

Partially Autonomous Belief

Lukas Schwengerer

Penultimate Draft – Please cite the published version at

<https://link.springer.com/article/10.1007/s12136-023-00559-y>

Abstract

Adam Carter (2022) recently proposed that a successful analysis of knowledge needs to include an autonomy condition. Autonomy, for Carter, requires a lack of a compulsion history. A compulsion history bypasses one's cognitive competences and results in a belief that is difficult to shed. I argue that Carter's autonomy condition does not cover partially autonomous beliefs properly. Some belief-forming processes are partially bypassing one's competences, but not bypassing them completely. I provide a case for partially autonomous belief based on processing fluency effects and argue that partially autonomous beliefs only amount to knowledge in some cases. I finally suggest how to adjust the autonomy condition to capture partially autonomous belief properly.

Introduction

Adam Carter (2022) recently proposed that any successful analysis of knowledge needs to include an autonomy condition. Especially in light of technological artefacts and the possibility of cognitive integration and extended minds (Clark & Chalmers, 1998; Menary, 2007; Heersmink, 2015) it seems some justified true beliefs fail to be knowledge because they are only products of such artefacts and not products of epistemic agents. Carter shows that falling short of knowledge in these cases is best explained with a new autonomy condition for knowledge. Autonomy is analysed negatively by looking at conditions in which a belief is not formed autonomously. Autonomy is violated if a belief is formed in a way that bypasses the agent's competences and the agent has substantial difficulty in shedding the belief in reflection. Building on Carter's analysis I suggest that bypassing competences can come in degrees. Hence, autonomy also comes in degrees. Some beliefs are only partially autonomous.

They do not satisfy the autonomy condition fully, but they also do not fail to satisfy it completely either. This poses a question: can partially autonomous beliefs amount to knowledge? I suggest that the answer is ‘sometimes.’

The paper is structured as follows: I start with a short summary of Carter’s notion of epistemic autonomy. I then argue that sometimes competences can be partially bypassed by pointing to the effects of processing fluency on our belief formation before I present the case for partial autonomous belief. In the final part, I ask whether a partially autonomous belief can amount to knowledge and suggest an adjustment to Carter’s autonomy condition.

Autonomous Belief

Carter’s autonomy condition is motivated by considering different versions of Lehrer’s Mr.

TrueTemp:

TRUETEMP: Suppose a person, whom we shall name 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 that the tempucomp has been inserted in his brain, is only slightly puzzled 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. Thus, he thinks and accepts that the temperature is 104 degrees. It is. Does he know that it is? (Lehrer, 1990, pp. 163-164)

The almost universally shared intuition here is that Mr. Truetemp does not know that the temperature is 104 degrees. However, there is no consensus on the diagnosis why exactly he does not know. In engaging with versions of TRUETEMP and Bonjour's (1980) related Clairvoyance case plenty of solutions have been suggested. To name a few, some propose that the agent lacks a form of internalist justification (e.g. a version of Alston (1988)). Some propose that the issue is that the process generating the belief is not sufficiently coherent with the understanding the agent has of that belief generation (Kvanvik, 2003). Others suggest that the problem is that the belief is not sufficiently related to Mr. Truetemp's abilities (Pritchard, 2012). Carter carefully shows that for every proposed diagnosis one could form a different version of TRUETEMP that cannot be explained. All of these attempts are epistemic in some sense, but according to Carter, no epistemic condition will do the trick. Something different is called for: the autonomy condition.

Carter's autonomy condition is drawing on Mele's (2001) notion of autonomy. Mele's notion is based on first identifying what compulsion is and then analysing autonomy in terms of lack of compulsion. Carter follows these steps and suggests that a compulsion history of a belief can be identified with two conditions:

- 1) a bypass condition—a condition pertaining to whether the attitude in question was acquired in a way that 'bypassed' the subject's cognitive faculties; and

- 2) an unsheddability condition—a condition pertaining to whether the subject is able to give up, or at least attenuate the strength of, the relevant attitude. (Carter, 2022, p. 45)

