

Reflective meta-attitudes and (in)compatibilism¹

Abstract This paper contains a defense of the orthodox view that knowledge entails belief. I begin with distinguishing two ways in which one can deny the knowledge-belief entailment claim, then I present intuitive arguments against both kinds of denials. I close the paper with a discussion over some possible objections against the presented arguments.

Keywords Knowledge, Belief, Higher Order Attitudes, Compatibilism, Incompatibilism

As Jakko Hintikka described the matter in his seminal work: “It is indeed reasonable to assume that whatever one *actively* knows one also *actively* believes” (Hintikka, 1962, p. 50). Call this a traditional picture of the knowledge-belief relationship. According to this traditional picture, knowledge entails belief. After Gettier (1963), it became part of common philosophical knowledge that the idea of knowledge that comprises this traditional view must be wrong. Most philosophers find the source of its weakness in the notion of justification. Replacing this notion with a concept that appeals to contextually relevant standards (DeRose, 1992) or counterfactual dependency of belief on circumstances of evaluation (Nozick, 1981) are among the most popular strategies for dealing with the weaknesses. Other strategies attempt to supplement the idea of knowledge with the fourth condition that is not met in Russell-Gettier cases (cf. the “no false grounds” view of Feit and Cullison, 2011). There remains, however, a smaller group of philosophers who look for a remedy elsewhere—they assent to the very possibility of knowledge without belief² and, therefore, question the traditional picture of the knowledge-belief relationship. Below I will present considerations that support the claim that they must be wrong, at least if among cognitive agents there are subjects capable of entertaining reflective meta-beliefs and reflective meta-knowledge states. Such an argument may be important since, as some critics have complained, the dominant strategy of arguments for the correctness of the traditional picture is a 'wait-for-counter-examples' approach (Myers-Schulz & Schwitzgebel, 2013).

The structure of the paper is as follows. In the first section I shall contrast two ways in which one can deny the knowledge-belief entailment claim. In the second and the third sections, I shall present intuitive arguments against both kinds of denial. In the fourth and final section I shall address some possible objections against the arguments from sections two and three.

¹ The work in this paper was funded by a National Science Center, Poland grant, under award number 2013/11/B/HS1/03947. The author is grateful to Piotr Stalmasczyk for his encouragement and patience. I also would like to thank the anonymous referee for suggestions.

² Yet an even smaller group of philosophers denies the factivity of knowledge. I do not discuss this issue in this paper.

I. Compatibilism and Incompatibilism

Let us borrow a standard notation from epistemic and doxastic logic (I shall say more about the commitment to such logics in the final section of the paper). Let us, therefore, abbreviate 'x knows that p ' as ' $K_x p$ ' and 'x believes that p ' as ' $B_x p$.' In this notation, the entailment thesis may be formulated in the following manner:

$$(KB) \quad \Box \forall x (K_x p \Rightarrow B_x p)$$

which (assuming the converse Barcan formula) immediately gives the following claim about actuality:

$$(AKB) \quad \forall x \Box (K_x p \Rightarrow B_x p)$$

The denial of (KB) is:

$$(NKB) \quad \Diamond \exists x (K_x p \wedge \neg B_x p)$$

which, obviously, does not have to be exemplified by any actual cognitive agent (merely a possible one, suffices). As such, (NKB) does not say much about the source of the possibility of knowledge without belief but, as equivalent to (if the accessibility relation is serial):

$$(NKB^*) \quad [\Diamond \exists x (K_x p \wedge \neg B_x p) \wedge \Diamond \exists x (K_x p \wedge B_x p)] \vee \neg \Diamond \exists x (K_x p \wedge B_x p)$$

it embraces two distinct claims about the relationship between knowledge and belief. The first one, which I will refer to as *compatibilism*, states that:

$$(COM) \quad \Diamond \exists x (K_x p \wedge \neg B_x p) \wedge \Diamond \exists x (K_x p \wedge B_x p)$$

while the other, which I will refer to as *incompatibilism*, is:

$$(INCOM) \quad \neg \Diamond \exists x (K_x p \wedge B_x p)$$

Compatibilism is the claim that the co-occurrence of knowledge and belief is a purely contingent

matter, while incompatibilism states that co-occurrence is impossible.

