

Autonomous knowledge: radical enhancement, autonomy, and the future of knowing

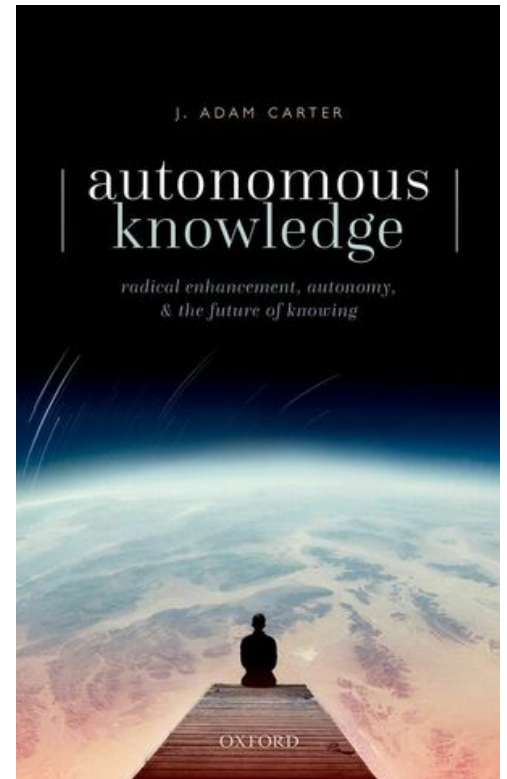
J. Adam Carter, *Autonomous knowledge: radical enhancement, autonomy, and the future of knowing*, Oxford University Press, 2022, 176pp., \$70.00 (hbk), ISBN 9780192846921.

Reviewed by Chris Tweedt, Christopher Newport University

2022.11.11

J. Adam Carter argues knowledge is more than justified, true, non-Gettiered belief; knowledge also requires an additional component that is not reducible to the others: autonomy. Knowledge must be the subject's knowledge; it must have come about by the subject, not compelled by an external device. Carter argues for the need for an independent autonomy component to both propositional knowledge and know-how (Chs. 1, 4), articulates what the autonomy condition involves (Chs. 2, 4), provides an account of knowledge defeat that is tied to autonomy (Ch. 3), and gives reasons concerning how autonomy contributes additional value to knowledge (Ch. 5).

This work is not only an important contribution to the analysis of knowledge; it has practical import. Technological advances such as Elon Musk's Neuralink device are meant to be implanted in people's skulls to interface their brain with powerful computers via the internet. Neuralink's device will be designed to bring information to subjects' minds and, possibly, to help subjects acquire skills. If the device works as intended, we will be able to acquire information via a device that is not a natural part of us. Will the acquisition of this information give us knowledge? Will we really know how to do whatever it is that Neuralink's device assists us in doing? How we answer these questions is important. Knowledge is, arguably, an important human achievement, and if Neuralink's device (or some other technological assistance) makes it so that we acquire information without knowing, or if it makes it so we do not know how to do what we do as a result of the technological assistance, then we have thereby reduced our ability to succeed in ways important to our humanity. Further, knowledge is arguably required for



moral action. Performing a charitable act in full knowledge (and knowing how to perform the act) is more praiseworthy than performing the act in ignorance. If, however, our reliance on new technology diminishes our ability to act knowledgably, then we are less praiseworthy for having relied on that technology.

Carter begins by arguing that our reliance on an external device such as Neuralink's device for acquiring information thereby prevents us from knowing that information (Ch. 1). Even if we have a justified, non-Gettiered, true belief acquired via the device, we nevertheless do not know that information, because the information was acquired heteronomously, not autonomously. Carter's argument centers on Keith Lehrer's Truetemp case, in which Truetemp, unbeknownst to him, has a device implanted in his head that makes him reliably have thoughts about the temperature without checking a thermometer or other temperature-indicating devices. Mr. Truetemp fails to know, or at least that is the consensus among contemporary epistemologists. Although many of these epistemologists argue that the justification condition on knowledge is not met in the Truetemp case, Carter argues that there are variations of the Truetemp case in which the justification condition is met but in which Truetemp fails to know. For example, Carter gives cases in which the device compels Truetemp not only to believe propositions about the temperature but also to believe other propositions that are good reasons to believe the temperature propositions (11). The device could compel Truetemp to believe everything he believes (12), or the device could remap Truetemp's cognitive architecture such that the device controls his other cognitive dispositions, all of which are integrated with Truetemp's temperature beliefs.