The bypass condition captures the history of the belief in question. A belief that is generated without going through my cognitive faculties in the proper way does not have the right history for an autonomous belief. The unsheddability condition then adds a further condition about my ability to change and adjust my belief. If a belief could be easily revised, then even the compulsion history would not be much of a problem. It does not matter how the belief originates if I can just give it up as soon as I evaluate it critically. A truly compulsive belief is one that I cannot give up, or at least cannot give up easily. It sticks, whether I like it or not. In Carter's picture, a compulsive or non-autonomous belief satisfies both bypass and unsheddability conditions. Moreover, the unsheddability is a result of bypassing cognitive faculties. Hence, Carter gives us the first version of his autonomy condition by suggesting that a belief cannot be compulsive in the relevant sense:

History-sensitive externalism about epistemic autonomy (HSEEA): *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 *S*'s cognitive faculties, and (ii) the bypassing of such faculties issues in *S*'s being unable to shed *P*. (Carter, 2022, p. 46)

HSEEA seems promising if we fill in some lacking details of TRUETEMP about the sheddability of the temperature belief. Mr. Truetemp acquires his belief in a way that clearly bypasses his cognitive faculties. And in the original case, it seems to be implied, or at least not implausible

to assume, that Mr. Truetemp cannot easily shed his belief. In case one is not willing to simply read that into TRUETEMP Carter also provides us with a contrasting version in which the agent explicitly is able to shed the belief easily.

TRUETEMP-SHEDDABLE: In the original version of the case, it's not said explicitly, but it is implied, that Mr. TrueTemp can't easily shed this belief (by eradicating it or attenuating its strength). He's at any rate stuck with it. Let's now suppose that, on the present variation of the case, this is explicitly not so. Mr. TrueTemp can easily shed the belief, by simply judging the content to be false (e.g. in light of other things he believes) or otherwise attenuating its strength. Finally, let's suppose he elects not to revise this 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. (Carter, 2022, p. 48)

In TRUETEMP-SHEDDABLE the agent seems to acquire knowledge. The intuition pumped differs from the original TRUETEMP. If this is right, then it supports Carter's diagnosis that bypassing cognitive faculties alone is not sufficient for a non-autonomous belief. Sheddability is relevant as well and HSEEA provides the correct results.

Carter makes some further small changes to HSEEA, of which I want to mention only one that is especially important for building on Carter's account later. He proposes that formulating the bypass condition with cognitive faculties in mind will not work for all cases. Sometimes a belief formation occurs through our cognitive faculties, but these faculties are in that moment not functioning well enough to secure autonomous belief. This can occur either because the agent and their faculties are in bad shape or in a bad situation. Carter presents two respective cases for those problems:

BAD SHAPE: An enhancement team is working to make Julian 'NASA material'. Because Julian is too lazy to properly study, the enhancement team attempts to incept a range of basic aeronautics facts in his head. Out of respect for Julian, the enhancement team permits him to critically reflect on each to-be-implanted fact before it is inserted (and to use any books and online resources when doing so for fact-checking purposes), and to 'veto' any that he does not accept as true or otherwise does not want implanted. Julian, unbeknownst to the enhancement team, is extremely drunk. After critically drunkenly reflecting on each item, he says 'Yes, implant it!'

BAD SITUATION: An enhancement team is working to make Julian 'NASA material'. Because Julian is too lazy to properly study, the enhancement team attempts to incept a range of basic aeronautics facts in his head. Out of respect for Julian, the enhancement team permits him to critically reflect on each to-be-implanted fact before it is inserted, and to 'veto' any that he does not accept as true or otherwise does not want implanted. However, due to time sensitivity, the team will not let Julian consult any books or computers in order to check any of the facts. He is required to rely on only what he already knows. Sober and clear-headed, but feeling under pressure, he does his best to accept for implantation all and only those aeronautical propositions that sound plausible enough to him in light of what he already knows, which isn't much. (Carter, 2022, p. 50)

In both BAD SHAPE and BAD SITUATION, the cognitive faculties are not bypassed, but Julian cannot exercise the cognitive faculties sufficiently. Either because he is too drunk, or because the situation is not suited for a proper exercise of these faculties. One route to go here would be to spell out the bypassing of cognitive faculties such that only properly functioning cognitive faculties matter. It bypasses properly functioning cognitive faculties; hence it

