LUND
UNIVERSITY

Is Externalism Compatible with the KK-thesis?

Balder Ask Zaar

Supervisor: Erik J. Olsson

Magister's Thesis (15 Credits)

FTEM01

Department of Philosophy

Lund University

Table of contents

1. Introduction

The KK-thesis is equivalent to saying that if one knows, one also knows that one knows. It is also known as the principle of positive introspection, meaning simply that one has introspective access to one's states of knowing. Knowledge is, however, not always regarded as an internally accessible state and so positive introspection does not follow trivially from being in a state of knowing. Externalism of justification posits that what makes a true belief into knowledge is something that is external to the mind of the epistemic subject. One should then be skeptical towards the idea that the KK-thesis holds within an externalist framework. If we add the fact that we are human, however, the KK-thesis and the nature of the human being's epistemic situation may make the matter less obvious. This paper thus aims to explore the KK-thesis in relation to externalist conceptions of epistemic justification. The attempt will be to answer the question: Is externalism regarding epistemic justification compatible with the KK-thesis? At first glance, as established, this seems highly implausible. Surprisingly, though, their compatibility has nonetheless been argued for by some proponents of externalism. The goal of this paper can be said to be threefold: (1) to evaluate the arguments posited for and against the KK-thesis' compatibility with externalism, which then allows a discussion of (2) whether knowledge entails a high-order state accessible to the knowing subject, and if (2) is answered negatively in some sense, then (3) figure out how the KK-thesis can be modified and what the externalist needs to posit in addition to their truth-conditions for knowledge in order to make the KK-thesis plausible given their theory justification.

First and foremost, this holds relevance to epistemic logic and whether principles such as the principle of positive introspection hold, and generally what can be logically entailed from being in a state of knowing. If the KK-thesis were compatible with externalism, it could possibly bear on the value of knowledge as well, since knowing would then entail that a knowing subject were also in a high-order state where the relation between subject and world is accessible to the knower. States of knowing that are accessible, or known, to the knowing subject seem to be more valuable than knowledge in that they allow for more rational and responsible decision-making – when the latter is based on one's epistemic situation. Of course, this is context dependent. Oftentimes knowledge is implicit or latent; we may be in situations where we act on instinct before we are consciously aware of the reasons for our actions without there being any punishments for such instinctual behavior; we may be in situations where we are not facing significant risk were our epistemic states not reflected upon. In these kinds of scenarios, the value of knowing that you know may be on par with merely knowing. The value

of knowing that you know is best seen in high-risk situations that require certainty in order to best decide on how to act. But one could also regard the value of knowing that one knows through the virtues inherent to such high order states. As Sosa (2001) points out in his discussion of the difference between reflective and animal knowledge:

> *[W]e are often interested not only in having the truth but in discovering it, which involves not just being visited with the truth by sheer happenstance or through some external agency, but to arrive at the truth through our own intelligent doings, by relying on our own reliable abilities, skills, and faculties.*
>
> *[…]*
>
> *Prominent among values that constitute the higher, reflective level [of knowledge] is that of understanding. But this does not preclude a correlative level of knowledge allied to such understanding. It is in part because one understands how one knows that one's knowing reaches the higher level. A belief constitutive of such reflective knowledge is a higher epistemic accomplishment if it coheres properly with the believer's understanding of why it is true (and, for that matter, safe), of how the way in which it is sustained is reliably truth-conducive.*

Knowing that one knows, seems to be close to this kind of higher state of reflective understanding, as will hopefully seem clearer through the discussions of this paper. A given theory of knowledge's relation to the kind of reflective knowing involved in having knowledge of one's state of knowing then seems relevant for the sole purpose of understanding what it means to be in a state of knowing. Whether it is possible to derive that one knows that one knows from the fact that one knows (as a human being) should then concern anyone concerned with understanding what it means to know. Exploring the relation between externalist theories of knowledge[1] and the KK-thesis should shed light on what it means to be in an epistemic situation, answering the question: what are some of the cognitive implications for a self-aware entity that is in a state of knowing?

The paper will begin with some background of internalism and externalism and how and why they can be deemed more or less compatible with the KK-thesis. Three externalist theories will briefly be discussed, a causal theory of knowledge, a reliabilist theory of knowledge and Williamson's theory of knowledge that posits that knowing is a mental state (yet still externalist). The KK-thesis is first discussed without assuming its independent plausibility by

[1] 'Externalist knowledge' and 'externalist theories of justification' are in this paper taken as equivalent.

exploring Williamson's anti-luminosity arguments and some responses to it. It is concluded that Williamson's anti-luminosity argument ultimately fails to consider a crucial distinction between reflective and perceptual knowledge. The paper then turns to the independent plausibility of the KK-thesis given the different forms of externalism brought up in section 1.1. It is found that reliabilism and a causal theory of knowledge are more plausibly compatible with the KK-thesis than Williamson's theory (given the nature of knowledge and human metacognition). It ends with a discussion of the possible ramifications this can have for the internalist, and further remarks on the distinction between animal and reflective knowledge not being a distinction regarding knowledge itself, but a distinction that stems from whether or not one has the capacity for metacognition, and thus understanding. The ultimate claim is that, for human beings in relatively normal conditions, the KK-thesis is valid even given some externalist theories of justification.

1.1 Background – Internalism, Externalism, and their Relationship to the KK-thesis

First, in order to understand in some depth why the KK-thesis and externalism likely are incompatible, a brief explanation of the KK-thesis and externalism is required. The KK-thesis as a logical principle was originally defended in Hintikka (1962). With Hintikka's notation, it is expressed as follows (where K stands for 'knows', a is a subject, and *p* a true proposition):

(KK) $K_a p \rightarrow K_a K_a p$

This is read as 'If *a* knows that *p*, then *a* knows that *a* knows that *p'*. The KK-thesis is coupled with what Hintikka calls a 'strong sense of knowing'. The latter is what one has if one denies that any new information could lead to the alteration of one's view (ibid., p. 20). The weak sense of knowing is what we get by viewing knowledge as mere true belief (Hintikka, 1970). Considering that many externalist theories of justification may be regarded as a kind of middle ground between reflective internalist conceptions of knowledge and knowledge theories that reduce knowledge to true belief, it is not entirely clear whether the KK-thesis holds given externalist assumptions about knowledge (as well as certain other facts about being a knowing human being, including self-awareness, background knowledge about one's situations, and so on).

In any case, the principle has some clear implications. The first implication is perhaps that it assumes that justification is something internal to the mind, but also that the mind itself is Cartesian. What is now called Cartesian is not clearly attached to its original source, so a

rather brief overview of the Cartesian view of mind and justification should prove useful as a contrast to externalist point of view. We can then see how they each relate to the KK-thesis.

What does it mean for the mind to be Cartesian? And what does it mean for a theory of justification to be Cartesian? First, it seems to mean that we have privileged access to the contents of our minds, which makes this the most secure foundation of all knowledge claims. Mental states are thus said to be transparent, self-evident, self-presenting, self-intimating, accessible, luminous, and so on. Of course, these are used somewhat differently, but the key terms here are 'self-intimation' and 'transparency'. The self-intimation of mental states means that if one has a certain mental state, then this is made known to the possessor of that mental state. So that, if we *are* sad, this state is self-intimating in that it is brought to our conscious awareness and so we *feel* sad. Likewise, this would then mean that if we were in a state of knowing and knowing involves some mental states (like being justified and having a belief), then we would in principle have access to the contents of these states. The only evidence of the idea of self-intimation being specifically Cartesian (that I could find) comes from Descartes in Objections and Replies (2008, p. 158) where he writes:

> *Moreover, that there can be nothing in the mind, in so far as it is a thinking thing, of which it is not conscious, seems to me to be self-evident, because we understand there to be nothing in the mind, considered from this point of view, that is not a thought, or dependent on thought; for otherwise, it would not belong to the mind, inasmuch as it is a thinking thing; and there can be no thought in us of which, at the moment at which it is in us, we are not conscious.*

The relevance of Cartesianism has further significance for knowledge, specifically, by serving as the distinguishing factor for two contrasting epistemic starting points. In Our knowledge of the internal world (2010, p. 2-4) Stalnaker highlights the contrast between Cartesian internalists and anti-Cartesian externalists. The former begins all theorizing with the contents of our minds and attempts to construct an external world (or a conception of the external world), the latter 'starts in the middle' and thus begins with some notion of the external world in order to explain how various mental states the world can be about (or correspond to) the world. Internalists are Cartesian, presumably, because they aim to construct concepts from the infallible self-intimating contents of their mind, in line with what Descartes writes in his Meditations (2008, p. 21):

> *Is there any of these things that is not equally true as the fact that I exist—even if I am always asleep, and even if my creator is deceiving me to the best of his ability? Is there any*

> *of them that can be distinguished from my thinking? Is there any that can be said to be separate from me? For that it is I that am doubting, understanding, wishing, is so obvious that nothing further is needed in order to explain it more clearly. But indeed it is also this same I that is imagining; for although it might be the case, as I have been supposing, that none of these imagined things is true, yet the actual power of imagining certainly does exist, and is part of my thinking. And finally it is the same I that perceives by means of the senses, or who is aware of corporeal things as if by means of the senses: for example, I am seeing a light, hearing a noise, feeling heat.— But these things are false, since I am asleep!—But certainly I seem to be seeing, hearing, getting hot. This cannot be false. This is what is properly meant by speaking of myself as having sensations; and, understood in this precise sense, it is nothing other than thinking.*

Noteworthy is also the quote from *A Discourse on the Method* (Descartes, 2016, p. 17), where Descartes explicitly says the first principle of his philosophical method is the following:

> *The first [principle] was never to accept anything as true that I did not incontrovertibly know to be so; that is to say, carefully to avoid both prejudice and premature conclusions; and to include nothing in my judgements other than that which presented itself to my mind so clearly and distinctly, that I would have no occasion to doubt it.*

If we bring together the self-intimation of mental states and the Cartesian internalist and foundationalist methodology, the KK-thesis seems to follow quite naturally. This can perhaps be made more understandable with an argument. Given an internalist conception of knowledge and an anti-realist conception of the world (P2–P4 are presumed to be internalist assumptions), the following argument can be presented:

> (1) Knowledge, in the most general sense, is a correspondence relation between the mind of subjects and the world.
>
> (2) The correspondence relation, given internalism, amounts to a comparison between a conception of the world and a cognitively accessible state of knowledge.
>
> (3) The world is conceptualized through mental means (we can perhaps say that, to the internalist, both the methodology of constructing a conceptualization of the world and the building blocks themselves are mental).

(4) Epistemic mental content is self-intimating,[2] that is, if a subject is in a state of knowing, the subject cannot fail to have access to the fact that they are in such a state.

(Conclusion) Therefore, if a state of knowledge is instantiated, it implies that there is a relation between the mind and the world such that the mind, the world, and the relation between them are all accessible to the mind in a self-intimating manner. Since this presentation itself is what constitutes second-order knowledge, knowledge entails knowing that one knows.

This, I believe, is the underlying argument from introspection that buttresses the notion that 'knowing that *p*' is "virtually equivalent" (Hintikka, 1962, p. 104) to 'knowing that you know that *p*' from a Cartesian point of view. It seems to hold true because second-order knowledge, given internalism, is not revealing any additional information about the subject's mental state (since knowledge itself is an awareness of the correspondence relation between mind and world). Saying that someone knows that *p* and that they know that they know that *p* amounts to more or less saying the same thing. If $K_a p$ is equivalent to $K_a K_a p$, then of course $K_a p \rightarrow K_a K_a p$ holds.

