On the Autonomy of (Some) Knowledge

Kurt L. Sylvan

Introduction

J. Adam Carter's great new book *Autonomous Knowledge* argues that a kind of epistemic autonomy is necessary for knowledge. More specifically, Carter defends 'the need for…an *autonomous belief condition* on propositional knowledge—a condition the satisfaction of which…is *neither* entailed by nor entails the satisfaction of either a belief condition or, importantly, an epistemic justification condition' (viii). A little later in the book, Carter puts the claim more strongly, suggesting that his autonomy condition neither entails nor is entailed by 'any epistemic condition on knowledge' (22). This stronger claim, which I will call the *Main Claim*, is really the bold new idea in the book. Carter describes the idea as 'revisionary', and suggests we need an 'update' of epistemology in light of 'the fact that the nature of cognition is rapidly changing via the latest science' (6).

Carter defends the Main Claim in chapter 1 via a series of variations on the TrueTemp thought experiment (see §1 for some cases). He then offers an externalist account of the autonomy condition in chapter 2, after maintaining that 'internalism about epistemic autonomy is a non-starter' (37). After these key chapters, Carter explains in chapter 3 how the autonomy condition predicts the existence of some unrecognized forms of epistemic defeat, considers how his view might extend to knowledge-how in chapter 4, and seeks to understand the value of autonomous belief in chapter 5, where he takes a surprising turn toward a Kantian view, inspired by Korsgaard's work on self-constitution.

In this critical notice, I will focus on arguments and ideas in chapters 1, 2 and 5. I will make three sets of criticisms, and end with some questions about the place of autonomy in epistemology. Firstly and most importantly, I will cast doubt on Carter's defense of the Main Claim in chapter 1. I will argue that the autonomy condition follows from familiar conditions on knowledge in virtue epistemology and is better explained by views that are more consistently Kantian than Carter's. More generally, I will suggest that the condition follows from a proper basing requirement on reflective knowledge. I will also raise some doubts about the way Carter sets up the argument for his Main Claim. Even if he were right that the autonomy condition is independent of a *justification* condition, I think he doesn't do enough to show that it is independent of a 'de-Gettierization' condition.

I'll then explore whether Carter is right to claim in chapter 2 that an internalist account of epistemic autonomy is a 'non-starter'. I'll suggest that Carter doesn't consider all the options available to internalists, especially when we consider more Kantian forms of internalism. I will also suggest that if Kantian internalism really is incompatible with Carter's view (which is not as obvious as it may seem!), it is preferable to his view. Finally, I'll respond to Carter's attempt to show that Sylvan (2018)'s account does not explain the epistemic value of autonomy. I will suggest that Carter's Kantian account of the value of autonomy is compatible with (and may need scaffolding from) this kind of account.

Despite these reservations, I think the book is a major advance in epistemology. It shows convincingly that autonomy can play an important role in traditional projects of analysis in epistemology, paving the way for new Kantian approaches. The task now is to understand exactly where and why autonomy plays this role. Is its role limited to the theory of justification and reflective knowledge? If so, what then is the place of unreflective, 'animal', knowledge? If, as I think is true, reflective subjects can have mere animal knowledge and our animal side sometimes *properly* bypasses

our reflective side, can autonomy have as deep a place as Carter hopes? I look forward to seeing more work on these questions inspired by Carter's book.

1. Against the Main Claim: Autonomy Follows from Other Conditions on Knowledge

Carter's argument for the Main Claim proceeds by a series of iterations on Lehrer (1990: 162-3)'s celebrated TrueTemp thought experiment. Here is the original case, as worded by Carter (p.8):

> TRUETEMP: 'Suppose a person, whom we will call Mr. Truetemp, undergoes brain surgery by an experimental surgeon who invents a small device which is both a very accurate thermometer and a computational device capable of generating thoughts. The device, call it a tempucomp, is implanted in Truetemp's head so that the very tip of the device, no larger than the head of a pin, sits unnoticed on his scalp and acts as a sensor to transmit information about the temperature to the computational system in his brain. This device, in turn, sends a message to his brain causing him to think of the temperature recorded by the external sensor. Assume that the tempucomp is very reliable, and so his thoughts are correct temperature thoughts. All told, this is a reliable belief-forming process. Now imagine, finally, that he has no idea about why he thinks so obsessively about the temperature, but never checks a thermometer to determine whether these thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp.'