Roughly speaking, incompatibilism is a radical thesis and it is difficult to see how one can find any reasons for believing it. It goes without saying that we may utter a sentence p and, for instance, truly say afterwards: “I believe and know that to be true.” If one takes incompatibilism to be true, one has to deny the intuitive truthfulness of such assertions. This clearly requires an argument as well as a theory that is going to explain why we should treat them as perfectly correct and possibly true. Furthermore, one may note that if incompatibilism is true, then—since in such a case, belief that p would entail ignorance about p and knowledge would entail the lack of belief—we have to face the fact that particular instances of conjunctions ' $B_x p \wedge \neg K_x p$ ' and ' $K_x p \wedge \neg B_x p$ ' are redundant as is generally true of conjunctions of the form ' $p \wedge q$ ' in cases where p entails q or vice versa³. And, that it is difficult to recognize the redundancy or triviality of such utterances.

It may be worth noting, however, that there exists an interesting connection between incompatibilism and skepticism about the external world. Bertrand Russell noted once that between the content of the mental state and its objective (i.e. the fact that makes it true or false):

“(...) there is sometimes a very wide gulf, for example in the case of 'Caesar crossed the Rubicon.' This gulf may, when it is first perceived, give us a feeling that we cannot really 'know' anything about the outer world. All we can 'know,' it may be said, is what is now in our thoughts. If Caesar and the Rubicon cannot be bodily in our thoughts, it might seem as though we must remain cut off from knowledge of them.” (Russell, 1992, p. 234)

As I understand him, Russell interprets external world skepticism (he is not taking this position, of course) as defending a version of restricted incompatibilism: beliefs about the external world are not (and cannot be) knowledge states, and (equivalently) knowledge states about the external world cannot be belief states. This is usually interpreted as the claim that there are no knowledge states (about the external world) at all, but Russell's description of the theory fits a different interpretation very well: knowledge states are not a species of belief states⁴.

I must admit that I know of no author who firmly defends incompatibilism (external worlds skeptics aside). The only person who defended a view *relatively* close to it is Raziel Abelson (Abelson, 1968), who claimed that *in cases of first-person sentences* it is pragmatically inappropriate to utter a belief sentence and the corresponding knowledge sentence, as the correctness of the former pragmatically entails incorrectness of the latter. In his analysis Abelson

³ I would like to thank Tomasz Puczyłowski for stressing that point in the discussion.

⁴ Does this interpretation extend to radical skepticism? Since Russell explicitly relies on the distinction between knowledge of our mental states and knowledge of external facts, the reasons he describes seem inapplicable to skepticism, appealing to Cartesian-Demon-like scenarios (excluding states that are constitutive for *cogito* if one takes them to be directly knowable). However, the radical skeptic clearly defends incompatibilism too, as the radical skeptic thinks that the Russellian gulf extends to every belief.

was not making use of the notion of implicature, but one may interpret his suggestion as the claim that in normal contexts a belief that the speaker does not know that p is a scalar implicature of the utterance 'I believe that p .' Since this implicature can be canceled in particular contexts, utterances like 'I believe and know that to be true' are perfectly correct and possibly true in such contexts. Interpreted in such a way, Abelson's theory can be reconciled with both compatibilism and the traditional picture of the knowledge-belief relationship.

Compatibilism, on the other hand, has been defended by several prominent thinkers. One can mention here, for instance, Radford (Radford, 1966), Black (Black, 1971), Mannison (Mannison, 1976), and Lewis (Lewis, 1996). Some recent results in experimental philosophy (cf. Myers-Schulz & Schwitzgebel, 2013) may also be interpreted as supporting compatibilism. One of the arguments for compatibilism postulates a strict symmetry between cases of intuitively correct utterances of sentences like 'A believes and knows that to be true' and intuitively correct utterances of sentences like 'A knows this, but s/he doesn't believe it.' Consider, for instance, the following example from Myers-Schulz and Schwitzgebel (which is a variation of the example discussed originally by Radford⁵):

Kate has spent many hours studying for her history exam. She is now in class taking the exam. Everything is going quite well, until she comes to the final question. It reads, "In what year did Queen Elizabeth die?" Kate has reviewed this date many times. She even recited the date to a friend just a few hours earlier. So, when Kate sees that this is the last question she feels relieved. She confidently looks down at the blank space, waiting to recollect the answer. But before she can remember it, the teacher interrupts and announces, "All right, the class session is almost over. You have one more minute to finalize your answers." Kate's demeanor suddenly changes. She glances up at the clock, now flustered and worried. "Oh, no. I can't perform well under this kind of pressure." Her grip tightens around her pencil. She strains to recall the answer, but nothing comes to her. She quickly loses confidence. "I suppose I'll just have to guess the answer," she says to herself. With a sigh of disappointment, she decides to write "1603" in the blank space. This is, in fact, the correct answer.