satisfies a so-modified bypass condition. Carter's solution is similar in principle but takes inspiration from Sosa's virtue reliabilism to analyse cases like BAD SHAPE and BAD SITUATION in more detail. Carter replaces talk of cognitive faculties in the bypass condition with talk of *competences*. Building on Sosa (2010) competences are understood as dispositions to perform well in a given domain. Competences have a 'triple-S' constitution: *seat, shape, and situation*. These are best explained with an example. Take the competence in archery. To be a good archer I need the relevant skill of how to move my arms, aim at the target, release the string, and so on, stored in my brain and muscles. This is the seat of my competence. However, this innermost seat of skill alone is not enough. I also need to be in an appropriate shape for that seat to manifest itself in my shooting of the arrow. I cannot be drunk or asleep. An appropriate shape for manifesting archery competence is being awake and sober. And finally, the environment also has to be one in which I can manifest my skill in shooting the arrow. There needs to be enough light to see the target and not too much wind affecting the arrow. I can only manifest my archery competence if everything is right: seat, shape, and situation fit the triggering condition for my competence (Sosa, 2010).

It is easy to see how Sosa's framework gives Carter the tools to deal with BAD SHAPE and BAD SITUATION. If one defines the bypass condition with competences in mind, then both BAD SHAPE and BAD SITUATION turn out to be cases in which the competences are bypassed in virtue of the triggering condition for the competences not being met. Hence, Carter gives us a final account of the autonomy condition:¹

History-sensitive externalism about epistemic autonomy (HSEEA)*:** S's belief that p is epistemically autonomous (viz., autonomous in the way that is necessary for

¹ As mentioned earlier I am jumping from HSEEA to HSEEA***. Carter adds some smaller adjustments that I do not want to reconstruct fully here, as they are less important and require more space.

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 pre-empts 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 . (Carter, 2022, p. 53)

With HSEEA*** in place I can now make the case for partially autonomous belief. In the next section, I show the bypass condition (i) for non-autonomous belief can be partially satisfied. This occurs when cognitive competences are not fully bypassed, but only partially so.

Partially Autonomous Belief

Competences come in degrees. Manifesting competences even more so. There can be ideal conditions for an archer to hit a target, and there can be environmental conditions that are barely good enough for the archer to manifest their archery competence. Not a storm that makes it impossible to hit the target, but a bit of a breeze such that a very skilful archer can still hit most of the time, but a lesser archer cannot. Similarly, being fully awake makes it easy to manifest the competence to its full extent. Being asleep makes it impossible to manifest the competence. But being only a little bit sleepy might still be good enough to manifest the competence, even though perhaps not to its fullest potential. The slightly sleepy archer will hit the target a little less often than the fully awake one. This is something that Sosa already mentions in his virtue theoretic framework. The aptness comes in degree and one shot might be more apt than another (Sosa, 2015, p. 91). The aptness of a shot depends (among other things) on the degree of competence manifested. If this is right, then one should expect this also to be the case for cognitive competences. And if the manifestation of cognitive competences comes in degrees then it seems that whether the bypass condition (i) is satisfied

should not be seen as a binary issue. If the manifestation of a competence comes in degrees, then bypassing competences likely also comes in degrees.

I am going to provide cases of beliefs that are bypassing cognitive competences to some degree, but not fully. These are based on the phenomenon of processing fluency – the speed and ease with which we process information. There is solid empirical evidence that processing fluency impacts the belief-forming processes of agents. I suggest that in doing so processing fluency leads to a partial bypass of cognitive competences.

Let me start with some examples and experiments of the effects of processing fluency on our beliefs. A good overview of the processing fluency literature is presented in Alter & Oppenheimer (2009), I limit myself to three particular experiments to illustrate the issue.

Start with Reber & Schwarz (1999) who tested the influence of the readability of information on the assessment of truth. They presented subjects with simple geographical statements such as “Lima is in Peru” and varied the visibility of these statements by changing the colours used. The blue or red text was highly visible and easy to read on white background. On the other hand, yellow or light blue text was difficult to properly read, because of the lack of contrast to the same white background. The experiment ensured that statements in either readability condition were balanced and no side included statements that were more obviously true. After each presentation of a statement, the subject was asked whether the statement was true. To rule out any sort of manipulation by the subjects Reber & Schwarz came up with a story about reaction times based on colour as the goal of the experiment. What they were actually looking for was a difference in the truth judgments by the subjects that was correlated with or caused by the differences in the colour of the statements. And they found exactly that. Statements that were easier to read were significantly more

frequently endorsed as true. Subjects judged a statement to be more likely to be true when it was written in a colour with a high contrast to the background. Reber & Schwarz's proposed explanation is that processing fluency makes a difference in the assessment of truth. It was easier and faster to read the well-visible colours and that likely caused the higher probability of assigning 'true' to a statement.