It is, as Carter notes, common to deny that TrueTemp has justified beliefs about the temperature, and hence to explain TrueTemp's lack of knowledge about the temperature via a justification condition. But Carter observes that the case can be modified so that TrueTemp does satisfy a justification condition. He begins with the following revised case (p.11):

> TRUETEMP*: 'TrueTemp* has a highly sophisticated device implanted in his head that regularly causes him to form true beliefs about the ambient temperature. And whenever the device causes him to form a temperature belief, p, it also compels him to believe another proposition, q, where q is a good reason for p, and, finally, the device then compels TrueTemp* to believe p on the basis of a good reason, p.'

Carter argues that the justification condition on knowledge can be satisfied *heteronomously* in such iterations of TRUETEMP. Now, Carter concedes that TRUETEMP* does not establish this point alone, since only a very simple reasons-based justification condition is satisfied in this case. But Carter points out that further variations can be constructed in which more sophisticated justification conditions are satisfied. Here, for example, is a case in which a variety of much more demanding conditions (including some virtue-theoretic and reliabilist ones) would be satisfied, but where the heteronomy intuition (allegedly) remains (p.16):

> TRUETEMP***: 'TrueTemp*** has a highly sophisticated device implanted in his head. Once implanted, the device—through an immediate and highly advanced form of neuromodulation—remaps TrueTemp***'s cognitive architecture in such a way that, while the device is still causing him to believe what he does about the ambient temperature, the process that leads to this…is *itself* auto-integrated with the rest of his…cognitive architecture. A consequence of the auto-integration is that the process that controls his temperature belief formation…is not insensitive to other dispositions governing the formation and evaluation of his beliefs, but *this* is only because the device is *also* controlling these other dispositions that govern the formation and evaluation of his beliefs.'

Although Carter focuses primarily on showing that any reasonable justification condition can be heteronomously satisfied in such variations, he thinks his point 'can be generalized to apply not just to traditional epistemic justification conditions on knowledge but to *any epistemic* condition on knowledge, including modal epistemic conditions, anti-Gettier conditions, and the like' (10). Hence

he is committed to the stronger claim that the autonomy condition is independent of 'any epistemic condition on knowledge' (22).

Carter doesn't do enough to support this stronger claim. There are two options he doesn't sufficiently consider. One is to argue that the autonomy condition follows from the best solution to the Gettier problem. Here we can consider Sosa (2007, 2011, 2015)'s view that knowledge is apt belief, where apt belief is accurate belief whose very accuracy *manifests* cognitive competence. As far as I can see, in the TrueTemp iterations Carter considers where TrueTemp fails to know, Sosa can claim that the accuracy of TrueTemp's belief does not manifest competence. He can make this claim for the same reason why Carter denies that TrueTemp's belief is autonomous in chapter 2: namely, because it 'bypasses' S's cognitive competences.

Given Carter's reliance on concepts from Sosa's epistemology in chapter 2, it is surprising that he insists that the autonomy condition is separately needed, since it would be simpler to derive this condition from the requirement that the belief's accuracy *manifest* competence. To see this, consider Carter's eventual account of autonomous belief:

> 'S's belief that p is epistemically autonomous (i.e., autonomous in the way that is necessary for propositional knowledge) at a time, t, if and only if [this belief] has a compulsion-free history at t; and this is a history it has if and only if it's not the case that S came to acquire her belief that p in a way that: (i) bypasses or preempts S's cognitive competences, and (ii) the by-passing or pre-emption of such competence issues in S's being unable *to easily enough* shed p' (53).