The scenario has two versions: one with the summarizing question "Did Kate know that Queen Elizabeth died in 1603?" and the other with the question "Did Kate believe that Queen Elizabeth died in 1603?" The answers showed a statistically significant difference in positive replies to both questions. This might be taken as evidence of a case of knowledge without belief. If this is correct,

⁵ A similar scenario is mentioned by Lewis: "I even allow knowledge without belief, as in the case of the timid student who knows the answer but has no confidence that he has it right, and so does not believe what he knows." (Lewis, 1996, p. 556).

the argument goes, since we take cases of 'A believes and knows that to be true' as instances of correct and possibly true utterances, exactly the same has to be said of utterances like 'Kate knows that Queen Elizabeth died in 1603, but she doesn't believe that Queen Elizabeth died in 1603.' Hence proponents of the standard view have to explain why cases of the latter kind are not in fact similar to cases of the former kind. I shall go back to this challenge in section three below.

II. An intuitive argument against incompatibilism

In spite of the fact that incompatibilism is a radically unattractive view that has unwelcome consequences, it might be useful to have a direct argument against it. Here is one that seems convincing. Assume (INCOM) and the factivity of knowledge. This gives us (for a particular agent a):

$$\square [K_a p \Rightarrow (p \wedge \neg B_a p)]$$

which, if p is any sentence stating that ' $B_a q$,' gives:

$$\square [K_a B_a q \Rightarrow (B_a q \wedge \neg B_a B_a q)]$$

Now suppose, uncontroversially, that:

$$\diamond [K_a B_a q \wedge (B_a q \Leftrightarrow B_a B_a q)]$$

i.e. that there is a possible scenario in which a knows that s/he believes something and that (in that scenario) s/he believes that s/he believes what s/he believes. Now it follows immediately that:

$$\diamond (B_a B_a q \wedge \neg B_a B_a q)$$

i.e. that in this scenario s/he is simultaneously holding and not holding a particular meta-belief. Allowing such an impossibility constitutes, in my opinion, a decisive argument against incompatibilism.

III. An intuitive argument for the traditional picture

In order to convincingly argue for (KB), the following condition has to be met: in the argument,

neither the (KB) nor the corresponding inference rule ("K_xA entails B_xA") might be used. However, this does not generally apply to every *instance* of (KB) if we have *independent* reasons to think that a particular instance is justified. Consider, for instance, the following statement:

(MKB) $\Box \forall x (K_x K_x p \Rightarrow K_x B_x p)$

Suppose that you have reflective meta-knowledge about your knowledge that *p*. This requires a certain repertoire of epistemic constraint regarding your attitude towards *p*—*p* has to be true as well, as well as having to meet other conditions that knowledge has to meet. The same applies to reflective meta-knowledge that requires a certain repertoire of epistemic constraint regarding your attitude towards *knowledge that p*—it has to be the case that you know that *p* as well, as well as having to be the case that other constraints on knowledge hold. As noted above, the nature of such additional conditions remains controversial. However, one may argue that no matter what the conditions are, they are roughly met in the case of knowledge about our own mental states and, in particular, in the case of knowledge about our knowledge. If you think you know that you know something, then, if it is, in fact, true that you know this, then your reflective grounds for thinking that ensure that the other constraints imposed on knowledge are met. And, if they are met (e.g. if you are justified in thinking that you know that you know that *p*, or if this thought tracks the truth, or if no false grounds for thinking that are involved), then you clearly cannot know that you disbelieve that *p*. So, one may have independent reasons for believing (MKB) even if one does not accept (KB).

The simple reasoning I am going to present below applies to agents that are capable of entertaining reflective meta-belief and reflective meta-knowledge⁶ for every belief or knowledge state they have. Thus, a version of the traditional picture we are going to argue for looks like this:

(KB*) $\Box \forall x [x \text{ is capable of having reflective meta-attitudes} \Rightarrow (K_x p \Rightarrow B_x p)]$

In fact, it is sufficient even to assume that they are capable of entertaining reflective meta-belief and reflective meta-knowledge for every belief or knowledge state *that is not a meta-belief or meta-knowledge state*. This may be important as one may doubt if allowing meta-beliefs about meta-beliefs, meta-knowledge about meta-knowledge, meta-beliefs about meta-knowledge, or meta-knowledge about meta-beliefs, makes sense⁷. Below I shall argue for (KB*) directly, but the only assumption the argument requires is that the agents have a disposition to entertain attitudes that are

⁶ Occasionally I shall speak of meta-attitudes here.