A similar effect can be observed with information put forward in the form of a rhyme. McGlone & Tofiqbakhsh (2000) presented subjects with aphorisms that subjects were not familiar with. The list of aphorisms included some in rhyming and some in non-rhyming forms. For instance, "Woes unite foes" and "Woes unite enemies" were both part of the list. As a control measure the list also included pairs in which both aphorisms did not rhyme at all. For each presented aphorism the subjects had to rank the aphorism on a scale based on how accurate they thought the respective aphorism was. The results show that – at least when the subjects were not warned about any potential rhyming effect – the aphorisms that did rhyme were perceived to be more accurate than those that did not rhyme. Similar to the results in Reber & Schwarz a promising explanation for this result is that rhyming aphorisms are easier to process for the subjects. Hence, McGlone & Tofiqbakhsh suggest that "[...] this effect is a product of the enhanced processing fluency that rhyme affords an aphorism such as 'What sobriety conceals, alcohol reveals' relative to a semantically equivalent nonrhyming version" (2000, p. 427).

To end with a third, slightly different example, let me look at Oppenheimer (2006) who argues with experimental evidence that texts that are easier to process are deemed to be written by more intelligent authors. Being easier to process is determined by looking at the complexity of the words used in a text. The more complex words are used, the more difficult a text is to process. Experimental results show that indeed texts with less complex words give rise to

judgments that take the respective author to be more intelligent, compared to judgments based on texts with more complex words. Again, processing fluency seems to influence judgments that seem at least indirectly related to the assessments of truth and competence (Oppenheimer, 2006, p. 151).

Alter & Oppenheimer's (2009) meta-analysis includes many more similar cases, all pointing towards the effect of processing fluency on human judgment. Many of those judgments can be easily linked to belief formation. If a statement is taken to be true in Reber & Schwarz (1999) in part because it is written in high-contrast colours, then that will be part of the causal origin of a corresponding belief. An agent might believe that Lima is in Peru in part because the information was easy to process. There can be similar stories given for statements provided with less complex wording, or in the form of a rhyme. Perhaps the most well-known case of this phenomenon is the illusory truth effect that occurs when the mere repetition of a claim makes the claim appear more likely to be true (Hasher, Goldstein, & Toppino, 1977; Hassan & Barber, 2021). In any such case, the effect on the judgment of truth can also influence the formation of a belief. Hence, at least in some cases, humans form beliefs in part because of processing fluency effects.

With this empirical background, I can come back to partially autonomous belief. It seems clear that fluency effects can often bypass the competences that Carter's autonomy condition requires – at least to some degree. Competences in the sense relevant for belief are dispositions that lead an agent in good shape to truth in appropriate situations. Beliefs aim at truth, and competent beliefs hit that target sufficiently reliably. They need not hit all the time, but often enough. Fluency effects have no relation with truth themselves. The colour of a statement does not make it more likely to be true. As such, the fluency effects cannot be part of a competence that leads to successful beliefs. A competent believer is sensitive to truth-

relevant factors, but not to anything that does not impact the truth of a proposition in question. Hence, a belief that is caused only by fluency effects would not be caused by a competence of the agent. There just is no competence that could lead from fluency to true beliefs reliably enough, because there is no such connection between fluency and truth. A belief that is *solely* based on fluency effects will therefore bypass the agent's competences completely. Fortunately, that is not how fluency effects show up in human life. Processing fluency shows up when we process information and that information itself can go through our competences even though processing fluency influences our judgments. I do not believe a statement only because it is easy to read. I can evaluate the statement, mobilizing my competences of reasoning and memory. The tricky issue is that the empirical evidence shows that mobilizing my competences when thinking about the truth of p cannot keep the fluency effects out completely. My judgment of whether Lima is in Peru is likely a combination of competences and fluency effects. It is not one or the other. It is both that play a causal role in forming the belief that Lima is in Peru. If this is right, then this is a clear example of a case in which a belief partially bypasses my competences. It does not fully bypass my competences, because those competences play a role. However, the belief still bypasses my competences to some degree, because the processing fluency influences whether I form the belief or not (and to what degree). It is possible to create cases in which my competences on their own would barely not form a belief that p, but then the processing fluency gives the final push such that the combination of competences plus processing fluency effects result in me forming the belief that p. These are cases that are paradigmatic candidates for partially autonomous belief.

