. $\Pr(j(B^*,c^*)) = \Pr(j(B,c)) = 1$, in which case $\Pr(B^* \mid j(B^*,c^*)) = \Pr(B^*)$ and $\Pr(B) = \Pr(B \mid j(B,c))$.[4] This argument does not beg the question against the Naive Coherence Theory because it does not beg the question against that theory to suppose that degree of coherence is a non-contingent property of sets of propositions considered in the abstract.

A final question arises: if when the Naive Coherence Theory is applied to the Dunnit example it can be argued that $\Pr(j(B^*,c^*)) = \Pr(j(B,c)) = 1$, then perhaps it can be argued in the context of BonJour's theory that $\Pr(J(B^*,c^*,s^*,r^*)) = \Pr(J(B,c,s,r)) = 1$, so that justification and truth are probabilistically independent in that context after all. Nothing doing. A claim of the form $J(B,c,s,r)$ reports the existence of an actual agent whose current belief set is represented by B and whose actual belief history satisfies requirements of coherence, stability, and observational input.[5] Since

[4] I include below the independence calculation for the belief set represented by B. The other calculation is similar. Assume $\Pr(j(B,c)) = 1$; then we have $\Pr(B \,\&\, j(B,c)) = \Pr(B)$ and hence the following:

$$\Pr(B \mid j(B,c)) = \frac{\Pr(B \,\&\, j(B,c))}{\Pr(j(B,c))} = \frac{\Pr(B)}{1} = \Pr(B).$$

[5] By contrast, a claim of the form $j(B,c)$ does *not* assert the existence of an actual agent whose belief set is represented by B. That is, *being the belief set of an actual agent* is not part of what makes a belief set justified on the Naive Coherence Theory. It is true that when considering claims of the form $j(B,c)$ we normally restrict the range of B to conjunctions of belief sets of agents, but this does not mean that $j(B,c)$ itself asserts that B represents the belief set of an actual agent.

the existence of an agent with such a belief history is a contingent matter, no claim of the form $J(B, c, s, r)$ can plausibly be said to have unit unconditional objective probability.

Klein and Warfield were therefore right about the Naive Coherence Theory, though, as we have seen, their argument was incomplete: it omitted an important premiss, namely (PrInd). The situation is different as regards BonJour's theory: the truth conduciveness of justification on BonJour's theory is not refuted by the Dunnit example or, in general, by the fact that a coherent set of beliefs will often turn out to be more likely to contain a falsehood than some of its less coherent subsets. Since this latter fact does not constitute a reason to reject BonJour's theory, it does not constitute a reason to reject the very idea of a coherence theory of justification.[6]

The University of Georgia
Athens, GA 30602-1627, USA
ccross@ai.uga.edu

References

BonJour, L. 1985. *The Structure of Empirical Knowledge*. Cambridge, Mass.: Harvard University Press.

Cross, C. 1995. Probability, evidence, and the coherence of the whole truth. *Synthese* 103: 153–70.

Field, H. 1977. Logic, meaning, and conceptual role. *Journal of Philosophy* 74: 379–409.

Klein, P. and T. Warfield. 1994. What price coherence? *Analysis* 54: 129–32.

Klein, P. and T. Warfield. 1996. No help for the coherentist. *Analysis* 56: 118–21.

Merricks, T. 1995. On behalf of the coherentist. *Analysis* 55: 306–309.

[6] This research was supported in part by the University of Georgia Research Foundation. The author would like to thank Peter Smith, Stephen Hetherington, and an anonymous referee for comments leading to improvements in this essay.