While this is what allows the KK-thesis to hold almost trivially, let us also look at some classically internalist but not so radically Cartesian conceptions of knowledge and see how they relate to the KK-thesis. Being an internalist does by no means assure the validity of the KK-thesis, but on some conceptions of the JTB-variety, it is valid. Take for instance Chisholm's definition of knowledge in Theory of Knowledge (1989, p. 98):

[2] Chisholm (1981, p. 80) has a very useful formulation for this Cartesian style self-intimation (only he calls it self-presenting), it goes: "We will assume that every self-presenting property is necessarily such that, if an individual has it, if he considers his having it (that is, if he thinks of himself as having it), then, ipso facto, he will directly attribute it to himself. […] Thus we could say that a thing is conscious if and only if it has a self-presenting property." He goes on with some more precision, adding the notion of certainty (ibid., p. 82): "For every x, if (i) x has the property of being sad, and if (ii) x considers his being sad, then it is certain for x that he then has the property of being sad." Gilbert Ryle in his The Concept of Mind (2009) gives another early and very clear account of self-intimation: "The states and operations of a mind are states and operations of which it is necessarily aware, in some sense of 'aware', and this awareness is incapable of being delusive. The things that a mind does or experiences are self-intimating, and this is supposed to be a feature which characterises these acts and feelings not just sometimes but always. It is part of the definition of their being mental that their occurrence entails that they are self-intimating. If I think, hope, remember, will, regret, hear a noise, or feel a pain, I must, ipso facto, know that I do so."

> h is known by S $=_{df}$ (1) h is true; (2) S accepts h; (3) h is evident for S; and (4) if h is defectively evident for S, then h is implied by a conjunction of propositions each of which is evident for S but not defectively evident for S

Chisholm writes (ibid., p. 11) that "The evident is that which, when added to true belief, yields knowledge." It becomes important now to also define what it means for a proposition to be evident. Chisholm defines this in the following way (ibid., p. 11):

> *p* is evident for S $=_{df}$ for every proposition *q*, believing *p* is at least as justified for S as is withholding *q*.

This definition merely points out that if a proposition is evident then it is more justifiable for one to accept the proposition than it would be to withhold it. Since the definition of knowledge Chisholm posits says "h is evident ***for*** S", there seems to be an implicit awareness of the evident nature of the proposition. Since the evident is what makes a true belief into knowledge, and what is evident is specifically *for* the epistemic subject, it seems to follow that knowledge also entails knowing that one knows. But let us show this in detail with inspiration from Hilpinen (1970). Hilpinen's article Knowing That One Knows and the Classical Definition of Knowledge identifies two general ways a standard JTB theory of knowledge can be interpreted. The main difference between the justificatory aspect of knowledge conditions in different JTB theories seems to be the knowing subject's awareness of their own grounds for knowing. The question is whether one can be justified without knowing that one is justified, and this entirely depends on the definition of knowledge. If *p* is evident *for* one who knows, then it seems likely that the evident is meant to be regarded as a self-intimating state. To see what happens if evidence is not stipulated as being *for* the knowing subject, take an earlier definition by Chisholm (1957, p. 16):

> "S knows that *h* is true" means: (i) S accepts *h*; (ii) S has adequate evidence for *h*; and (iii) *h* is true.

The difference here is that having adequate evidence does in no way necessitate that one takes this evidence into account. It does not mean that having adequate evidence is self-intimating, nor that there is nothing blocking S's access to the evidence for *h*. Hilpinen (1970) formalizes the idea that the evident is self-intimating by the implication $Kp \rightarrow BEp$. If one knows that *p*, then one also believes one has evidence for *p*. It should be illustrative to show formally how the KK-thesis holds under this assumption (as well as a few more rather plausible assumptions),

so let us follow Hilpinen's (1970, pp. 115-116) presentation of the validity of the KK-thesis given evidentiality that is self-intimating:

> Having added BE*p* as a consequence of knowing, knowing that *p* can be defined with the following equivalence relation (where K = knowledge; B = belief; E = evidence for):
>
> (DK) K*p* ↔ B*p* ∧ E*p* ∧ BE*p* ⊿ p
>
> The KK-thesis entails the knowledge of each conjunct (assuming that the knowledge operator K distributes over a conjunction, that is if one knows a conjunction of propositions one knows each conjunct):
>
> (1) KK*p* ↔ KB*p* ∧ KE*p* ∧ KBE*p* ⊿ Kp
>
> This in turn implies the cumbersome:
>
> (2) KK*p* ↔ K*p* ∧ BB*p* ∧ EB*p* ∧ BEB*p* ∧ EE*p* ∧ BEE*p* ∧ *BBE*p ∧ EBE*p* ∧*BEBEp*
>
> With this expanded notation, the KK-thesis amounts to the following:
>
> (3) B*p* ∧ E*p* ∧ BE*p* ⊿ p → BB*p* ∧ EB*p* ∧ BEB*p* ∧ EE*p* ∧ BEE*p* ∧ *BBE*p ∧ EBE*p* ∧ BEBE*p*
>
> Another set of plausible assumptions can now validate the KK-thesis. First assume (4) B*p* → EB*p*, that is, that a belief in *p* entails that the belief in *p* is evident. This follows from various rational principles such as *one should only believe that which one has evidence for*. Then there is (5) E*p* → EE*p*. That is, if *p* is evident, it is also evident that *p* is evident. This comes from the idea that if one arrives at the notion that *p* is evident, one has sufficient information to think that it is evident that *p* is evident. Finally (6) B*p* → BB*p*, which is said to be self-sustaining through positive introspection (Hintikka, 1962, pp. 109-110). So far, given (4)–(6), the antecedent of (3) gives us five out of the eight conjuncts, that is B*p* ∧ E*p* ∧ BE*p* ⊿ p entail:
>
> BB*p* ∧ EB*p* ∧ EE*p* ∧ BBE*p* ∧ EBE*p*
>
> It remains to be seen how BEE*p* ∧ BEB*p* ∧ BEBE*p* are implied by knowledge. Hilpinen adds the following principle in order to make this possible: (7) If *p* → *q* is valid, then B*p* → B*q* is valid. Now from (5) and (7) we can also derive (8) BE*p* → BEE*p*. From (4) and (7) we get (9) BB*p* → BEB*p*. Since B*p* implies BB*p* by

(6), we also get (10) Bp → BEBp (by transitivity). Substituting p for Ep in (10) gives us the last conjunct: (11) BEp → BEBEp.

We can perhaps safely say that, despite the intuitive notion of there being a relation between the KK-thesis and classical internalist conceptions of knowledge, it is nonetheless far from trivial to spell this relation out in a more detailed fashion. Regardless, this further illustrates the point that what makes the KK-thesis plausible for the internalist is that justification is intertwined with the epistemic agent's (conscious) mind.

Let us turn now to externalist theories of justification. Externalism about epistemic justification posits that the justification for our knowledge is external to the mind of the knowing subject. That is, justificatory externalism is the view that in order for a true belief to amount to knowledge, the true belief has to come about in a certain way. Various theories of non-inferential knowledge (Grice, 1961; Goldman, 1967; Armstrong, 2000) based on causal criteria is one form of externalism. For instance, Goldman (1967, p. 369) in A Causal Theory of Knowing gives the following truth-conditions for "S knows that p":

> *S* knows that p if and only if the fact that p is causally connected in an "appropriate" way with **S's** believing that p.

The contrast with Cartesian theories of knowledge is clear. Here our knowledge is not determined by the contents of our minds. For knowledge to be instantiated in the causal theory of knowledge, certain conditions in a mind-independent external world have to obtain. That is, a causal connection between the belief that p and fact that p has to be in place. One does not need to be aware of this causal connection in order to be in a state of knowing. If one wants to know that they know, however, this seems to be a requirement. In order to know that one knows given a causal theory of knowledge, we have to have knowledge of the causal process(es) that result in the formation of the corresponding belief. Since these causal processes are in part external to the mind, it would be absurd to call them self-intimating in the Cartesian sense. The question would be whether a causal process connecting a mind and a fact somehow carries information not only of the fact through corresponding belief, but information about the connection between the fact and belief itself. That is, does the causal process carry information about the connection between the belief and the fact? It does not appear to be that way, at first glance, since the knowing subject (given a causal theory of knowledge) only has transparent access to the belief; that is, the belief is the only truly self-intimating aspect of the state of knowing. To find out whether there is in fact a causal connection between the fact that p and the belief that p, one would first have to reflect on possible ways in which such a connection

can arise, and then whether it really has been instantiated. The difference between obtaining second-order knowledge for internalists and externalists, then, seems to be that given externalism, epistemic agents have more work to do over and above obtaining knowledge in order to obtain second-order knowledge, whereas the knowing subject given internalism gets a form of epistemic free lunch (second-order knowledge) as soon as the truth-conditions for knowledge are fulfilled (at the very least, if one is in a state of knowing then one is in a position to know that one knows by way of a priori reasoning).

Reliabilism effectively illustrates the idea that second-order knowledge requires something over and above the state of knowing itself, at least in some formulations. One early formulation of reliabilism comes from Ramsey (1931, p. 258) who writes that: "I have always said that a belief was knowledge if it was (i) true, (ii) certain, (iii) was obtained by a reliable process." Ramsey connects certainty with knowing that one's process of obtaining a true belief is reliable *upon reflection*. He says (ibid.): "We say 'I know', however, whenever we are certain, without reflecting on reliability. But if we did reflect then we should remain certain if, and only if, we thought our way reliable." Here Ramsey seems to be getting at the same idea mentioned above, namely, that externalist knowledge requires something over and above the state of knowing in order to obtain second-order knowledge. Note that Ramsey is not saying that reflection is necessary in order to be certain, only that once one reflects upon one's state of knowing (and thus attempts to figure out whether one know that one knows), one is forced to consider the conditions necessary for the acquisition of a true and certain belief (that is, the reliability of the way one obtained the true belief). Two epistemic modes can be identified here corresponding to knowledge and second-order knowledge. Namely, a non-reflective mode enough to obtain knowledge and a reflective mode that is necessary for obtaining second-order knowledge. The difference between internalism and externalism can then be further specified with these different epistemic modes. The Cartesian internalist wants a state of knowing to be itself reflectively accessible, whereas the externalist does not take this kind of reflection to be necessary in order to obtain knowledge (this is explicitly said in Goldman (1979), where he writes that "I have denied that justification is necessary for knowing, but here I had in mind 'Cartesian' accounts of justification'"). The emphasis on reflection can also be said to stem from private versus public views on knowledge. From the Cartesian perspective knowledge is first personal and private. To the externalist, our purely private mental states cannot amount to knowledge, knowledge is instantiated as a relation between entities in a world, regardless of whether the relation is apprehended by some mind.

Perhaps this can suffice to explain the difference between internalism and externalism as they relate to the KK-thesis. It is a matter of the degree of reflection involved in obtaining knowledge which seems to determine the plausibility of the KK-thesis. Externalist theories of knowledge require very little (if any) reflection of their epistemic agents aiming to obtain knowledge, whereas internalist theories require that their epistemic agents reflect on their epistemic situation in order to obtain knowledge. Considering that externalists do not require this kind of reflection and that the truth-conditions for externalist theories of knowledge involve mind-independent elements (that then are not self-intimating), it seems that a lot more would have to be said in order for the KK-thesis to be compatible with externalism.

What has to be in place in order to make them compatible? The KK-thesis could be made weaker, so that it no longer is presented as being a relation of logical entailment between S's knowledge that *p* to S's knowledge that S has knowledge that *p*. It could also be possible to give arguments for the perhaps perceptive or reflective salience of reliable processes of belief-formation and causal connections. That is to say, perhaps these processes are not self-presenting in a Cartesian manner, but so salient in the environment that any epistemic agent that is in a state of knowledge and is self-aware has been brought to awareness of the fact that the conditions of knowledge are fulfilled (either by being able to identify the reliability of the process of obtaining a certain belief or the causal process that led to the formation of the belief). Both strategies seem rather dubious, but as we shall see they are precisely the type of strategies that are adopted in defense of the compatibility between the KK-thesis and externalism.