Both negative conditions follow, Sosa could argue, from the positive condition that the accuracy of S's belief must manifest competence. To see this for both conditions, we should separately consider whether the *formation* of a belief manifests competence and whether its *retention* manifests competence. If S acquired her belief in a way that bypasses or preempts her cognitive competences, then the accuracy of her belief cannot manifest those competences at the time of formation. But, as Carter I think rightly suggests, it should be possible for a belief that was acquired in a non-autonomous way to become autonomous if its retention after critical scrutiny manifests competence. This is in effect what he suggests in TRUETEMP-SHEDDABLE (47-48), where TrueTemp 'elects not to revise [his] belief in any way, despite having the power to, after subjecting it to (non-compelled) rational scrutiny, including scrutiny by which he comes to find out that the mechanism he's using is a reliable one' (48). To address these subtleties, a defender of Sosa's view can say that for a belief to constitute knowledge, it must be formed or maintained in a way that manifests competence. Satisfaction of this condition entails satisfaction of Carter's autonomy condition. Hence Sosa doesn't need a separate autonomy condition.

We can make similar points about structurally similar accounts of knowledge, like Wedgwood (2020)'s and Sylvan (2020: 20-21)'s. Wedgwood (2020: 5363) holds that 'if you have an outright belief in a proposition *p*, this counts as a case of your *knowing* p if and only if it is a case of your believing *correctly* precisely *because* it is a case of your believing *rationally* (that is, a case of your manifesting rational virtues to a sufficient degree).' It is open to Wedgwood to more explicitly claim that your correctly believing p must either in formation or retention manifest your rational capacities. This yields Carter's epistemic autonomy condition. Similarly, Sylvan (2020: 20-21) suggests that reflective knowledge consists in *complying* with objective accuracy-relevant reasons for belief, where one complies iff one believes accurately in virtue of *strongly respecting* accuracy, where this, in turn, involves believing accurately as a manifestation of a disposition to respond to objective accuracy-relevant reasons. One can elaborate on this by saying that it must be the case that your correctly believing p either in formation or retention manifests strong respect for accuracy. This more explicit statement yields Carter's epistemic autonomy condition.

These points show that Carter should have done more to consider whether existing epistemic conditions beyond the justification condition can address his cases. But related points go for doxastic

justification, since there are views that understand it by appealing to ideology from virtue epistemology, like Lord and Sylvan (2020)'s, Mantel (2018)'s, and Sylvan and Sosa (2018)'s. These accounts imply that doxastic justification requires proper basing on reasons, where such basing is understood as holding the attitude supported by these reasons *as a manifestation of reasons-sensitive competence*. In Carter's cases, it won't be true that one holds the reasons-favored attitude as a manifestation of reasons-sensitive competence. Hence it will not be the case that one has a doxastically justified belief on these views.

Carter cannot respond here, as he does elsewhere in chapter 1, by insisting that he can *stipulate* that the proper basing requirement is satisfied heteronomously in some extended TrueTemp case. This would be like stipulating that one can satisfy the conditions in Carter's account of autonomy heteronomously. That is impossible, since it would be incoherent to imagine both that a belief was formed or maintained in a way that doesn't bypass or preempt the believer's competences, and that this formation or retention was brought about by compulsion. But defenders of the foregoing views can similarly claim that it is incoherent to imagine both that an accurate belief is formed or maintained as a manifestation of reasons-sensitive competence, and that this formation or retention owes to compulsion.

So, there is, it seems clear, a problem with Carter's argument: whatever prevents TrueTemp cases from undermining his view will also prevent it from applying to these views of doxastic justification.

2. Internalism and Externalism about Epistemic Autonomy

Let's consider chapter 2. One thing that happens here is that Carter quickly dismisses internalism about epistemic autonomy. He gives two kinds of arguments. One is that internalism gives the wrong predictions in allegedly possible cases such as the following:

> 'PSYCHOLOGICAL TWINS*: Ann and Beth are psychological twins. They are identically mentally constituted. Both believe that Cicero's scribe was named Tiro. Ann believes this because she read it in a book. Beth believes it because scientists want her to be psychologically identical to Ann, and so they brainwash her until her psychology—as it pertains to all matters of Roman history—matches Ann's exactly' (35).