⁷ I am not claiming that it does not. The point is simply that we do not have to assume that in order for the argument to work. More about this below—see: *Problem 2* in section four.

not themselves attitudes about attitudes.

So here is the first simple, albeit controversial, argument. Assume that one knows that p :

[1] $K_a p$

Now, one may assume that for reflective agents the following holds as well⁸:

[2] $\Box \forall x (K_x p \Rightarrow K_x K_x p)$

This is, of course, the (in)famous KK principle and we have more to say about its relevance and justification (see below). [2] and [1] immediately give:

[3] $K_a K_a p$

and this (assuming (MKB)) gives:

[4] $K_a B_a p$

which immediately entails (assuming the factivity of knowledge):

[5] $B_a p$

Since all the additional assumptions used in the argument (that is: (2), (MKB), and the factivity of knowledge) are necessarily true, it follows that (KB*) is true. The argument is simple, intuitive, and presupposes neither (KB) nor the corresponding inference rule. As such, it appears a good candidate for a reply to those philosophers who complain about the 'wait-for-counterexamples' strategy of arguing for (KB). Its strength depends, of course, on the legitimacy of the assumptions used.

The last of the assumptions is factivity (in our version: ' $\Box \forall x (K_x p \Rightarrow p)$ '). I take this assumption to be uncontroversial (the best argument in its favor might be the nonexistence of convincing counterexamples). Another assumption is (2), or the (in)famous KK principle. This assumption is controversial. It has been defended, for instance, by Hintikka (1962, 1970) and criticized by Williamson (2000) (see Hemp (ONLINE) or a useful overview of the controversy—Hemp suggests that *both* Hintikka and Williamson may be correct as they are describing distinct notions of knowledge). I do not aim to resolve the controversy here. On the

⁸ This is the only principle regarding the reflective meta-attitudes that one has to assume in this version of the argument.

contrary, I believe that the controversy over the KK principle clearly indicates the need for an alternative argument⁹.

I believe that the following principle of “introspection” may help us here:

$$[6] \quad \Box \forall x (B_x p \Leftrightarrow B_x B_x p)$$

The principle seems to apply to reflective agents that are capable of having reflective meta-beliefs. (It is, for instance, a modalized version of one of the axioms of the logic LB (the logic of conscious beliefs), cf. Tokarz, 1990.) Note that the fact (noticed by numerous philosophers) that one may be mistaken if one is in the belief-that- p state is not a counterexample to [6]: even if I am wrong that I am brave enough to do a parachute jump (I have never been in the situation in which my disposition or indisposition to jump has been tested), this proves at most that *I do not know* if I really believe something. Another statement that should be true of reflective agents is as follows:

$$[7] \quad \Box \forall x (K_x p \Rightarrow B_x [(K_x p \wedge B_x p) \vee (\neg K_x p \wedge B_x p)])$$

[7] states that knowledge implies the belief that either one knows while believing or one believes without knowing. This is a simple formulation of view that knowledge states and belief states are subjectively indistinguishable. I think that [7] is a specific case of a more general principle stating that if two attitudes of the agent differ at most in external (i.e. subject-independent) credibility, then the subject has to believe that s/he is at least in the state of having fewer strict and demanding credibility standards.

Now [7] (by propositional logic and distribution of belief over conjunction) gives:

$$[8] \quad \Box \forall x (K_x p \Rightarrow B_x B_x p)$$

and this (by [6]) directly entails (KB*). Just as in the previous case, the argument assumes neither (KB) nor the corresponding inference rule. The only required additional assumptions are [6] and [7] which, I think, can be independently justified in the case of reflective and conscious agents.

IV. Problems

The argument cannot be left here, however. Below I shall concentrate on four possible objections that one could raise against the considerations presented.

⁹ The minimal conclusion that can be derived is that accepting the KK principle should give one good reason for believing in the traditional picture.

Problem 1. The first problem we have to face directly is the interpretation of apparent “knowledge without belief” scenarios while keeping the contrast between them and both “knowledge with belief” and “belief without knowledge” ones.