One might be tempted to say that in these latter variations of Truetemp cases, Truetemp does in fact have knowledge. After all, the device seems to be part of Truetemp, integrated with Truetemp in important ways. Just as someone with an artificial heart might have a healthy circulatory system and is counted as pumping blood, so someone with an artificial, implanted cognitive device might have a healthy cognitive structure and be counted as knowing. One might think this is especially true if Truetemp were to have consented to the implantation of the device (as people will with Neuralink's device). After all, if I agree to attend a class with a professor whom I take to be reliable and, as a result, passively receive and instinctively endorse the information the instructor tells me, I still know on the basis of the instructor's testimony. One might think the deliverances of a device are not any different—I might take the device to be reliable and thus passively receive and believe the information the device delivers on the basis of the device's testimony. Nevertheless, Carter maintains that in these latter variations the device

provides Truetemp with processes that we cannot credit to Truetemp and which bypass his cognitive dispositions and traits; because of this, Truetemp fails to know. Further, Truetemp would fail to know even if Truetemp consented to the implantation of the device (31).

Another way one might object to Carter's argument is to maintain that subjects' justification for their belief must belong to them—so autonomy is a condition on justification, not on knowledge as such. Or one might think that to have a belief, the belief must be brought about by the subject's natural processes, so autonomy is a condition on belief rather than on knowledge as such. Such a view would, pace Carter, make it so that the autonomy condition on knowledge is not independent of the justification and belief components of knowledge. Carter addresses extant articulations of the justification component of knowledge but does not provide reasons against the possibilities addressed in these last two paragraphs—that Truetemp does in fact know in the latter variations of Truetemp cases, that autonomy is a condition on the justification component, or that autonomy is required for the belief condition to be met.

What does the autonomy condition on knowledge require? Carter answers this question in Chapter 2. He argues that the autonomy condition requires a causal history that is free of compulsion. To argue for this account, Carter first gives the case of Ann and Beth, psychological twins, who each believe that Tiro is Cicero's scribe. Ann read that fact in a book, but scientists brainwashed Beth into believing it. Ann exercised epistemic autonomy and so knows, but Beth did not exercise epistemic autonomy so does not know (35). Beth's acquisition of the information bypassed or preempted her cognitive competences. If Beth's competences were bypassed or preempted, Beth did not exercise her inner skill to acquire beliefs, was not in proper shape when acquiring the belief (e.g., she was in some way incapacitated or forced into acquiring the beliefs), or she was not in suitable environmental conditions for acquiring the belief (51–2).

Carter's description of autonomy is externalist; the causal history of the belief (being free of compulsion) is external to the subject's present mental structure. Carter says this is how it should be—both internalist and non-historical externalist conditions can all be manipulated by external agents. For some examples, a subject could be compelled to have many beliefs, all of which are mutually supporting, or a subject could be brainwashed to positively endorse a compelled belief were the subject to reflect on it.