A slightly different way to describe the case that leads to the same result is to remember that the manifestation of a competence can come in degrees. In some situations, a competence can be fully manifested. In others it can still manifest, but only to a lesser degree. In a situation

that allows an archer to fully manifest their competence, the archer's shot will hit the target (if it hits) solely because of their competence. In a situation in which the archer can manifest their competence only to some degree – perhaps because the weather conditions are very unstable – the archer's shot will hit the target (if it hits) partially because of their competence, but partially because of other factors. A favourable wind might be partially responsible as well. The archer hitting the target is still creditable to the archer to some degree, but they are not fully creditable.² In the same way, situations that generate strong processing fluency effects constitute environments in which one's information processing competences can only partially manifest. One can still manifest one's competences to some degree, but not fully. And forming beliefs based on information processing here is also only to some degree creditable to the agent. The belief bypasses the relevant competences in part.

At this point, I have argued that in some cases relevant cognitive competences can be partially bypassed. And I have illustrated this with cases in which processing fluency influences the formation of beliefs. However, in Carter's autonomy condition bypassing competences is only half of what non-autonomous belief requires. Even if a belief partially bypasses the agent's competences, this may not be an issue at all as long as the belief can be easily enough shed afterwards. What I need for a partially autonomous belief is a case in which the belief formation partially bypasses an agent's competences and that belief is not sheddable. Such cases can be constructed. One way to generate a fitting case is to assume very strong processing fluency effects in a particular case combined with relatively low competences. If the influence of the fluency effects is strong and leads to very high confidence in a belief, then

² Some competitive sports attempt to capture and compensate for the role of the environment in an action that partially manifests competences. For instance, in competitive ski jumping favorable wind has a big impact on the distance jumped by the athlete. To reduce that effect wind is compensated with additional points to the athlete's score.

that belief will be difficult to shed.³ Moreover, if the agent has only low competences that can be reapplied in reflecting on the belief after it is formed, it will be more difficult to shed the belief even if that belief is false. Imagine the following case:

RECESSION Suppose Tom is looking for a job in a country with a general shortage of jobs. He wants to find out why the job market is in such a bad state. Fortunately, in his country elections will take place soon, so politicians with candidate explanations are easily accessible. Some politicians provide very elaborate and complex explanations about an underground conspiracy that controls the job market. These explanations are so intricate that Tom struggles to understand them. Other politicians give a clear, simple explanation: a recession is responsible, they say. This latter explanation is repeated over and over again in the most accessible forms. They even provide their explanation in the form of rhyming slogans all over the country, impossible to miss. “Recession in session” can be read at every truck stop. Tom is interested in political issues and he tries to think the arguments from each party through. The contents of the arguments by themselves are not enough to convince him. However, for Tom, the simple explanations are very easy to process. And given his particular psychology, these explanations just feel obviously correct. On the other hand, the more complex explanations seem confusing to Tom. They appear false to him. Hence, Tom forms the belief that the recession is responsible for the bad job market, and because the arguments for this side seem so much stronger to him than for the other side, he is

³ Note that sheddability is not always related to confidence. One could lack confidence in a belief and it could still be unsheddable. All that matters here is that a high-confidence belief can be difficult to shed. Even if that approach does not work, one could create structurally similar cases in which the unsheddability is established by something external to the processing fluency effect.

very confident in his belief. Given this high confidence, it will be almost impossible to change his belief.

Tom's belief partially bypasses his competences. His competences of (broadly speaking) reasoning and critical thinking are active. He evaluates the arguments from each side. He does not believe blindly. But his evaluation is influenced by processing fluency effects. One explanation is easy to process. It is often repeated, presented in simple words and even in the form of rhymes. Tom's psychology is such that these effects play a large role in his cognitive processes. Hence, the belief formed is partially based on the competences used in evaluating the candidate explanations, but also partially based on processing fluency effects. Both are involved in the causal history of the belief. Moreover, because of his psychology and the far easier-to-process explanation on one side, his belief ends up with a high degree of confidence. So high, that it will be very difficult to change the belief in virtue of reflecting on the belief. A belief held with such confidence is not easy to shed. Tom's belief is therefore both partially bypassing his competences and cannot be shed easily. It is a partially autonomous belief.