1.2 Knowledge as a State of Mind

Before discussing a few attempts to defend the compatibility, first a couple of arguments against knowledge being luminous, as argued by Timothy Williamson, will be presented. The compatibility between the KK-thesis and externalism does not only have to countenance the general aversion towards reflection seen in some of the original externalist theories of knowledge, but they also have to deal with some convincing arguments against the idea that mental states are luminous (transparent, self-intimating, etc.) altogether. Williamson argues specifically from the point of view of an externalist conception of knowledge – which will prove relevant for the discussion regarding the compatibility between the KK-thesis and externalism – but he does this in a particular sense. Knowledge, in his view (Williamson, 2000), is a most general factive propositional attitude. His theory remains externalist, however, since he also argues for externalism regarding propositional attitudes. He claims propositional attitudes are so-called *broad* states of mind. To understand what a broad state is, we first need

to understand what Williamson calls a 'system' and what he calls a 'case'. A system is an agent at a given time. A case is a system paired with its environment. A case, then, can be said to contain an internal (system-internal) part and an external (environmental, system-external) part. For each case, we then have *conditions* that either obtain or fail to obtain. Now the difference between narrow and broad conditions can be explained schematically in the following way (quoting Williamson, 2000, p. 52):

> *A case α is internally like a case β if and only if the total internal physical state of the agent in α is exactly the same as the total internal physical state of the agent in β. A condition C is narrow if and only if for all cases α and β, if α is internally like β then C obtains in α if and only if C obtains in β. […] narrow conditions supervene on or are determined by internal physical states: no difference whether they obtain without a difference in that state. C is broad if and only if it is not narrow.*

Knowledge, on this conception, is a broad condition. This means for example that if two agents S_1 and S_2 have visual experiences that *p*, but S_1 is hallucinating whereas S_2 truly sees that *p*, these conditions can be exactly the same internally without thereby implying that the two subjects are in the same mental state. Seeing that *p* and hallucinating that *p* are simply two different conditions.

Williamson emphasizes this with his notion of primeness (ibid., chap. 3). Not only are knowing states of mind broad, but they are prime. That a propositional attitude is prime means that it is not composed of internal and external conditions. A composite state, on the other hand, is a conjunction of an internal and external state. The point of primeness seems to be to show that knowledge, as a prime condition, is something over and above something like a true belief, or any other accidentally combined external and internal condition. Knowledge being prime, can more specifically be said to mean that if there are three cases α, β and γ, where γ is internally like α and externally like β, there is a condition that obtains in α and β but not γ. An example of this is if we take α and β to be cases in which a person sees a glass of water, and assume γ is the combination of the internal condition of α and the external condition of β. Then, let us say α is a condition where there is a glass of water on the right and a glass of gin on the left, but the perceiver has a brain lesion that causes her to only visually register what is on the right. In case β the is gin on the right this time, and water on the left, with a brain lesion that causes the perceiver to only register what is in her left field of vision. If γ takes the internal state of α, we get a brain lesion that prevents vision of all things on the left. If we combine this with the

external condition from β, that there is gin on the right, the condition of seeing water no longer obtains, since the perceiver is now perceiving a glass of gin, not water.[3]

With this Williamson takes himself to have shown that knowledge can both be non-composite (that is, prime) and a mental state while also maintaining broadness and thus externalism. If knowledge is a mental state yet stays within the purview of externalism, perhaps the KK-thesis and externalism are not so incompatible. We have relatively privileged access to our propositional attitudes, after all (compared to access to, let us say, the propositional attitudes of other people).[4] This seems to be one of the better opportunities for externalists to argue for the compatibility between externalism and the KK-thesis.

Yet the idea that we have privileged access to our mental states is not uncontroversial. Around 50 years before Williamson's anti-luminosity arguments there was already the deeply anti-Cartesian account of the mind espoused by Gilbert Ryle. There are many notions in Ryle's The Concept of Mind (2009, pp. 138–145) that show the difficulty in reconciling self-knowledge of mental states with the KK-thesis. For instance, he rejects the idea that being conscious of one's mental state is having knowledge of them to begin with. He claims that it would be unreasonable to add accusatives expressing mental states to a verb phrase containing 'to know' (or 'knows that…). For instance, Ryle claims it is nonsensical to say that one knows a feeling of pain or knows a colored surface. Knowing, to him, is knowing that something is the case. That is, knowing that a colored surface is a cheese-rind, that the feeling of pain is sensory information carried by nerve fibers, and so on. In Ryle's view, then, as opposed to Williamson's view, sense experience itself is not a way of knowing. The contrast between Williamson's and Ryle's view on sense experience serves to highlight the distinction between propositional and non-propositional ways of perceiving. This is also a distinction that has been discussed in some recent literature on the meaning of perception verbs such as 'sees' or 'hears' (French, 2012). Specifically, French claims that 'sees that…' may have a visuo-epistemic sense that entails propositional knowledge ('knows that…'). But then we are no longer talking about sense experience as in the process of visual perception itself, but a way of perceiving that entails the accurate conceptual categorization of the objects of perception. So that in the case of simply seeing a colored surface, we do not say that this is propositional perception, but when we identify the colored surface correctly as cheese-rind, we may have achieved propositional

[3] See Williamson (2000), pp. 66-72 for the full account of primeness.
[4] This is not uncontroversial, however. For instance, Gopnik (1993) claims that self-knowledge of one's mental state is acquired through a process similar to what one would use to inquire about the mental state of other people.

perception in a way that entails knowing by way of a visual experience that there is some cheese-rind in front of us.

Along similar lines, Ryle notes that we often fail to recognize a given state of mind for what it is, as well as often think we know something when we in fact do not. We can, for example, be surprised when a clock stops ticking even though we were previously not aware of it ticking (the implication being that we were processing it after all, only without being aware of it). We may also think we are dreaming we are awake, and vice versa. Mental states cannot, given these facts, be seen as (at least not perfectly) self-intimating. The proviso, then, is that *if we recognize the mental state we are in for what it is,* then we are perhaps in a position to know that we know (if we are in a state of knowing and knowledge is a mental state).

If the mental state in question is a propositional attitude, however, then it involves a proposition and thus Ryle's objection to the self-intimation of mental states does not seem to have bearing on the type of mental states Williamson regards as states of knowing. A propositional attitude can perhaps be said to have a higher degree of attentional focus and conceptual sophistication than the background noise one hears while absorbed in some more immediately relevant task (i.e., the ticking clock). The latter kinds of experiences are perhaps best described as non-propositional or inattentional. Given that propositional attitudes seem to be at least more explicit than mere sense experience and imply a kind of attention directed towards the proposition (one could perhaps call it intentional, in the phenomenological sense), the following question still seems pertinent: is knowledge a state (either mental or a state consisting of mental and non-mental conditions) that entails that a knowing subject recognizes what kind of state they are in? If so, is this recognition a way of knowing that one knows?

Williamson both does and does not seem to think that propositional attitudes have this added attention and recognition. Despite the relative plausibility of a propositional attitude being self-intimating, he presents two arguments against luminosity. While Williamson argues against the KK-thesis, he does so not by expanding on what it would mean to know that one knows given that knowledge is a mental state. His approach is a reductio ad absurdum argument aiming to dismiss the KK-thesis outright. The following section will present both of his arguments against the KK-theiss, after which some attempts to salvage the compatibility of externalism and the KK-thesis will be discussed in more detail.

2. Anti-Luminosity and the Vagueness of Mental States

The reason Williamson argues against the KK-thesis seems to stem from the fact that he wants to reconcile the idea that knowledge is a mental state with some of the externalist assumptions

regarding knowledge. The internalist presumably thinks that, like hinted at above, there are mental states that are so central to a subject's cognitive apparatus that they become fully self-intimating. Williamson believes this is false. That is, he believes there are no core set of types of mental states that are such that, once one is in a mental state of one of these types, one is always in a position to know that one is in such a mental state. The reader may notice that this is a slightly different version of the KK-thesis. That is to say, there is a difference between the formulations "If S knows that *p* then S knows that S knows that *p*" and "If S knows that *p* then S is in a position to know that S knows that *p"*. In Williamson's formulation (if S knows that *p* then S is in a position *to* know that S knows that *p*), there is no longer a requirement on the epistemic subject that they know that they know that *p.* They merely have to be in a position to know that they know that *p.* But this does not mean we have abandoned the KK-thesis. The distinction seemingly only due to the fact that in Hintikka the KK-thesis is based on an ideally rational agent, whereas in Williamson this has likely been rejected as too demanding of any real person (see also Stalnaker, 2019, p. 38). In the one case, the epistemic agent reflecting on their epistemic situation is taken for granted, in the other it is not. The formulations can be said to denote situations that differ only in whether a given epistemic agent has asked themselves what kind of mental state they are in at any given moment. The kind of reflection then required to know that one knows once one is in a position to know that they know should be free of obstacles (given somewhat normal cognitive conditions), even given Williamson's formulation. It is maybe possible to claim that there is no substantial difference between knowing that one knows that *p* and being in a position to know that one knows that *p*, considering the relative ease with which one inquires into one's own mental states. The interpretation of 'S being in a position to know that *p*' will then be that *there are no obstacles preventing S from knowing that p, given S's interest in obtaining this knowledge as well as inquiry into whether they have knowledge of p.*[5] Pertaining to the aim of this paper, then, we see that if an argument is posited that shows that one can know without being in a position to know that one knows, then this is tantamount to showing that knowing does not entail knowing that one knows (considering that Williamson's formulation is a weaker – in these sense of 'less demanding' – version of the KK-thesis). If one knows and is in a position to then know that they know, it also means that if they were an ideal rational agent, they would always, given that they know that *p*, know that they

[5] Williamson's way of putting this in Knowledge and its Limits (2000, p. 95) is the following: "To be in a position to know p, it is neither necessary to know p nor sufficient to be physically and psychologically capable of knowing p. No obstacle must block one's path to knowing p. If one is in a position to know p, and one has done what one is in a position to do to decide whether p is true, then one does know p. The fact is open to one's view, unhidden, even if one does not yet see it."

know that *p*. One could also see the KK-thesis as an ideal of epistemic responsibility (see for instance Cresto, 2012, p. 927), that is, as something separate from a rational ideal. Epistemic responsibility, in relation to the KK-thesis, would simply mean that one is obliged to accept one's first order states by way of a second order state. A similar idea is expressed by Levi (1997, p. 41) when he writes that:

> *[W]e are committed at that time t to have full beliefs that are logically consistent and to fully believe all the logical consequences of what we fully believe at that time. We no doubt fail to fulfill this commitment and, indeed, cannot do so in forming our doxastic dispositions and in manifesting them linguistically and in our other behavior. We are, nonetheless, committed to doing so in the sense that we are obliged to fulfill the commitment insofar as we are able to do so when the demand arises and, in addition, have an obligation to improve our capacities by training, therapy or the use of prosthetic devices provided that the opportunity is available and the costs are not prohibitive.*

And in his (1980, p. 10), Levi writes:

> *If X is committed to a standard for serious possibility, he should live up to that commitment to the extent that he is able. Such ability depends, in part, on the extent to which X is aware of that commitment. […] A normative account of the improvement of knowledge should prescribe no more than persons and institutions are capable of implementing. Thus, it would be foolish to require that rational X identify all the logical consequences of the assumptions he explicitly makes. We can, at most, expect him to identify those consequences insofar as he is able.*

If the KK-thesis is logically valid given some assumptions, we can then say with Levi that, as rational agents, then when are in a state of knowing we are also committed (as in rationally obligated) to knowing that we know that we are in a state of knowing. For Cresto, this is a commitment limited epistemic responsibility in the sense of "embracing" or "making sure you own" your own beliefs. For Levi (1997) this would be a principle motivated by the more general principle of having doxastic commitments to the logical consequences of one's full set of beliefs at any given time. In Levi (1980) one is committed to the logical consequences of one's standard for serious possibility based on one's capabilities. The principle from Levi (1980) could serve to explain why the KK-thesis possibly holds for human beings and not for animals, since the former are capable of metacognition, whereas the latter are not and thus cannot be expected to know that they know, given that they know (even if the KK-thesis is logically valid).