But I think this case is either not possible as described, or, if it is possible, that there is no reason to deny that Beth has knowledge.

One kind of internalist would claim that when Ann believes this proposition with doxastic justification at t, this is due to her belief's manifesting her rational capacities at t. They would also claim that these features supervene on intrinsic features of her mind. If it is true that Beth at t* is an intrinsic mental duplicate of Ann at t, then it must also be true that she believes this proposition in a way that manifests her rational capacities. If so, then although it may be true that Beth only ended up doing this 'because of' her interaction with the scientists, it is not true that her belief is *symptomatic* of brainwashing.

We can instead treat this case like some irrelevant influence cases are treated by theorists who claim that rational belief and knowledge is still possible in them. The sense in which Beth believes 'just because' of her interactions with the psychologists is like the sense in which, according to Stanley (2015: Ch.5), Stanley knows the value of democracy only because of his ideology and upbringing. Hence, if the case is supposed to be one where the belief manifests brainwashing, we can deny that it is possible (assuming it is a conceptual truth that brainwashed beliefs cannot be knowledge). A happier description of the genuinely possible case lurking here is that Beth comes to have a rational belief and, indeed, knowledge despite the fact that her belief is held in some sense 'only because' of an irrelevant influence.

If one likes the autonomy condition, one should conclude that not all irrelevant influences undermine its exercise. If one doesn't like that conclusion, one shouldn't be so sure that rational belief requires relevant autonomy. For it is possible for brainwashed subjects to have some knowledge that they wouldn't have had were it not for the influence of their brainwashers. For example, if a dutiful subject in North Korea seeks to learn about events through state TV, they can gain *some* knowledge about restricted subject matters where it speaks the truth (e.g., about local weather), even if their general trust owes to patriotism.

So much for Carter's first argument against internalism. His second is a spinoff of a standard objection to hierarchical accounts of autonomy. He writes that '[t]he most promising internalist approaches…tell us that the autonomy of a given belief is ultimately a matter of the relationship that the belief bears to one's 'higher-order' attitudes, such as those that feature in the process of reflective identification' (35-36). He then suggests that if first-order beliefs can be heteronomous, so can the relevant higher-order attitudes. While I agree that this is a good objection to hierarchical approaches, internalist approaches needn't be hierarchical. A more promising internalist account would be a Kantian one, which sees epistemic autonomy across time as a matter of complying with norms of theoretical reason in one's processing of the evidence at every time, where such compliance consists in manifesting one's rational capacities at every time, and where one's rational capacities are understood (plausibly, I think) as intrinsic properties of one's mind. Reflective knowledge is then a matter of accurate belief that manifests rational processing, which is necessarily epistemically autonomous.

This kind of account fits better with the Kantian story about the value of epistemic autonomy Carter tells in the last chapter. Moreover, it is unified and helps to explain why the conditions in Carter's view matter. Carter's view grounds epistemic autonomy in something negative: believing in a way that does *not bypass or preempt* S's cognitive competences. But something positive explains why this condition is met when it is: rationality.

We shouldn't be confused here by the fact that the negative condition is in some broad sense 'historical'. It is true that whether a belief manifests epistemic autonomy depends on *why* one holds it. But when I believe autonomously, this 'why' picks out *my reason*, which sustains my belief through my cognitive spontaneity at the very time it is held. The 'history' of my belief doesn't refer to the past of my belief considered as a mere occurrence in the world, but to the *rational basis* of my belief. In the first instance, we would speak here of 'history' only in the way that we speak of 'priority' in metaphysics.