I have now shown in this section that there can be partially autonomous beliefs. These are beliefs that bypass an agent's competences to some degree, but not completely, and are not easy to shed. I have made the case for these by pointing to processing fluency, which seems to be a good candidate for a source of partially autonomous beliefs. These fluency effects play a large role in our everyday life and are as such a potential danger to our epistemic autonomy (cf. Schwengerer (2022)) and are therefore worth our attention. However, processing fluency effects are likely not the only way to end up with partially autonomous beliefs. All that is needed for partial bypassing of relevant competences is a shape or situation in which the competence can still be manifested to some degree, but not fully. To end up with a partially autonomous belief one then merely adds a story for how the partial manifestation of the

competence leads to a belief that is unsheddable (or at least not easily sheddable). This gives a basic formula for how to create cases of partially autonomous belief. With a recipe for these partially autonomous beliefs in place, I now want to ask whether partially autonomous beliefs can still amount to knowledge.

Can Partially Autonomous Beliefs be Knowledge?

Carter proposes the autonomy condition as one part of the more complex knowledge condition. He remains somewhat neutral on the other conditions, but let me assume with Carter that knowledge is something close to justified, true, autonomous belief + x, where x is a condition dealing with Gettier cases. Let me look now at whether partial autonomy can be enough for knowledge.

To answer this question, I need to look at partially autonomous cases in which JTBx is satisfied. In many cases of partial autonomous belief, one of the conditions will not be satisfied. Take Tom from RECESSION and let me add a little more to the story. Suppose that the easily understandable explanation of recession is good. It is an explanation that is well reasoned for by its proponents and these proponents are trustworthy testifiers.⁴ Tom evaluates the arguments properly and therefore has justification for his belief. Furthermore, suppose that the explanation was not only good but also correct. The recession is in fact responsible for the bad job market in Tom's home country. In this adjusted RECESSION case Tom's belief is justified and true. Hence, Tom has a justified, true, partially autonomous belief. Does Tom know that the recession is responsible for the bad job market? It does not seem that way. It seems to me that Tom was lucky to be right. The easier-to-understand explanation might have

⁴ I will ignore potential difficulties of a virtue epistemic framework (such as Sosa's view on epistemic competences) with regard to testimony here.

been wrong, and Tom would still have believed it because it was easier to understand. In cases like this JTB is satisfied, but the anti-Gettier condition indicated by the placeholder 'x' is not.

Carter himself argues that there will always be cases in which an agent fails to know even though all JTBx conditions are met. The agent fails to know because the autonomy condition is not satisfied. Even common virtue conditions are insufficient for knowledge without autonomy (Carter, 2022, pp. 15-17). To look into JTBx plus partial autonomy I will use a variation of Carter's case⁵ and adjust it towards partial autonomy.

PARTIAL TRUETEMP: Mr. 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 the device supports his biological systems to detect the room's exact temperature. The device enhances and provides input to the usual cognitive mechanisms for detecting temperature. Moreover, it does so in a way that comes with a high degree of processing fluency, enhancing the processing power and speed of TrueTemp*'s neurological mechanisms. Detecting the exact temperature of the room comes very easily to TrueTemp* thanks to this device. The device is itself auto-integrated with the rest of his (recently neuromodulated) cognitive architecture. A consequence of this 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 influencing these other dispositions that govern the formation and evaluation of his beliefs.

⁵ The base version is Carter's case against virtue theories without an autonomy condition (Carter, 2022, p. 16).

In PARTIAL TRUETEMP the belief is not caused by the device alone but by a combination of the device and TrueTemp*'s biological features. If the neuronal mechanisms were not present, then TrueTemp* would not form a belief about the temperature. However, if the device were not present, then TrueTemp* would not be able to form a belief about the temperature with the same exactness and reliability as he does with the device. TrueTemp*'s belief is not fully coerced, but given the stipulation of especially high processing fluency and TrueTemp*'s psychology, it seems that TrueTemp* is almost guaranteed to form the belief. Does TrueTemp* in this case know that the room has, say, 22.32°C? The answer does not seem obvious to me. The belief is justified, true and not due to relevant luck. Moreover, it is autonomous according to Carter. The way TrueTemp* acquired his belief does not bypass or pre-empt his cognitive competences. His competence in temperature detection is still in play. Nevertheless, the belief is generated in a way in which the cognitive competences are strongly influenced by the device. Carter has to say that we have a clear case in which TrueTemp* knows. But the intuition pumped by PARTIAL TRUETEMP does not support that. It is not clear at all whether TrueTemp* knows – whether the belief was autonomous enough. I take it that the absence of a clear intuition for PARTIAL TRUETEMP indicates that something is missing in the description. It is underdescribed in a way that shows something about the autonomy condition. What we look for is not merely whether the cognitive competences are bypassed or not (plus a sheddability constraint). If it was just bypassing that mattered, then PARTIAL TRUETEMP should pump a clear intuition. I suggest that what is missing is some information about the degree to which the cognitive competences determine the belief that is generated. Let me illustrate this by comparing the earlier case of RECESSION to a different version.