Now we can return to the point of contention for Williamson. For to him, there are no mental states such that a subject has this obstacle-free access to them. As he says, there are no cognitive homes. The stance is that not only are we often prevented from knowing that we know that *p* even if we know that *p*, but, given knowledge that *p*, we may not even *be in a position* to know that we know that *p*.

Knowing now what it means to be in a position to know, we can understand Williamson's (2000, p. 95) definition of luminosity:

> A condition C is defined to be luminous if and only if (L) holds:
>
> (L) For every case a, if in a C obtains, then in a one is in a position to know that C obtains.

We can also see clearer that luminosity is a far less demanding than self-intimation. In no way does luminosity necessitate that a subject is aware of the luminous condition obtaining if it in fact does (which is what self-intimation does). (L) merely necessitates that luminous conditions are 'bright enough' to be visible (knowable) if one directs their attention towards them. Despite being less demanding both by adding the qualification 'being in a position to' and using luminosity instead of self-intimation, Williamson gives two arguments against this weaker version of the KK-thesis.

2.1 The Anti-luminosity of a Paradigmatic Mental State

The first argument pertains to the condition of feeling cold (see ibid., section 4.3). While 'feeling cold' is not a case of knowing, it is a condition we would normally regard as luminous (if not even self-intimating). If we feel cold, we surely are in a position to be conscious of the fact that we feel cold. If this paradigmatic mental state were to not be luminous, presumably there are slim chances for states of knowing to be luminous.

Williamson asks us to consider a day that feels freezing at dawn but slowly warms up and by noon it starts to feel hot. There is then a corresponding change in the second-order states about one's mental state (Williamson assumes these changes occur simultaneously with the corresponding feelings of being hot or cold). One starts the day at dawn in a position to know that one feels cold, but ends it not being a position to know that one feels cold (since one now feels hot). Williamson then asks us to suppose that the feelings of hot and cold change so gradually that it is not possible to notice a difference over just one millisecond. This leads to a gradual decrease in the confidence of what it is one feels (so one goes from being sure that one feels cold, to hesitating, to then finally being sure that one no longer feels cold). The argument proper now runs as follows:

Assume a time series, $t_0, t_1, \ldots, t_n$ with one millisecond intervals going from dawn (t_0), when it is cold, until noon when it is warm (t_n). Then let α_i be a case at t_i ($0 \leq i \leq n$) where one knows that one feels cold. If one knows that one feels cold at α_i, then, Williamson claims, we have to accept that one feels cold at α_{i+1}, otherwise the knowledge at α_i would be far too unreliable (let us call this rule 1_i, following Williamson). If we now assume that feeling cold is luminous, we can construct the reductio argument:

(1) In α_i one feels cold.

(2) Given luminosity one also knows that one feels cold in α_i.

(3) If one knows that one feels cold in α_i, one also feels cold in α_{i+1} (given 1_i)

The reader may see where this is going. If one feels cold at α_{i+1}, then this state is equally luminous, and so one knows that one feels cold at α_{i+1}, which given 1_i means that one also feels cold at α_{i+1+1} which again leads to knowledge that one feels cold at α_{i+1+1}, and so on it goes until one reaches α_n and still know that one feel cold. The reductio is then complete, since at α_n it is noon, and at noon one feels warm. We can then either deny that 1_i is true, or that luminosity is true. Williamson decides to reject luminosity. This argument seems to work mainly based on the assumption that if we know that *p*, *p* needs to be such that it is not going to change from millisecond to millisecond. That is, it seems to assume that only truths not subject to unreliable and rapid changes are apprehensible by a subject.

The second argument Williamson presents against luminosity uses a paradigmatic case of knowing as an example instead of the condition of feeling cold. This is an important addition because while it may be true that feeling cold is not a luminous condition, we can still conceivably view knowledge as a condition that is not subject to similar kinds of subtle gradual changes. Williamson clearly believes that knowledge can be acquired and lost through gradual processes if the knowledge is of a mental state subject to gradual and sometimes unnoticeable changes, but it is not clear just how subtle one can allow the state one has knowledge of to be in order to say one genuinely has knowledge of it. For even if we accept that knowledge can be acquired gradually, we do not have to accept that such a process possesses the same vagueness or subtlety associated with temperature changes and the corresponding bodily sensations they engender. Many mental states are not perfectly luminous, even mental states that share certain similarities with knowledge, but it is not yet clear why knowledge could not be luminous. Feeling cold is also not a state of knowing in Williamson's view, since feeling cold is not a factive mental state (one can feel cold without being cold, as when one has a fever). As we shall see below (Section 3.3.1), introspection is likely not factive, which means that the knowledge

that one feels cold is not possible to begin with in Williamson's view. In any case, his second argument meets the doubt that feeling cold is not an appropriate example to show the implausibility of luminosity for knowledge. The vagueness in this argument relates not to the object of knowledge, but instead to the vagueness inherent to some factive propositional attitudes (that is, vagueness inherent to some states of knowing).

2.2 Knowledge and its Margins for Error

Williamson's second argument against the KK-thesis relies on knowledge of an estimated measurement of some object in the distance. Imagine a scenario involving a man called Mr Magoo. He looks out of his window and spots a tree in the distance and wonders how tall it is. He does not know the precise height of the tree, but there is a particular type of knowledge he does gain. In reality the tree is 666 inches tall, and while Mr Magoo cannot know this, he can know that the tree is not 60 inches or 6000 inches tall. However, since he cannot know the precise height of the tree, there is always a natural number *i* such that Mr Magoo does not know that the tree is not *i* inches tall. That is, Mr Magoo knows that for any estimate *i*, that the tree could really be *i*+1 or *i*–1. Similarly, if the tree is *i+1* inches, he does not know that it is not *i* inches tall. The scenario, according to Williamson, gives rise to the following rule that Mr Magoo follows once he reflects on his own situation:

> (1_i) Mr Magoo knows that if the tree is *i*+1 inches tall, then he does not know that the tree is not *i* inches tall.

Williamson then asks us to accept a closure principle for Mr Magoo (not by way of a general principle, but because Mr Magoo has 'attained reflective equilibrium over the propositions at issue', ibid., p. 116):

> (C) If *p* and all members of the set *X* are pertinent propositions, *p* is a logical consequence of *X*, and Mr Magoo knows each member of *X*, then he knows *p*.

For the reductio, again assume that the KK-thesis is true and consider the following two sentences:

> (2_i) Mr Magoo knows that the tree is not *i* inches tall.
>
> (3_i) Mr Magoo knows that he knows that the tree is not *i* inches tall.

We get (2_i) from Mr Magoo knowing that the tree is, for instance, not 0 inches tall. From (2_i) and the KK-thesis we get (3_i). Now let *q* be the proposition that the tree Mr Magoo looks at is

i+1 inches tall. Assuming *q*, then from (1_i), ¬ (2_i) is derived and Mr Magoo can derive the implication $q \rightarrow \neg (2_i)$. ¬ *q* is then inferred from $q \rightarrow \neg (2_i)$ and (2_i) by modus tollens. Since Mr Magoo is aware of these logical consequences by (C), and he knows that (2_i) through the KK-thesis, he also knows ¬ *q*, and we get*:*

(2_{i+1}) Mr Magoo knows that the tree is not *i*+1 inches tall.

We can see that the argument can be repeated until we eventually reach the fact that Mr Magoo knows that the tree is not 665+1 inches tall which leads to a contradiction because knowledge is factive and the tree is actually 666 inches tall. The decision then is whether to keep epistemic closure for Mr Magoo or the KK-thesis. Williamson opts for the former. Since we do not need to accept a general closure principle in order to accept Williamson's argument, it would be necessary to give some argument for why Mr Magoo could not be aware of these logical consequences. In principle the argument that arrives at (2_{i+1}) seems perfectly accessible to him (with some effort, perhaps, which, as stipulated, Mr Magoo has put into understanding his own situation).

The argument takes advantage of the fact that perceptual knowledge oftentimes brings with it information about various more precise measurements. Hearing, he claims (ibid., 119), brings with it information about loudness in decibels, felt temperature brings with it information degrees in Celsius, vision brings with it certain estimates in length or volume of perceived objects. Fundamentally, it seems that these are different kinds of knowing involved here. On the one hand, there is perceptual knowledge that gives us an idea of the presence of a tree. On the other, from the veridical perception of a tree, we can safely assume that the tree has some extension (that it is measurable). Through this Mr Magoo can say that he knows that the tree is not 0 inches tall. But this is not arrived at through perception, it is arrived at through reason (all real objects are extended in space, so if this tree I see is real, it cannot be 0 inches tall). From the perception of the tree we can make a very broad claim about the measurement of the tree, but since no measuring device is used, it seems that this is arrived at through reason. It is therefore unclear if the distinction between perceptual and non-perceptual knowledge can invalidate Williamson's argument. The following section will attempt to discuss this in some more detail by examining two articles aiming to exploit the distinction between knowledge acquired by perception and knowledge acquired by reason.

2.3 Defending the KK-thesis – The difference between Knowledge and Knowledge

A curious fact about two articles (Sharon & Specrtre, 2008; Dokic & Égré, 2009) that set out to defend the KK-thesis from Williamson's argument is that they both reformulate some of the crucial premises in his argument. In both cases this leads to a reformulation of the entire argument. Reformulations complicate matters somewhat since it becomes harder to see if the counterargument against Williamson still holds when it does not discuss either the principles or the premises Williamson posits directly. This section will nonetheless examine two of these attempts to see whether the reformulations remain applicable to Williamson's original argument (as it was presented in the previous section). Sharon & Spectre (2008) and Dokic & Égré (2009) both argue that Williamson fails to distinguish between two types of knowledge in his argument and that once this distinction is heeded, the reductio fails and the KK-thesis is salvaged.