Carter maintains that there are some cases in which a subject's cognitive competences are bypassed or preempted but the subject still meets the autonomy condition. These are cases, Carter maintains, when the subject could easily shed the externally-acquired belief if presented

with compelling counterevidence—that is, if the subject is reasons-responsive (48). For this reason, Carter adds: a belief is autonomous if it is not the case that: it is acquired in a way that bypasses or preempts the subject’s cognitive competences and the subject is unable to easily shed the belief. Here’s the final version of the account:

History-Sensitive Externalism about Epistemic Autonomy: S’s belief that p is epistemically autonomous (viz., autonomous in the way that is necessary for propositional knowledge) at a time, t, if and only if p 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 bypassing or pre-emption of such competences issues in S’s being unable to easily enough shed p. (53)

The sheddability addition seems to present a problem for Carter’s project. The sheddability condition amounts to the claim that everyone who is reasons-responsive (so could easily shed the belief) believes autonomously, even if the belief is implanted by an external means. Truetemp’s temperature beliefs would be autonomous if Truetemp would shed those beliefs were Truetemp provided with compelling counterevidence. Truetemp can meet this reasons-responsive condition even in a case in which Truetemp’s cognitive architecture is entirely remapped by an external device in a way not authorized by Truetemp. In that case, the device could make it so that Truetemp sheds beliefs when presented with sufficient evidence against those beliefs (the belief is easily sheddable). Carter, however, says that Truetemp does not believe autonomously in such a case (40). It thus seems to me that Carter should stick to the historical condition and eliminate the sheddability condition.

In Chapter 3, Carter argues that the addition of an autonomy condition on knowledge makes it so that knowledge can be defeated in a novel way. Knowledge is not just defeated by undercutting and rebutting defeaters; there are also heteronomous epistemic defeaters: a new belief that indicates that one’s belief is not epistemically autonomous. For example, a subject has a heteronomous epistemic defeater when the Director of Behavioral Sciences at the FBI tells the subject that his belief is the result of an implanted memory, part of a multi-year experiment.

In Chapter 4, Carter extends the argument to conclude that we do not exercise know-how if the skill is induced via external means such as Neuralink’s device. For example, someone might have a chip implanted in his skull that causes him to make a particular move in chess without knowing why. This action is non-autonomous, and, Carter maintains, the person does not possess know-how. For this reason, Carter states that know-how cannot just involve “some

suitably specified success-oriented disposition” (i.e., skill, 89), because skill can be caused by a mechanism that the subject doesn’t own, but know-how cannot. The skill must have been caused in the right way (autonomously) to provide the subject with know-how. So know-how requires, in addition to skill, an independent component: autonomy.

In a way similar to the objection to Carter’s autonomy condition on propositional knowledge above, someone could argue that subjects do possess know-how even if the skill the subject knows how to perform is caused by a mechanism that the subject doesn’t own, as long as the skill is adequately integrated with the rest of the subject’s abilities. Or someone could maintain that the Neuralinked subject does not exercise know-how, but the failure to exercise know-how is not due to an independent autonomy condition; instead, the subject does not, despite appearances, have the relevant skill, where a subject has the skill only when it issues from the subject’s own abilities.

What does the autonomy condition on know-how require? (Here I assume that know-how is not entirely reducible to propositional knowledge.) Carter maintains that to bring about an action via know-how, the subject must be responsible for bringing about the action. This responsibility requires guidance control: it is caused by a reasons-responsive mechanism that the subject owns. A subject’s mechanism is reasons-responsive if the subject would act differently when presented with good reasons to do so. What does it mean for a subject to own a mechanism? Carter answers: a subject owns a mechanism when the subject 1) reasonably takes herself to be the agential source of the outcomes of the mechanism and 2) reasonably takes herself to be the fair target of reactive attitudes regarding the outcomes of that mechanism.

Carter’s answer here also seems to create a problem for his view, because these ownership conditions can be met by device-induced cognitive dispositions. A device could remap Truetemp’s cognitive dispositions so that he takes himself to be the agential source of the outcomes of the device-induced mechanism, and the device could provide Truetemp with reasons to take himself to be the fair target of reactive attitudes to the outcomes of the device-induced mechanism. A device could make it so that Truetemp takes a competent perspective on the reliability of the mechanism that gives rise to an action in order to see how the action is reliably brought about. These conditions on mechanism ownership seem to me to be internalist, and, as Carter maintains, internalist conditions can be manipulated by external agents, such as implanted devices. One consequence of this is that a device could make it so that the device belongs to the subject. To avoid this consequence, it seems Carter should endorse an external condition on mechanism ownership to maintain consistency with the rest of the project. For

example, Carter could maintain that a subject owns a mechanism when the subject is the agential source of the outcomes of the mechanism and the subject is the fair target of reactive attitudes regarding the outcomes of that mechanism.