ECONOMIST Suppose Ella is an economist looking for a job in a country with a general shortage of jobs. She wants to find out why the job market is in such a bad state.

Fortunately, in her country elections will take place soon, so politicians with candidate explanations are easily accessible. Some politicians provide very elaborate and complex explanations about an underground conspiracy that controls the job market. These explanations are so intricate that even Ella struggles to understand them. Other politicians give a clear, simple explanation: a recession is responsible, they say. This latter explanation is repeated over and over again in the most accessible forms. They even provide their explanation in the form of rhyming slogans all over the country, impossible to miss. “Recession in session” can be read at every truck stop. Ella is not just interested, but as an economist, she is well qualified to analyse the situation. She tries to think the arguments from each party through and the content of the arguments by themselves *are* enough to convince her that recession is the best explanation. In addition, for Ella, the simple explanations are very easy to process. And given her particular psychology, these explanations just feel obviously correct. On the other hand, the more complex explanations seem confusing to Ella. They appear false to her. Hence, Ella forms the belief that the recession is responsible for the bad job market, and because the arguments for this side seem so much stronger to her than for the other side, she is very confident in her belief. Given this high confidence, it will be almost impossible to change her belief.

In ECONOMIST Ella has a partially autonomous belief, just like Tom had in RECESSION. It is partially autonomous because it is partially based on her cognitive competences, but at the same time also influenced by processing fluency effects. Moreover, the belief is difficult to shed. I take it that Ella clearly knows that the recession is responsible for the bad job market. In this case, the fact that it is merely a partially autonomous belief does not stop her from knowing. The important difference is in the role of the processing fluency effect. For Ella, the

processing fluency influences the belief generation, but it does so in a way that makes no difference to the resulting belief. Ella forms the belief because of her competence, not because of the processing fluency. On the other hand, Tom formed the belief because of the processing fluency, even though his competences were still actively involved in evaluating the arguments. What is needed to distinguish these cases is a clause that identifies when the cognitive competences do not play a decisive role in forming a belief. Carter's bypass condition cannot do that. The bypass condition only asks whether the cognitive competences are involved or not. It does not tell us to what degree or in what form they have to be involved. And it seems important that they are involved such that the cognitive competences on their own determine whether to believe that p or not.

I suggest that this can be adjusted as follows:

History-sensitive externalism about epistemic autonomy (HSEEA)**:** 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 pre-empts S's cognitive competences **or leaves these competences doxastically inert**, and (ii) the bypassing, pre-emption, **or inertness** of such competences issues in S's being unable to easily enough shed P.

I call doxastically inert such cognitive competences that do not determine on their own whether a belief is formed or not. Hence, Tom's competences in RECESSION are doxastically inert, because they do not generate a belief by themselves, even though they are not bypassed. Ella's cognitive competences on the other hand are not doxastically inert. Her competences generate a belief regardless of processing fluency effects. In a way, this

adjustment builds a virtue theoretic idea into the condition for epistemic autonomy. To be autonomous the belief has to be caused by the agent's competences or be sheddable by the agent. It is not enough that competences are not bypassed. They need to be involved in a specific way – as determining factors for belief. This formulation gives the right verdict for all cases and fits into Carter's picture of a view resembling justified, true, autonomous belief +x.

We can use this to now adjust and fully describe PARTIAL TRUETEMP:

PARTIAL TRUETEMP*: Mr. 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 the device supports his biological systems to detect the room's exact temperature. The device enhances and provides input to the usual cognitive mechanisms for detecting temperature. Moreover, it does so in a way that comes with a high degree of processing fluency, enhancing the processing power and speed of TrueTemp*'s neurological mechanisms. Detecting the exact temperature of the room comes very easily to TrueTemp* thanks to this device. ***Because of the ease and speed of judging the room's exact temperature with the device TrueTemp* forms the belief that the room has 22.32°C. If TrueTemp*'s attempt to figure out the exact temperature would have been slower and required more cognitive effort, he would be unsure and not have formed the belief.*** The device is itself auto-integrated with the rest of his (recently neuromodulated) cognitive architecture. A consequence of this 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 influencing these other dispositions that govern the formation and evaluation of his beliefs.