The fundamental argument against (1_i) takes more or less the same form in both articles. Dokic & Égré (2009, p. 4) is slightly easier to read notation-wise, so let us start with this. They use a different margin of error principle than Williamson assigns to Mr Magoo, namely:

(KME) $K\,(K\neg p_i \rightarrow \neg\, p_{i+1})$

Where p_i stands for the proposition that the tree is i inches tall. Compare this with the formal version of (1_i):

(1_i) $K\,(p_{i+1} \rightarrow \neg\, K\neg\, p_i)$

We can see that (KME) is derived from contraposing the imbedded conditional of (1_i). The new formulation is called Williamson's Proposition (WP) by Sharon & Spectre (2009, p. 290). They also claim it is a more intuitive formulation of the principle, as well as being logically equivalent to (1_i) (which it is, given (C)). Reconstructing the full margin for error argument with (KME) instead of (1_i) results in the following argument that assumes the KK-thesis, (KME), and (C) to be true (taken from Dokic & Égré, 2009, p. 4):

(1) $K\neg\, p_i$, hypothesis
(2) $K\,(K\neg\, p_i \rightarrow \neg\, p_{i+1})$, by (KME)
(3) $K\,K\neg\, p_i$, by (1) and (KK)
(4) $K\neg\, p_i, K\neg\, p_i \rightarrow \neg\, p_{i+1} \vdash \neg\, p_{i+1}$, by propositional reasoning
(5) $K\neg\, p_{i+1}$, by (2), (3), (4) and (C)

So, the same conclusion is reached with (KME) as with (1_i). But let us mainly discuss (1_i), since it is the most familiar. Dokic & Égré discuss the different types of knowledge that can be identified in (1_i). They quickly conclude that the type of knowledge one has of the conditional

cannot be perceptual. It does not seem possible prima facie to perceive a conditional, so this seems perfectly acceptable. If we assume instead that the knowledge is non-perceptual, we get the following formulation of (1_i) (where "Kρ" denotes non-perceptual knowledge*)*:

$$K_\rho\ (p_{i+1} \rightarrow \neg\ K_\rho \neg\ p_i)$$

This, however, leads to a very strange scenario in the case of $i = 0$. In this case we would have to say that one knows logically that if a tree is *0+1* inches, then one does not know that the tree is not 0 inches. But non-perceptual knowledge is not by necessity subject to any margin of error principle (the same goes for Mr Magoo's perceptually based margin for error principle, however; it is not this way by necessity, but simply contingently the case for Mr Magoo). Knowledge we arrive at through reason does not seem to need the same margin for error. While vagueness pertaining to vision is very common, it is not as clearly true for reasoning. Thus, the knowledge operators cannot both be said to be either perceptual or non-perceptual. The resulting compromise for Dokic & Égré is the following ("K_π" denotes perceptual knowledge, K non-perceptual knowledge):

(KME′) $K\ (K_\pi \neg\ p_i \rightarrow \neg p_{i+1})$

This yields the same conclusion of $K\neg\ p_{i+1}$, but it can no longer be iterated. Take a look at the new argument:

(1) $K_\pi \neg\ p_i$, hypothesis

(2) $K\ (K_\pi \neg\ p_i \rightarrow \neg\ p_{i+1})$, by (KME′)

(3) $K\ K_\pi \neg\ p_i$, by (1) and (KK′)

(4) $K_\pi \neg\ p_i, K_\pi \neg\ p_i \rightarrow \neg\ p_{i+1} \vdash \neg\ p_{i+1}$, by propositional reasoning

(5) $K \neg\ p_{i+1}$, by (2), (3), (4) and (C)

Note now that the knowledge we have of $K \neg\ p_{i+1}$ is not perceptual. This means that the argument can no longer be repeated because we have lost the initial premise for every possible iteration of the argument (that is, $K_\pi \neg\ p_{i+1}$ cannot be derived*)*. Without perceptual knowledge, (4) cannot be derived. Note that this is not Williamson's original argument, however. With (1_i) modified along the lines suggested by Dokic & Égré, assume again (KK′), and (C). Let us see what this results in:

(1) $K_\pi \neg\ p_i$, hypothesis

(2) $K\ (p_{i+1} \rightarrow \neg\ K_\pi \neg\ p_i)$, by ($1_i'$)

(3) $K\ K_\pi \neg\ p_i$, by (KK′)

(4) $\neg\ p_{i+1}$, by (1), (2), and modus tollens

(5) K ¬ p_{i+1}, by (C), (1_i) and (3)

With the distinction between perceptual and non-perceptual knowledge in place, the original argument cannot be iterated. This hinges on the assumption that reflective knowledge does not inherit the need for a margin for error. Dokic & Égré effectively argue that a margin of error is relative to the methods for obtaining knowledge, or relative to the kind of knowledge one is talking about. This kind of maneuver by Dokic & Égré seems acceptable. For even if we accept Williamson's idea that knowledge is some kind of genus under which various species of knowledge are included, such as reflective and perceptual knowledge, we do not have to accept that they all share similar characteristics and follow similar principles. When it comes to the logico-syntactic properties of knowledge, species of knowledge may be homogeneously grouped as instances of knowing, but we have little reason to believe that this means there are no further distinctions between instances of knowing relating to the kind of principles Williamson posits for Mr Magoo's discriminatory capabilities. The KK-thesis may then be salvaged; the main culprit seems to be Mr Magoo's failure to distinguish between two different species of knowledge and not realizing that they do not follow the same kinds of principles.

This is only one way to approach the relation of the KK-thesis to one kind of externalist knowledge. To use it in a reductio along with several other premises will always leave open the possibility to simply deny one of the other premises. One could even bring into question Williamson's homogenization of different types factive mental states under the heading 'knowledge' as a result of the Mr Magoo argument, before one denies the KK-thesis. Williamson says very little about the plausibility of the KK-thesis in relation to the idea that knowledge is a mental state, so he is of little help when it comes to spelling out the reasons for which the KK-thesis may be viable even for an externalist. So, another approach to the problem is to follow Hilpinen in seeing if it is possible that externalist theories of justification support or make plausible the notion that knowledge entails knowing that one knows. Specifically, let us see whether the conjuncts that make up knowledge that *p* (or if knowledge is taken as a broad mental state) entail the conjunction of knowing that one knows that *p*. This requires exploring some ways in which the KK-thesis can be said to be compatible with externalist theories of knowledge, which is what the next section will attempt to do.

3. The Plausibility of Positive Introspection for Externalist Theories of Knowledge

So far we have seen that the adoption of the KK-thesis does not lead to any immediate contradictions relating to various safety principles of perceptual knowledge. But why should we think that the KK-thesis is compatible with externalism? In broad terms, the strategy is to show that self-knowledge of one's state of knowing comes with knowledge. That is to say, given that we are human beings with certain self-reflective capabilities and that we *normally* understand the situations we are in as well as our place in them (as well as propositions relating to these situations) we also generally have an ability to form higher order states about lower order states. This section will look at a defense of the KK-thesis with these caveats in mind, namely, Conor McHugh's (2010) article *Self-knowledge and the KK principle*. He defends the following version of the KK-thesis:

> (KK) For any subject S and proposition *p*, if S knows *p*, and S grasps the proposition that she knows *p*, and the normal conditions for psychological self-knowledge are in place, then S is in a position to know that she knows *p*; and this is true for principled reasons having to do with the nature of knowledge.

This does not seem to be too different from Williamson's formulation of the KK-thesis. It again seems to relativize the thesis to non-ideal rational agents in normal psychological conditions. McHugh claims that, given the nature of knowledge and related epistemic phenomena, the KK-thesis holds a priori. What is the nature of knowledge, then, that makes it so that the KK-thesis is valid even with an externalist conception of knowledge? Let us first go through McHugh's account of the KK-thesis and see in which way it can be said to hold for an externalist conception of knowledge.

McHugh uses the term 'warrant' to stand for those concepts that, when undefeated and added to a true belief, amount to a state of knowledge. Within this concept he does not want to differentiate between externalist and internalist conceptions, and he includes reliabilism within the term 'warrant', so for the purposes here let us read 'has warrant' as 'the belief that *p* has been formed by a reliable process'.

The first fact about knowledge is that if a true belief has been arrived at through a reliable process, this means nothing more has to be done in order for the true belief to be considered reliably acquired. That is, no more evidence asked of the knower when they claim that they know something, as long as their true belief is reliably acquired (that is, as long as they have warrant). This, I take it, corresponds to the implication E$p \rightarrow$ EEp, put forth by Hilpinen. If the true belief that *p* is acquired through a reliable process, then it is also the case that this process is reliably reliable (or else it would not be reliable to begin with). Using similar notation as

Hilpinen, the implication would be something like RB*p* → RRB*p* (if a belief that *p* is obtained through a reliable process, then the reliable belief forming process is reliable). This comes from the idea that warrant has to be undefeated in order to avoid various Gettier examples. If undefeated warrant is indeed defeated, this is by overriding the warrant by a more forceful type of warrant (or in this case, a reliable process can only be overridden by an even more reliable process).

The second fact about knowledge according to McHugh is that one comes to believe that *p* under the right cognitive conditions (that is, under normal conditions) by way of the reliable process. This mainly means that the cognitive performance of the knower is not hindered by things such as distractions, lacking alertness, being under the influence of some kind of substance, the knower is not hallucinating as opposed to perceiving, and so on.

Continuing along this path, self-knowledge about one's beliefs also necessitate some kind of normal conditions. Again, the normal conditions for self-knowledge of one's mental states are met when there is nothing obscuring the belief from the introspective capabilities of the holder of the belief. This is the general idea captured by B*p* → BB*p*. But it also captures the idea that B*p* → RBB*p*. That is, if one is in normal cognitive conditions, one can reliably form a belief about one's own beliefs. Alternatively, we could say that these implications hold if one is in a position to inquire into one's beliefs since, under normal cognitive conditions, adult humans are usually in a position to ask themselves whether they hold a certain belief. On this point Gareth Evans (1982, p. 225) makes an equivalence between inquiring into whether one has the belief that *p* and merely inquiring into *p* when he writes: "I get myself in a position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*". There seems to be a parallel between the quote from Evans and the idea that considering whether one believes that they believe that *p* is tantamount to considering whether one has the belief that *p*. The idea of this parallel is developed significantly in Robert Gordon's (2007) idea of Ascent Routines (AR). He writes (ibid., p. 154): "whether in answer to a question or not, people optionally step up a semantic level from an assertion that *p* to a self-ascription of a belief that *p*. We allow ourselves to move from what common parlance and some of the philosophical literature calls expressing the belief that *p* to self-ascribing the belief that *p*". Ascent routines are generally taken to be procedures take that us from statements about the world to statements about self-ascribed states. These ascent routines can be seen as the underlying procedures that make the following principle valid (given that *p* → *q* is accessible): if *p* → *q* is valid, then B*p* → B*q* is valid. This is then effectively applied to a belief, as well, so that one can easily go through the procedure of going from a

statement about the state of one's mind to a self-ascription about the state of one's mind (Bp → BBp). Gordon also notes (ibid., p. 161) that self-ascription of this kind requires a conceptual competence regarding the propositional attitudes in question. So, it is possible to say "I believe that it is raining" while being linguistically competent in the sense that one can use such a sentence in the correct circumstances, but still not be conceptually competent in the sense that one knows that it means to believe that it is raining. If, however, one explicitly self-ascribes a propositional attitude, this implies a conceptual competence that very plausible could be applied to one's beliefs about one's belief that, for example, it is raining outside. Given that McHugh uses the caveat 'you grasp the proposition that you know p', it seems that this kind of ascent routine is valid and can be used to argue that KK is valid.

I will now first present McHugh's (2007, p. 241) argument for the KK-thesis and then test whether it can genuinely be said to hold for a properly externalist conception of justification. McHugh's argument runs as follows:

(1) You know *p*, you grasp the proposition that you know *p*, and the conditions for self-knowledge are met. (Assumption)

(2) You have a warranted belief in *p*. (1)

(3) You have an undefeated warrant to believe that you believe *p*. (1)

(4) You have an undefeated warrant to assume, of any warranted belief of yours, that it is warranted. (Assumption)

(5) You have an undefeated warrant to believe that you have a warranted belief in *p*. (2, 3, 4)

(6) You have an undefeated warrant to believe that you know *p*. (2, 5, closure)

(7) You are in a position to know that you know *p*. (1, 6)

(8) If you know *p*, you grasp the proposition that you know *p*, and the conditions for self-knowledge are met, then you are in a position to know that you know *p*, and the conditions for self-knowledge are met, then you are in a position to know that you know *p*. (1, 7)

So from *p* and one's warranted belief in *p*, the undefeated warranted belief that one has undefeated warrant to believe *p* is entailed. Can this argument be repeated for a reliabilist theory of knowledge? Let us also assume self-knowledge under normal conditions holds and that one grasps the proposition that one knows that *p*. Adopting Hilpinen's notation and assuming that knowledge is a true belief obtained through a reliable process, we get: Kp ↔ p ∧ Bp ∧ RBp and KKp ↔ KBp ∧ KRBp ∧ Kp ↔ BBp ∧ BRBp ∧ RRBp ∧ Bp ∧ RBRBp ∧ RBp ∧ p (where RBp means that Bp was acquired through a reliable process). The implications we arrive at through

McHugh's discussion of the nature of knowledge and normal conditions for self-knowledge are the following:

(a) $Bp \rightarrow BBp$

(b) $RBp \rightarrow RRBp$

The KK-thesis given reliabilism amounts to the following implication:

(KK) $p \nearrow Bp \wedge RBp \rightarrow BBp \wedge BRBp \wedge RRBp \wedge Bp \wedge RBRBp \wedge RBp \wedge p$.