These internalist conditions on device ownership eventually lead Carter to maintain that there are some cases in which a subject could have all of their cognitive faculties remapped and have beliefs implanted but still be epistemically autonomous—but only if the remapping occurred long ago. If a subject's cognitive structures have been remapped recently, the subject is not autonomous and so does not know. Why? Carter says that innate faculties, implanted or not, can be relied on in the same way as natural innate faculties (114). Subjects have time and opportunities to calibrate innate faculties, learn about misfires, and so on. So, if someone has had an implanted Neuralink device from birth, that person exercises know-how on the basis of the skill induced by the device. But if a subject has had the device implanted recently, the subject cannot exercise know-how on the same basis. If, after some time, the subject has calibrated, learned how to monitor for misfires, and so on, then the subject can exercise know-how on the basis of the device's contributions. One does not know how at first, but given time, one can acquire know-how on the basis of the device's contributions.

This position, however, is surprising—it seems to make it so that meeting the autonomy condition is merely a matter of meeting the skill condition. Since the subject is able to reasonably rely on the device, calibrate for misfires, and so on, the subject knows. But the subject's reasonable reliance on the device, calibrating, and adjusting for misfires seems to have more to do with tweaking a skill (adjusting for differing contexts, developing its reliability, etc.) than whether the subject owns those mechanisms. As a result, autonomy and skill do not seem to be independent. Here, it seems like Carter has two options. One is to maintain that mechanism ownership does not have to do with reliability, etc. In that case, the position in the paragraph above would need to be revised. The second option is to maintain that one achieves mechanism ownership to the degree to which one's skill has been developed (due to the mechanism's integration with other processes, reliability, etc.), in which case satisfying the autonomy condition is not independent from the skill condition.

Carter's final position also seems to suggest revisions to his positions at the beginning of the book—for example, it might be that Truetemp does have propositional knowledge when a sufficient number of Truetemp's cognitive structures are remapped, are reliable, and are reasonably endorsed upon reflection (Carter's earlier denies that Truetemp knows in this case). If one kind of knowledge (know-how) can be achieved as long as processes (skills) that lead to the target of that knowledge (performance) are made reliable, integrated with other processes,

etc., then it seems that another kind of knowledge (propositional knowledge) can be achieved as long as processes (epistemically justifying processes) that lead the target of that knowledge (belief) are made reliable, integrated with other processes, etc.

In the final chapter, Chapter 5, Carter returns to propositional knowledge, arguing that autonomy adds value to mere justified, true, non-Gettiered belief. This is because autonomy makes a belief ours. Autonomy makes it so that we are tied to particular beliefs—we are epistemic agents with respect to those beliefs. The resulting beliefs have relational value, value that is tied to their making us epistemic agents. Epistemically heteronomous beliefs do not have these features—they do not contribute to our being epistemic agents. Of course, an account of the value that a feature adds to knowledge does not require that the feature is an independent component of knowledge. It might be that autonomy is required for justification or belief, in which case the account of the value of autonomy provided in this chapter adds to the value of justification or belief, not to the account of knowledge as such.

Overall, *Autonomous Knowledge* is a valuable resource for anyone aiming to examine the relationship between knowledge and emerging cognitive enhancing technologies. Carter successfully applies current epistemological work and analyzes accounts of propositional knowledge, know-how, ownership, control, and epistemic value, among others. Carter also arrives at a positive account of propositional knowledge, know-how, and epistemic value that provides an important start to the conversation about whether and how new technologies such as Neuralink's device affect whether we know.