In PARTIAL TRUETEMP* it seems clear that TrueTemp* does not know that the room has 22.32°C. While his cognitive competences play a role in the belief formation, they do not play a decisive role at all. What ultimately determines whether he forms the belief is the processing fluency that comes from the implanted device. His competences stay doxastically inert. The intuition about TrueTemp*'s knowledge is pumped by the question of what ultimately determines his belief, not whether the competences are completely bypassed. If the case was stipulated such that TrueTemp* would form the belief because of his competences and irrespective of fluency effects, then TrueTemp* would know that the room has 22.32°C.

Conclusion

I made a case for partially autonomous belief – belief that is to some degree based on an agent's competences, but to some degree based on other factors. A good illustration of these partially autonomous beliefs is related to processing fluency effects and how these might influence belief formation even though these effects are not competences. I argued that some cases of partially autonomous belief do not seem to constitute knowledge, even when they are justified and true. I suggest that we look at different forms of partially autonomous belief and distinguish between forms that can amount to knowledge and forms that cannot. The relevant difference is to be found when we ask whether the cognitive competences of an agent are inert or play a decisive role in determining the agent's belief. Hence, I adjust Carter's autonomy condition: it is not merely bypassing or pre-emption of cognitive competences that are important, but also whether the competences are doxastically inert or not. Autonomous belief has to involve cognitive competences in a way that gives them the last word.

References

- Alston, W. P. (1988). An Internalist Externalism. *Synthese*, 74(3), pp. 265–283.
- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the Tribes of Fluency to Form a Metacognitive Nation. *Personality and Social Psychology Review*, 13(3), pp. 219-235.
- BonJour, L. (1980). Externalist Theories of Empirical Knowledge. *Midwest Studies in Philosophy*, 5(1), pp. 53–73.
- Carter, A. J. (2022). *Autonomous Knowledge*. Oxford: Oxford University Press.
- Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis*, 58(1), pp. 7-19.
- Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16(1), pp. 107-112.
- Hassan, A., & Barber, S. J. (2021). The effects of repetition frequency on the illusory truth effect. *Cognitive Research: Principles and Implications*, 6(38).
- Heersmink, R. (2015). Dimensions of Integration in Embedded and Extended Cognitive Systems. *Phenomenology and the Cognitive Sciences*, 13(3), pp. 577-598.
- Kvanvik, J. (2003). Simple Reliabilism and Agent Reliabilism. *Philosophy and Phenomenological Research*, 66(2), pp. 451–456.
- Lehrer, K. (1990). *Theory of Knowledge*. Abingdon: Routledge.
- McGlone, M., & Tofighbakhsh, J. (2000). Birds of a Feather Flock Conjointly (?): Rhyme as Reason in Aphorisms. *Psychological Science*, 11(5), pp. 424-428.
- Mele, A. R. (2001). *Autonomous Agents: From Self-Control to Autonomy*. Oxford: Oxford University Press.

- Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbounded*. Basingstoke: Palgrave Macmillan.
- Oppenheimer, D. M. (2006). Consequences of Erudite Vernacular Utilized Irrespective of Necessity: Problems with Using Long Words Needlessly. *Applied Cognitive Psychology*(20), pp. 139-156.
- Pritchard, D. (2012). Anti-Luck Virtue Epistemology. *Journal of Philosophy*, 109(3), pp. 247-279.
- Reber, R., & Schwarz, N. (1999). Effects of Perceptual Fluency on Judgments of Truth. *Consciousness and Cognition*, 8, pp. 338–342.
- Schwengerer, L. (2022). Promoting Vices: Designing the Web for Manipulation. In F. Jongepier, & M. Klenk (Eds.), *The Philosophy of Online Manipulation* (pp. 292-310). Routledge.
- Sosa, E. (2010). How Competence Matters in Epistemology. *Philosophical Perspectives*, 24(1), pp. 465–475.
- Sosa, E. (2015). Knowing full well: the normativity of beliefs as performances. *Disputatio*, 4(5), pp. 81-94.