Removing the conjuncts that are trivially true yields:

(KK) $p \nearrow Bp \wedge RBp \rightarrow BBp \wedge BRBp \wedge RRBp \wedge RBRBp$

Now we can see that from (a) and (b) and K we merely get BB*p* and RRB*p*. There are two conjuncts missing in order for KK to be true, which is the belief that one has obtained the belief reliably (BRB*p*) and that one has a reliably obtained belief about the reliability of one's belief that *p* (RBRB*p).* RBRB*p* seemingly implies BRB*p*, since if RBRB*p* is true, it is true because the belief in the reliability of the belief that *p* has itself been acquired through a reliable process. The question whether having a belief formed by a reliable process entails a reliable belief that one had gone through such a belief forming process is now forced upon us. Given normal conditions for self-knowledge, does having a RB*p* entail RBRB*p*?

The scope of self-knowledge first needs to be discussed. For nothing in reliabilism necessitates that one believes that one's true belief is acquired through a reliable process. Knowledge can be obtained, in this sense, without it being intimated to the knowing subject. The reliability of the belief-forming process does not have to be within the 'view' of our introspective capabilities. But does this mean we are not in a position to see what kind of process has led to our belief? That is, are we (human beings) ever able to obtain knowledge without an inkling as to how this belief arose? We do not come to believe that it is raining outside, for instance, from a belief that emerges from the void (presumably). We come to believe a proposition only through some process; a process that has to be in place in order for us to believe it in the first place. So, given normal conditions for self-knowledge, is a belief such, or is knowledge such, that we are automatically in a position to know what kind of events and perhaps intentional states that took place and thus engendered the given propositional attitude? If we borrow Williamson's (2000, p. 34) idea that some paradigmatic ways of knowing are seeing that…, hearing that…, remembering that… etc., we can ask: are these states not also directly accessible to us in a way that allows us to, at the very least, know that we are having a particular kind of experience? When one says "Yes, I believe it is raining" it is with background

knowledge of the evidentiality of the belief one ascribes to oneself. We are not only aware of a given proposition, but also the way in which we are aware of it. Alternatively put, we are simultaneously in a position to know what a given intentional object is, but also the intentional act. Or as Franz Brentano writes (1995 [1875], p. 68)

> *Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon includes something as object within itself, although they do not all do so in the same way. In presentation something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired and so on.*

This was further developed by Husserl (2002, §1 and §12; 2004, §87), with his distinction between an intentional act's quality and material (and later, in Ideas, between noesis and noema). One of the senses in which Husserl speaks of consciousness is as the term for all psychical or intentional experiences, these are characterized both by the intentional act (the noetic) and the intentional object (the noema). The self-knowledge one has under normal conditions, then, if Brentano and Husserl are right, should give one an idea of both the intentional act and object which underlie one's epistemic access to some given proposition. I believe this gives us RBp → BRBp. This is because if we are in a state of knowing, given that belief does not arise without intimation of the events that led up to the belief, we seem to be in a position to at least believe that we have arrived at a belief by way of a reliable process. This is only the case, however, if we associate some of our own intentional acts as reliably providing us with true beliefs (so that seeing that p is not something that often leads to contradiction, or to clashing beliefs with other people, and so on). If an intentional act and a correlated material, for instance 'sees that p', is what brings about a belief, and we associate this kind of intentional act with reliably providing us with true beliefs (in at the very least an implicit way, in the sense that we would be prone to sticking to believing what we see in the vast majority of case), it also seems that having a belief that p formed by a reliable process entails that we also believe that our belief was formed by a reliable process. RBp → BRBp seems to be valid with these ideas in mind.

Furthermore, if knowledge is the term for different kinds of factive (broad) mental states, like Williamson suggests, then we have even more reason to believe that RBp → BRBp holds. The reliable process is then, in most cases, simply the way one came to believe something and

there is no reason to think we would not believe that this process has occurred if it indeed did and normal conditions for self-knowledge were in place. That a belief that one has gone through a reliable process of obtaining another (first order) belief is also supported by the fact that human beings generally learn, either implicitly or explicitly, which kinds of processes are reliable. So that, if one propositionally sees that *p*, one generally also believes that one sees that *p* because one has, at least implicitly (and explicitly if one is under normal conditions of self-knowledge), a belief that seeing that *p* is a reliable process.

A similar point is argued by Daniel Greco (2014) for an informational theory of knowledge. One of the most well-known proponents of such a view is Fred Dretske, who in Knowledge and the Flow of Information (1981) gives an analysis of knowledge based on information states:

> K knows that *s* is F = K's belief that *s* in F is caused (or causally sustained) by the information that *s* is F.

Greco, inspired by Dretske's and David Lewis' contextualist theory of knowledge (the latter he believes has the same 'abstract structure' as Dretske's theory), claims that states carrying information about various facts also carry information about the information-carrying state itself, thus he writes (2014, p. 185):

> *If my stomach is growling, then my stomach is in a state (the growling state) that carries the information that it is in a state (the very same growling state) that carries the information that I am hungry. This is because whenever my stomach is growling, I am in a state - the growling stomach state - that carries the information that I am hungry. Higher-order information comes for free with first-order information*.

This can be illustrated formally with the conditions involved in S carrying information that *p*:

(1) S is in a state X such that $\Box_N$ (S is in X → *p)*, and
(2) N.

That is if S is in X, then given normal conditions ($\Box_N$ = all worlds where conditions are normal) then S's state of X is factive. KK on this theory involves being in an information-carrying state which carries information about (1) and (2). (KK) can then be expressed as follows:

(1') S is in a state X' such that $\Box_N$ (S is in X' → (S is in state X such that $\Box_N$ (S is in X → *p*) ∧ N)) and
(2') N.

The argument for KK is now rather simple. (2) entails (2') trivially since they are the same condition. Greco then assumes that X = X' is true (the state of one's growling stomach carrying information about one's hunger is the same as the state carrying information about one's growling stomach). The last N is also ignored in (1') by Greco, because, as he claims (ibid., p. 185), "any claim of the form $\Box_P (Q \rightarrow R \wedge P)$ is equivalent to $\Box_P (Q \rightarrow R)$". Now (1) becomes equivalent to (1'). Writing it out may help illustrate the tautological nature of the statement: If S is in a state X that is such that under normal conditions if S is in X it implies that S is in a state X such that under normal conditions if S is in X it entails that *p*. 'S is in state X' tautologically entails that 'S is in state X' and the rest of (1') is derived from (1).

From an information-based causal theory of knowledge, the KK-thesis can be derived given assumptions about normal conditions. Greco does not stipulate that normal conditions involve any kind of normal conditions for self-knowledge. He merely states that if one is in normal conditions where one sees a dog, for instance, the visual system is working as it should, while the object of perception (the dog) is precisely what it seems to be. But if 'normal conditions' has a wide enough scope that includes the cognitive being as well as her surroundings, then normal conditions for self-knowledge may conceivably be included in the broader notion of 'normal conditions'.

What is lacking on this analysis, however, is that there is no requirement that a subject be aware of the fact that they are in normal conditions in the higher order state. This possibly results from Greco's version of David Lewis' (1996) contextualist theory of knowledge being too much of an abstraction of the original. Lewis writes (1996, p. 561) "I say S knows that P *iff* P holds in every possibility left uneliminated by S' s evidence - Psst! - except for those possibilities that we are properly ignoring". Greco (2014, p. 183) reduces this to "S knows that P *iff* S is in an experiential state X such that in all normal (that is, not properly ignored) possibilities, if S is in X then P". But with this move we have lost what sustains the relation between the experiential state and P *(*cf. Dretske's definition of knowledge which includes 'causally sustains' or 'is caused by'), and we have no independent reasons to think that the experiential state includes information about the normal conditions. In fact, we do not even have reasons to believe that X is related to *P* other than by mere coincidence, or that X is related to P through the perception of some other state Q that entails P. For example, if one is looking at a tree under normal conditions, this entails that there is a tree, and the tree has underground roots. One can be in a state X of visually perceiving the tree, and when one does the proposition P (that the tree has roots) will also be true, which by Greco's definition would amount to knowing that the tree has underground roots. But one cannot see the underground roots and

does not really know how the tree is kept upright from the position one is in of merely visually perceiving the tree (at least, one does not know that one knows this).

As noted above, an experiential state is always an experience of something. So, it seems that in order to maintain the relation between X and *p*, some reformulations would be necessary. Namely, X*p*, as opposed to just X. That is, if X*p* and *p*, are both true under normal conditions, one can be said to know that *p* (this is more or less Williamson's idea that factive propositional attitudes are ways of knowing). If we insist that the experiential state has to be about what is instantiated in the world, however, Greco's argument for KK given an informational theory of knowledge does not work. Knowing that you know would then be written as: S is in a state $X_{p \wedge N}$ such that $\Box_N$ (S is in $X_{p \wedge N} \rightarrow$ (S is in state X_p such that $\Box_N$ (S is in $X_p \rightarrow p) \wedge N$)). No experiential state in and of itself includes information about one's general conditions (neither surrounding conditions, for instance that one is in fake barn country, nor information about the health or accuracy of one's perceptual system).

It seems then that the question remains whether RBRB*p* can be derived from RB*p* (or BRB*p)* and what exactly needs to be in place in order to make such an entailment plausible. Reliabilism does not require an idea of normal conditions in this sense. But one can perhaps assume that if one's true belief is acquired through a reliable process, then conditions are normal. If things were not normal, then there would be no room for reliable processes to flourish. Still, to figure out whether the last entailment holds, we have to ask: Is BRB*p* formed by a reliable process? That is, is our belief in the reliability of the process of acquiring a belief reliably acquired? If so, is this process instantiated whenever RB*p* or BRB*p* is instantiated? It is unclear if this is too different from simply asking if the reliable process itself is reliable. But perhaps there is a distinction to be made here. The difference between RRB*p* and RBRB*p* would be that RRB*p* only requires that whatever reliable process one has gone through in order to arrive at a true belief is itself reliable, which seems to follow trivially. RBRB*p*, on the other hand, requires, for example, that the belief that one's visual perception leading to one's belief that *p* is arrived at through a reliable process. So first of all, what is the process one goes through in order to arrive at a true belief that one has a reliably acquired true belief? It seems that the only process there is that can allow one to gain access to the character of one's belief forming processes is introspection. If introspection reliably can tell us what kind of process has formed one of our beliefs, it seems that RB*p* along with normal conditions for self-knowledge would entail RBRB*p.* A brief excursion into the reliability of introspection is due.

3.1 Is introspection a Reliable Belief-forming Process?

Let us first clear up what kind of processes can be deemed reliable. Goldman (1981, p. 92) seems to consider the same type of processes brought up here as reliable when he writes: "Justifiedness centrally rests on the use of suitable psychological processes; and by psychological processes I here mean basic, elementary processes, not acquired techniques that are mentally encoded and applied." The native psychological processes of cognizers "mark the province of primary epistemology" (ibid., p. 93). What makes these processes reliable is the fact that they, in the vast majority of instances, bring about true beliefs (they are truth-conducive processes). Again, quoting Goldman (ibid., p. 26): "An object (a process, method, system, or what have you) is reliable if and only if (1) it is a sort of thing that tends to produce beliefs, and (2) the proportion of beliefs among the beliefs it produces meets some threshold, or criterion, value. Reliability, then, consists in a tendency to produce a high truth ratio of beliefs". This is similar to definition of reliability by Peels (2016., p. 2464). Writing specifically about the reliability of introspection he writes: "I take belief formation on the basis of introspection to be reliable just in case it delivers true belief in most cases, that is, if the process yields true beliefs dependably enough, including across a range of nearby counterfactual cases". In Goldman (2006, p. 253) the claim is made that introspection is just like any other perceptual process, in the sense that transduction still occurs, that is, there is a neural input based on various first order states whose contents are redeployed in an introspective state. Combining this view with Goldman's earlier view that basic psychological processes are generally reliable, it seems that introspection inherits similar levels of reliability that can be found in first order perceptual systems.