3. The Value of Epistemic Autonomy

Let's consider Carter's discussion of the epistemic value of autonomous belief. Carter rejects pragmatic and instrumental explanations of this value and confines his attention to non-instrumental explanations. Drawing on a conceptual category introduced by Korsgaard (1983), he suggests that certain extrinsic properties of autonomous beliefs explain their non-instrumental epistemic value. In particular, the fact that these beliefs contribute to *epistemic self-constitution* is what gives them their non-instrumental epistemic value, according to Carter. Carter takes this value not to be derivative from familiar epistemic values, such as accuracy or knowledge. Hence, on his view, autonomous beliefs have extrinsic but non-instrumental epistemic value that is not derived from a relation to other epistemic values (besides epistemic self-constitution). It is noteworthy that his explanation does not look to the history of a belief, but rather forward, to its role in self-constitution, to explain its value. Epistemically autonomous beliefs are part of the larger activity of epistemic self-constitution, whose value 'flows back' to the beliefs.

This view is attractive, but I think that the view in Sylvan (2018) that Carter criticizes can incorporate much of what is good about it. To bring this out, let's briefly review this view. On this view, truth is the fundamental epistemic value, and all derivative epistemic value is explained via some connection to *respect* for truth, which has a distinctive kind of derivative value suggested by Hurka (2001). The

value of rational belief is explained in two stages: first, rational belief has derivative but non-instrumental value because it constitutes respect for the truth, but respect for truth itself gets *its* value by being an epistemically fitting response to the fundamental epistemic value of truth. The first part of this story is structurally like Carter's: one non-instrumentally epistemically valuable item gets its non-instrumental epistemic value by being constitutively related, in a forward-looking way, to a higher value. The story adds, however, that the epistemic value here comes from something—namely, respect for truth—that has derivative value relative to the more fundamental epistemic value of truth. This is like how a morally worthy act gets derivative value from a more fundamental moral value (e.g., personhood) by contributing to respect for that more fundamental value (e.g., respect for personhood).

With this story in mind, one would expect that the value of autonomous belief could be explained in a similar way to rational belief: epistemically autonomous beliefs contribute to (or manifest) proper valuing of truth and that they have derivative epistemic value for that reason. That seems promising, but more needs to be said about *why* epistemic autonomy manifests or contributes to proper valuing of truth. I think the full answer is complicated, but I will give a simpler answer here, which rests on the fact that I don't think it is possible to *genuinely value* anything heteronomously. One can *desire* something heteronomously, but one's *values* are themselves part of one's self-constitution. So, one cannot properly epistemically value truth in one's thinking unless this thinking is epistemically autonomous.

This isn't the story Carter considers. But I don't think either story is undermined by his discussion. His key objection assumes that it is possible to manifest respect for truth heteronomously:

> [I]t's one thing for a belief to be epistemically autonomous; it's another for a belief to manifest respect for truth autonomously. For the line [Sylvan defends] to work, it would need to be the case that the former kind of autonomy implies the latter. But it does not. In short, an epistemically autonomous belief could manifest respect for truth but do so heteronomously. (133)

These claims are mistaken. Heteronomous respect for truth is not possible: it isn't genuine respect if it is heteronomous. It is also not necessary to claim that the epistemic autonomy of proper belief 'implies' the epistemic autonomy of respect for truth if this is meant to track an explanatory relation. While epistemically autonomous beliefs will also be autonomously truth-respecting, the beliefs are epistemically autonomous because the respect is epistemically autonomous, rather than vice versa. Finally, I see no reason to agree, in the relevant senses, that it is 'one thing' to be epistemically autonomous and 'another thing' to manifest respect for truth autonomously: the former plausibly partly constitutes the latter.

If all these claims of Carter's fail, the alternative story I've suggested explains the value of epistemic autonomy at least as well as his. I also think this story does work that his does not do—namely, to explain why self-legislation in the doxastic realm has *epistemic* value, rather than value *simpliciter*. Ultimately, our accounts are not inconsistent: my account can be used to help underpin his account, and show how the kind of extrinsic non-instrumental value he attributes to autonomous belief can be explained using tools already in the literature.