While some authors (Schwitzgebel, 2008; Peels, 2016; Engelbert & Carruthers, 2010; Gopnik, 1993) argue that introspection is not reliable based on various failures to produce correct self-ascriptions corresponding to particular stimuli in experimental settings, it is far from generalizable to the obviousness with which one can for instance be aware of the fact that one is having a visual experience of something if one is looking at a tree. One might not be able to estimate the height of a tree, or doubt that a change in light strength has occurred if one is looking at some kind of light source, or one may fail to self-report on a visual stimulus relating to minor hue changes in colors or fast flashes of scenes and so on, but one rarely fails to be aware of the distinction between having a visual, as opposed to olfactory, experience. That is, one only needs to reliably be able to identify what kind of mental state one is in as one attempts to apprehend information about one's environment. Whether one does this by way of ascent routines, by being aware of one's own behavior, or through some kind of direct perception of one's mental state by way of introspection, it nonetheless seems that, given that one is in a first

order epistemic state and normal conditions for self-knowledge are satisfied, one has reliable access to what kind of perceptual state one is in. That is to say, under normal condition, when one has acquired a true belief through a reliable process, one has psychological access to the kind of psychological process it is that has brought on the true belief. We can perhaps say, then, that RBp → RBRBp is valid given normal conditions for self-knowledge.

3.2 Reliabilism and the KK-thesis

Is this all a reliabilist theory of justification requires in order to know that one knows? Is mere awareness (through introspection) that a psychological process has occurred, a process one generally regards as reliable, enough to validate the KK-thesis? Since this awareness is what constitutes introspection, and introspection likely is a reliable process, then given reliabilism, it seems to be enough. Given normal conditions for one's introspective ability, (KK) seems to hold even for a reliabilist variety of knowledge. That is to say, (KK) $p \nearrow$ Bp ∧ RBp → BBp ∧ BRBp ∧ RRBp ∧ Bp ∧ RBRBp ∧ RBp ∧ p is valid. Let us briefly summarize the particular entailments hold that in in order to show the validity of (KK) given normal conditions for self-knoweldge. I have now attempted to give argument in the support for the following entailments (given normal conditions for self-knowledge):

(a) Bp → BBp
(b) RBp → RRBp
(c) RBp → BRBp
(d) RBp → RBRBp

Assuming the correctness of these entailments, the full argument can now be given:

1.	$p \nearrow$ Bp ∧ RBp	(hypothesis)
2.	p	(∧E)
3.	Bp	(∧E)
4.	RBp	(∧E)
5.	RBp	(∧E)
6.	BBp	(3, a)
7.	RRBp	(4, b).
8.	BRBp	(4, c)
9.	RBRBp	(4, d)
10.	BBp ∧ BRBp ∧ RRBp ∧ Bp ∧ RBRBp ∧ RBp ∧ p	(2-9, ∧I)
11.	$p \nearrow$ Bp ∧ RBp → BBp ∧ BRBp ∧ RRBp ∧ Bp ∧ RBRBp ∧ RBp ∧ p	(1, 10, →I)

Now there is the question whether RRB*p* and RBRB*p* should be differentiated, and if they should, if both should factor into the conjunction of knowing that one knows. RRB*p* seems to mirror Hilpinen's EE*p*, that is, that the reliable process is itself reliable is about the same as saying as the evident is also evident. But since the notation chosen here is motivated by the intention to represent that reliability as coupled with the belief that *p*, it would likely be appropriate to remove the RRB*p* as a conjunct altogether. It has no genuine impact on the argument, however, so this can be left up to the reader's own preferences. If one can accept (a)-(d), then under normal conditions for self-knowledge, the KK-thesis seems to hold even for reliabilism. Of course, with 'under normal conditions for self-knowledge' we have significantly altered the original conception of the KK-thesis. On the other hand, it seems that this is an assumption that has to be in place when we are speaking of ideal rational agents. The conclusion for reliabilism is, then, that if one is a cognizer capable of metacognition and if one is prone to using this ability, then one is also in a position to know that one knows, if one is in a state of knowing.

3.3 The KK-thesis for other kinds of externalism

3.3.1 Knowledge as a Factive Mental state

A parallel argument can be constructed with Williamson's theory of knowledge. We thus ask, does having a first order factive stative propositional attitude imply that one is in a position to have a second order factive stative propositional attitude about the first attitude? This poses a slightly different challenge, because now introspection no longer has to be reliable, but has to be factive. This means that if one has a factive propositional attitude that *p*, then *p* is invariably true. This is certainly not true for introspection in general, as Ryle and the anti-introspection authors quoted above (Section 3.1) so aptly point out. But is introspection factive given normal conditions for self-knowledge? Here it becomes worthwhile to discuss the difference between external and internal perception. Seeing, for instance, is a broad mental state in Williamson's view. Seeing does not only depend on what is internal to a cognizer, but what is external to them. Can introspection be regarded as introspection only if what one introspects is actually there? For instance, can one introspect that one believes there is a cat on the mat without it being true that one believes there is a cat on the mat? Contrast this with seeing that there is a cat on the mat, we would not say that someone sees that there is a cat on the mat unless there really is a cat on the mat. But would it be strange to accept that one introspects a belief about a cat being on the mat without the cognizer having such a belief? I think it would be quite odd to

claim that one introspects if one is not aware of what it is one introspects. It seems perfectly possible to argue for the fact that, like perception, introspection is a state one is in only if there is something to introspect. Like with regular perception, one may be mistaken in some ways about the contents of one's mind, but it is rare that one completely misses the mark given normal conditions for self-knowledge. Furthermore, if one is mistaken about what one sees (by categorizing the object of perception incorrectly, perhaps), this just means that one did not perceptually what one thought one apprehended and it has no bearing on the factivity of the actual mental state of seeing what is actually there. These kinds of mistakes do not threaten factivity either since they, given broadness, ensure the falsehood of the perceptual state as a whole, not just what the perceptual state takes as its proposition (or its corresponding state of affairs).[6] Each instance of seeing is a way of seeing something in particular, if one is wrong about what one sees, then one cannot be said to really see it. Can the same be said for introspection? And are there ways in which the classification of a first order mental state by introspection can go awry, thus showing that introspection is not factive? Prima facie one could think that the category mistakes that sometimes affect objects of perception, and sometimes the kind of perception itself, is a mistake stemming from the non-factivity of introspection. But let us expand on this slightly.

First, on an eliminative materialist conception of the mind (or on Ryle's conception of the mind), it would seem that introspection could be said not to be factive. One could go so far as to say that introspection does not reveal anything about the state of our brains or about what it means to actually have a belief (in terms of what such a state would correspond to in the brain). Given such a view, however, the whole endeavor of arguing for the possibility of metacognition about first order mental states falls apart since one can then never know anything about the content of one's mind from the first-person perspective. The KK-thesis quite trivially fails on this view (even with normal conditions for self-knowledge, normal conditions in general, merely demanding that an agent be in a position to know…., and so on). If we assume that self-knowledge does say something about the state of one's mind (even indirectly, so that

[6] This is one of David Armstrong's worries in his (1963, p. 422) where he writes: "Introspective apprehension or awareness, like all other apprehension, is an apprehension that the thing apprehended is of a certain sort. The apprehension involves classifying the experience, in however rudimentary a way; that is, it involves the application of concepts. Now, surely, the notions of classifying and misclassifying are co-ordinate notions; surely the one can apply only when it is meaningful to apply the other? We can apply a certain concept to our experience only if it is possible to withhold that concept. Yet, according to the doctrine of incorrigibility, the application of any concept except the concept we do apply is logically impossible." It is logically possible that one can fail to apply the right concept to one's experiences, as Armstrong points out here, but in Williamson's view this would lead to the total negation of a sentence of the type "S sees that p*"*, not just *p* (the latter would lead to a counterexample to the factivity of seeing).

a state of mind corresponds to a state of the brain), then it is hard to see how introspection can fail to be factive while other mental states such as seeing that or hearing that do not. If these mental states are all part of the basic psychological processes of human beings, as Goldman was quoted saying above, then perhaps they also share the fact that they are factive, along the lines of Williamson's theory of knowledge.

Sosa's (2012, p. 175) principle that introspective 'seemings' are justified only when they are concurrent with a mental state that shares the same propositional content as some lower-level mental state holds some relevance here:

> *Any propositional content that appears introspectively to you (that thus seems to you correct) is thereby epistemically justified provided its propositional content corresponds to a mental state that you undergo at that time and that attracts your assent to that content (where this mental state is either the fact constituted by the truth of that propositional content or else is the truthmaker for that fact).*

Then, on the difference between introspective justification and perceptual justification, Sosa writes (ibid., p. 177): "What distinguishes introspective justification is that the truth of the content of the introspective seeming (or the truthmaker for that truth) exerts its attraction unaided". Finally, on what makes introspection factive (ibid., p. 178):

> *An introspective seeming on our view is a kind of direct (noninferential) attraction to assent to <p>, a seeming that p, where <p> appropriately concerns a mental state hosted by the subject. Such a seeming can get it right. If it does so in a way that manifests the subject's introspective competence, then it is an apt introspective seeming, which makes it factive.*

These quotes by Sosa seem to aim at the fact that states of introspection, when introspection carried out by a competent subject who only is attracted to assent to true content, is indeed factive in the sense that the introspection itself is what attracts one to assent to the content of one's mental state. If we assume that the competent subject is equivalent to the subject that has obtained normal conditions for self-knowledge, it seems that Sosa can also be said to subscribe to the factivity of introspection. If introspection is the apt attraction to assent to some propositional content of one's mind, then it seems that this direct attraction to assent to one's mental content can only arise if the content truly is there.

A problem arises, however, when we think about the scope of the factivity of introspection. If the mental content one introspects has to merely be there for a kind of introspective seeming (as in, 'it seems that I am looking at some coffee'), then it is hard to see how one can be mistaken about the general characteristics of such an experience. If the factivity

of introspection necessitates the truth of a proposition of a lower-level mental state like 'sees that *p*' (as it seems to do if introspection is factive in the Williamsonian sense), the KK-thesis is not as easily argued for with Williamson's theory of knowledge. This is because the broadness of the factive propositional attitude is not accessible through introspection unless something more is said about the nature of having a factive propositional attitude. If one has a factive propositional attitude, then one is in a position to introspectively apprehend the internal conditions of such a state, but in no way does one know from introspection whether one's perception is veridical.

We could still say that, whenever one has a factive propositional attitude, one is in a position to form a higher order propositional attitude about one's first order state in a way that can never be false unless the first-order propositional attitude is not factive. To an omniscient logician, then (one that can always distinguish between seeing and hallucinating), the KK-entailment would hold given Williamson's theory of knowledge. For human beings, on the other hand, it is not always possible to differentiate between having a factive mental state and a similar-but-not-factive mental state. An attempt to argue against this can be found in *Phenomenology of Perception* (2012, pp. 308-311&349-358) where Merleau-Ponty argues that the distinction between reality and appearance can only be an internally motivated distinction. That is, he claims we are able to verify the veridicality of our perception through perception itself, and likewise we can only spot an illusion by way of veridical perception. Moreover, he describes the ability of patients suffering from schizophrenia to distinguish between their hallucinations and their veridical perception. Through veridical perception, he goes on, we are capable of 'crossing out' (ibid., p. 311) our past illusions or hallucinations. On his view, genuine perception always risks being illusory, but likewise illusion is effectively overridden by increasing the carefulness with which one observes a phenomenon, that is, by increasing one's focus on the phenomenon, by spending more time observing the phenomenon, or whatever else that may aid veridical perception or prevent illusion. This, however, I see as at most speaking for the reliability of perception. Given enough time and focus, we are often able to distinguish between veridical perception and non-veridical illusions or perceptual mistakes, but such a tendency does not quite permit us to posit the factivity of introspection about factive propositional attitudes. As long as we are able to mistake genuine perception for illusion for some period of time, or vice versa, we cannot say that introspection is factive with a scope that ranges over the first order mental state. Specifically, we can introspect that we see that *p*, without it being true that we see that *p* (for instance when it is the case that we hallucinate that *p*). Considering then that knowing that one knows, in Williamson's theory of knowledge, would

be instantiated when a subject has a factive propositional attitude that they are having a factive propositional attitude, it seems that KK, after all, does not hold given his variety of externalism.

3.3.2 Causal theories of Knowledge and the KK-thesis

Recall the definition used to illustrate the causal theory of knowledge from Goldman (1967):

> *S* knows that *p* if and only if the fact that *p* is causally connected in an "appropriate" way with **S's** believing that *p*.

For simplicity's sake (again not subscripting the subject S), let us formalize 'S knows that *p'* given this view as K ↔ B*p* ∧ *p* ∧ CB*p* and the KK-thesis as (ignoring the conjuncts that hold trivially) B*p* ∧ *p* ∧ CB*p* → BB*p* ∧ CBB*p* ∧ BCB*p* ∧ CBCB*p*. From B*p* we get BB*p*, as have been established already. Given the nature of knowledge according to McHugh, causal processes carry not only information about the fact that *p* but information about the causal process itself, we seem to both get BCB*p* and CBCB*p* if we also have normal conditions for self-knowledge. That is, if we see our perceptual system as an appropriate causal connection between the world and our belief in various aspects about the world, which seems appropriate, then when we have visual experience, for instance, we will also believe that the appropriate causal connection (visual perception of something state of affairs) to the world has been formed (BCB*p*). Is BCB*p* causally connected to CB*p* in an appropriate way? If causal processes do divulge information about the kind of causal process is going on, then it seems that the corresponding belief also comes about in a causally appropriate way since one is simply forming a belief about the causal process based on the information provided by the causal process itself. CBB*p* similarly is implied by B*p* in that it entails BB*p* which is caused by some form of introspection or ascent routine. If one asks oneself if they believe that *p*, as one does if conditions for self-knowledge are satisfied, then BB*p* is caused by being faced (introspectively) with the fact that one believes that *p* (B*p*). So, the case could be made that even for causal theories of knowledge, given that causal processes are accessible to epistemic subjects and that the subject obtain suitable conditions for self-knowledge, the KK-thesis holds true.

4. Concluding remarks

The reader may have noticed that we ended up quite far from the original KK-thesis posited by Hintikka. What is interesting, however, is that only a few assumptions were needed in order to argue for the KK-thesis coupled with externalist theories of justification. Once we remove the requirement that a subject be ideally rational and simply require that the subject be in a position

to know that they know when they are in a state of knowing, we see that knowledge is a salient-enough state so that when one finds oneself in a state of knowing, one oftentimes is in a higher order mental state as well. Since human beings are metacognizing, self-aware, self-monitoring, etc., they are usually in a position to know what kind of experiences they have and what has formed them (at least to some extent). Oftentimes this is enough to be in a position to know that one has formed a belief by way of a (generally) reliable process, or through an appropriate causal process. It may not be enough, however, to put one in a position to distinguish between whether one has a factive mental state in every single case. It seems that reliabilism and causal theories of justification aptly allow for the fallibility of higher order epistemic states. This is crucial precisely due to the reasons for which Williamson's theory of knowledge fails to entail the KK-thesis (even given the added provisos). But of course, Williamson gets his sought-after result, despite the potential failure of his own anti-KK arguments.

Let us briefly summarize the results and discussions of this paper. The paper begun by identifying the difference between Cartesian (internalist) and anti-Cartesian (externalist) epistemology. The main difference between them seemed to be the level with which respective proponents of these theories demand that states of knowing are self-intimating and reflectively grounded. For the project to be at all plausible, first Williamson's objections against the KK-thesis had to be discussed. The second argument, taking the KK-thesis proper as a premise, was found to conflate two types of knowing. It was concluded that reflective knowledge and perceptual knowledge cannot be said to operate under the same types of principles, which prevented the iteration of Williamson's argument (which was necessary in order to derive an absurd conclusion that then permitted the rejection of the KK-thesis). The plausibility of the KK-thesis then had to be discussed on its own merits, requiring a discussion of introspection and how it differs from other types of basic psychological processes. It was found that introspection is generally reliable, even approximating the property of being factive (given a competent introspective subject) but since there is a chance to mistake illusion for genuine perception, introspection is not a genuinely factive state of mind. This fact is what then invalidates the KK-thesis on Williamson's account of knowledge.

What is interesting, I believe, is that internalists have to make similar compromises in order for the KK-thesis to hold for non-ideal subjects. That is, if one accepts that mental states are not perfectly self-intimating, one also has to add normal conditions for self-knowledge and remove the requirement that the subject know everything they are in a position to know. This would necessitate vastly more argumentation in order to be fully convincing, but perhaps through these results we can begin to see that externalism allows for the KK-thesis to hold for

human beings in normal conditions in similar ways to internalism. At the same time, externalism allows for more types of cognizers to obtain knowledge without thereby losing the value of reflective knowledge emphasizes by internalism. If internalism and externalism both validate the KK-thesis given similar assumptions about the nature of knowledge and conditions on the normality of the epistemic situation and the subject's introspective capabilities, what could then motivate the preference of a theory of justification that excludes most cognitive creatures?

One could argue that given similar assumptions, some externalist theories can amount to the same type of higher-level knowledge that internalism aims to define, given that it is a human being that obtains it. Internalism holds an implicit assumption that it is human knowledge one is speaking of (since requiring any reflection, again, excludes the vast majority of cognizers) anyway, so again these does not seem to be any difference between choosing an internalist theory of knowledge and adding another assumption to externalism. Perhaps results such as these can convince the internalist that externalism for human beings is not so different from internalist theories of knowledge. Our metacognition, in most cases, provide us with reliable information about our own states. But we also find that we lose the infallibility of knowledge, or the complete factivity of knowing and knowing that one knows, and in a sense we lose the hope for absolute certainty. It is unclear if the epistemologist can arrive at these lofty goals, but if that is one's goal, then externalist theories that, given few assumptions, validate the KK-thesis may not be preferable.

The internalist may still find some interest in the possibility to derive second order knowledge from the fact that one knows given an externalist theory of justification while using similar assumptions the internalist is forced to use (only implicitly) in order to arrive at first-order knowledge. Even if knowledge is ascribable to animals, human beings can maintain a certain privileged epistemic position in that they, with their metacognition, are able to reflect and gain knowledge of their own states, and thus *understand* their relation to the world in a way that goes beyond animal knowledge simply by using reliable processes or being in causal relations that are unique (for the most part) to humans.

References

Armstrong, D. M. (1963). Is introspective knowledge incorrigible? *The Philosophical Review*, *72*(4), 417–432.

Armstrong (2000). The thermometer-model of knowledge. In Bernecker, S., & Dretske, F. I. (2000). Knowledge: Readings in contemporary epistemology.

Brentano, F. (1995). Psychology from an empirical standpoint. Routledge.

Chisholm, R. (1957). *Perceiving*: *A philosophical study*. Cornell University Press.

Chisholm, R. (1981). The first person: An essay on reference and intentionality. The Harvester Press.

Chisholm, R. M. (1989). Theory of knowledge (3rd Edition). Prentice-Hall.

Cresto, E. (2012). A defense of temperate epistemic transparency. *Journal of philosophical logic*, *41*(6), 923–955.

Descartes, R. (2006). A Discourse on the Method. Oxford University Press.

Descartes, R. (2008). Meditations on first philosophy: With selections from the objections and replies. Oxford University Press.

Dokic, J., & Égré, P. (2009). Margin for error and the transparency of knowledge. *Synthese*, *166*(1), 1–20.

Dretske F. (1981). Knowledge and the Flow of Information. MIT Press.

Engelbert, M., & Carruthers, P. (2010). Introspection. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(2), 245-253.

Evans, G. (1982). The varieties of reference. Clarendon Press.

French, C. (2012). Does propositional seeing entail propositional knowledge? *Theoria*, *78*(2), 115–127.

Goldman, A. I. (1967). A causal theory of knowing. *The journal of Philosophy*, *64*(12), 357–372.

Goldman, A. I. (1979). What is justified belief? In *Justification and knowledge*, 1-23. Springer, Dordrecht.

Goldman, A. I. (1986). Epistemology and cognition. Harvard University Press.

Goldman A. (2006). Simulating Minds: the Philosophy, Psychology, and Neuroscience of Mindreading. New York, NY: Oxford University Press

Gopnik, A. (1993). 15 How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality. *Readings in philosophy and cognitive science*, 315.

Gordon, R. M. (2007). Ascent routines for propositional attitudes. *Synthese*, 159(2), 151–165.

Greco, D. (2014). Could kk be ok?. *The Journal of Philosophy*, *111*(4), 169–197.

Grice, H. P., & White, A. R. (1961). Symposium: The causal theory of perception. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *35*, 121–168.

Hilpinen, R. (1970). Knowing that one knows and the classical definition of knowledge. *Synthese*, 109–132.

Hintikka, J. (1962). Knowledge and belief: an introduction to the logic of the two notions (Repr. ed.). King's College Publications.

Hintikka, J. (1970). 'Knowing that one knows' Reviewed. *Synthese*, 141–162.

Husserl, E., & Cavallin, J. (2002). Logiska undersökningar. Thales.

Husserl, E. (2004). Idéer till en ren fenomenologi och fenomenologisk filosofi. Thales.

Levi, I. (1980). The enterprise of knowledge: an essay on knowledge, credal probability, and chance. MIT Press.

Levi, I. (1997). The covenant of reason: rationality and the commitments of thought. Cambridge University Press.

Lewis, D. (1996). Elusive knowledge. *Australasian journal of Philosophy*, *74*(4), 549–567.

McHugh, C. (2010). Self-knowledge and the KK principle. *Synthese*, *173*(3), 231–257.

Merleau-Ponty, M. (2012). Phenomenology of perception. Routledge.

Peels, R. (2016). The empirical case against introspection. Philosophical Studies, 173(9), 2461–2485.

Ramsey, F.P. (1931). *Knowledge*, in R.B. Braithwaite (ed.), Foundations of Mathematics and Other Logical Essays, Kegan Paul.

Ryle, G. (2009). The concept of mind. Routledge.

Sharon, A., & Spectre, L. (2008). Mr. Magoo's mistake. *Philosophical Studies*, *139*(2), 289–306.

Sosa, E. (2001). Human knowledge, animal and reflective. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *106*(3), 193-196.

Sosa, E. (2012). In Smithies, D., & Stoljar, D. (Eds.). (2012). Introspection and consciousness. Oxford University Press, 169–182.

Stalnaker, R. (2010). Our knowledge of the internal world. Oxford University Press.

Stalnaker, R. C. (2019). *Luminosity and the KK Thesis*. In Knowledge and Conditionals (31–48). Oxford University Press.

Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review*, *117*(2), 245–273.

Williamson, T. (2000). Knowledge and Its Limits, Oxford University Press.