4. The Place of Epistemic Autonomy

A final concern I want to raise concerns the place of autonomy in epistemology. Here some awkward questions seem to arise for Carter. While it is plausible that some knowledge is constituted by autonomous belief, there is a kind of knowledge—namely, animal knowledge—for which it is hard to believe that it is a requirement. Suppose, then, that reflective knowledge but not animal knowledge requires autonomy. What is it about reflective knowledge that makes it have this additional requirement? A plausible answer suggested by Sosa (2015) is that reflective knowledge requires justification for belief when belief is understood as a disposition to judge, and that autonomy is

necessary for justified judgment. This cannot be Carter's answer, since he has argued that autonomy is independent of justification. But what else about reflective knowledge makes the difference?

Here is another angle on this problem. Carter's account of epistemic autonomy collapses, I argued earlier, into a virtue-theoretic view that analyzes epistemic autonomy in terms of epistemic competence. But the notion of competence that Carter takes from Sosa applies to animals. There are animal beliefs that manifest competence and animal beliefs that don't: there are paranoid cats, and cats on too much catnip, for example. So, it must be something about the nature of the epistemic competences that humans have and animals lack that explains the difference. One might conjecture that the former competences are seated in reason, while the latter are not. This is the answer suggested by the Kantian internalist. Yet that answer is in tension with Carter's in chapter 2. What, then, is it about human epistemic competence that makes it suited to explain epistemic autonomy?

Conclusion

Let me conclude by emphasizing that I think this book should be widely read, despite its shortcomings. Although the book is packaged to appear applied and high-tech, at bottom it makes a fundamental contribution to contemporary epistemology that reanimates old, forgotten questions rooted in Kant. While Carter may be wrong that autonomy is a *novel* requirement on any central epistemic standing, he convincingly shows, I think, that it is a requirement on a central epistemic standing (viz., reflective knowledge). The significance of this point is far-reaching and invites a new confrontation between reliabilist virtue epistemology and Kantian epistemology. For even if it is true, as I've argued above, that reliabilist virtue epistemologists can claim already to make room for the autonomy condition, there is a question to consider about what best explains this condition. Given that animal knowledge is plausibly *not* subject to this condition, it is unclear how *reliabilist* virtue theory can make sense of, rather than simply take for granted, the interesting fact that manifesting reflective virtues entails autonomy. To explain this fact, I think we need a more consistently Kantian theory than Carter offers in the book. His gestures toward Korsgaard at the end hold promise, but they come too late in the book. *Sub specie aeternitatis*, however, it is never too late for Kant.

The University of Southampton
UK
k.l.sylvan@soton.ac.uk

References

Carter, J. A. 2022. *Autonomous Knowledge*. OUP.
Hurka, T. 2001. *Virtue, Vice, and Value*. OUP.
Korsgaard, C. 1983. 'Two Distinctions in Goodness.' *Philosophical Review* 92: 169-195.
Lehrer, K. 1990. *Theory of Knowledge*. Westview Press.
Lord, E. and Sylvan, K. 2020. 'Prime Time (for the Basing Relation)' in Carter, J. A. and Bondy, P. (eds.) *Well-Founded Belief*. Routledge.
Mantel, S. 2018. *Determined by Reasons*. Routledge.
Sosa, E. 2007. *Apt Belief and Reflective Knowledge, Volume 1*. OUP.
Sosa, E. 2011. *Knowing Full Well*. Princeton University Press.
Sosa, E. 2015. *Judgment and Agency*. OUP.
Stanley, J. 2015. *How Propaganda Works*. Princeton University Press.
Sylvan, K. and Sosa, E. 2018. 'The Place of Reasons in Epistemology' in Star, D. (ed.) *The Oxford Handbook of Reasons and Normativity*. OUP.
Sylvan, K. 2018. 'Veritism Unswamped.' *Mind* 127: 381-435.
Sylvan, K. 2020. 'An Epistemic Nonconsequentialism.' *Philosophical Review* 129: 1-51.
Wedgwood, R. 2020. 'The Internalist Virtue Theory of Knowledge.' *Synthese* 197: 5357-5378.