Reply. The first class of cases of apparent “knowledge without belief” scenarios encompasses situations in which one is disposed to assent to 'I do not only believe that: I know it' (or assent to a sentence that, in a given context, can be treated as equivalent to such locution). However, this hardly produces a counterexample to (KB), as uttering 'I do not only believe that: I know it' seems to be just a way of saying that one is not in a state of believing without knowing but rather in a state of *believing and knowing*. Moreover, this fits the implicature analysis sketched above well: by assenting to 'I do not only believe that: I know it' one simply cancels the scalar implicature of regular belief sentences¹⁰ (one may note immediately that such cases are instances of “knowledge with belief” situations). It goes without saying that it is hard to see how the implicature strategy might be applicable to the “belief without knowledge” scenario.

The second class of cases of apparent “knowledge without belief” embraces, for instance, the Radford example in the version presented by Myers-Schulz and Schwitzgebel. The scenario is interesting partially because (in contrast with the examples of the first kind) the implicature strategy seems to be inapplicable to it. As such, I think, it indicates that it is incorrect to claim that:

“It makes sense to say, 'I do not believe that; I know it', not because it is logically inconsistent to say that a woman believes what she knows but rather because this is an emphatic way of saying, 'I do not only believe that: I know it'” (Lehrer, 1990, p. 27).

The reason is that not every utterance of 'I do not believe that; I know it' can be taken as tantamount to 'I do not only believe that: I know it.' In the Radford-Myers-Schulz-Schwitzgebel scenario we clearly do not want to say that Kate does not only believe that Queen Elizabeth died in 1603. That being said, the interpretation of this (and other) empirical considerations used against the entailment thesis is very controversial.

The conclusions of Myers-Schulz and Schwitzgebel have been questioned by other authors who claimed that experimental considerations in fact support the view that knowledge entails *dispositional belief* (Rose & Schaffer, 2013) or that knowledge entails the so-called *thin belief*¹¹ (Buckwalter & Rose, 2015). Both views are taken by the authors to be the intended understanding

¹⁰ This corresponds well to what certain other authors think of such situations (cf. Lehrer, 1990).

¹¹ Roughly speaking, *thin belief* that *p* is the belief that requires nothing more than regarding *p* as true. The other category of *thick belief* requires additionally that emotional (and motivational) element is present. Buckwalter and Rose think that folk psychology distinguishes clearly two notions of belief that correspond to this distinction.

of the entailment thesis. I generally agree with the criticism presented in the two aforementioned papers. Notwithstanding this agreement, I would like to briefly indicate another possible weakness of Myers-Schulz and Schwitzgebel interpretation of the Radford scenario (or similar scenarios).

In actual fact, I want to claim that there is a sense in which *we* may *say* that Kate knows and disbelieves that Queen Elizabeth died in 1603. This, however, comes at a cost: if we grant that, *we* must claim that there is an unexplained asymmetry between the third- and first-person cases of attitude attributions. Consider the following version of the original story:

You have spent many hours studying for your history exam and you are now in class taking the exam. Everything is going quite well, until you come to the final question. It reads, "In what year did Queen Elizabeth die?" You have reviewed this date many times... (then the story continuous as in the original scenario). After the exam you meet a friend who asks, "What was your answer to the question: In what year did Queen Elizabeth die?" You reply: "1603." The friend applauds: "That is correct! I am glad that you knew it!"

What will be your *honest* reply to this comment? I have no doubt that it will be along the lines of: "No, I didn't. I had to guess and I was lucky. I knew it before the exam but when the teacher announced that the class session was almost over I simply couldn't recall the date."¹² (The answer to the belief question is going to be straightforwardly negative.) Now, the only difference between the scenarios is that we have switched the third person to the first person. It follows that, in fact, here we do not have a *knowledge without belief* scenario but rather an *ignorance with disbelief* scenario. I must admit that I have no idea how proponents of compatibilism are going to explain the asymmetry between the two cases. I prefer to think of both scenarios as analogous and the replies that confirm the hypothesis endorsed by compatibilists as incorrect. Compatibilism seems to be guilty of a one-sided diet that "nourishes one's thinking with only one kind of example" (*Philosophical Investigations*, p. 593)¹³.

Problem 2: The considerations presented above seem to be committed to (one or other) version of epistemic and doxastic logic. But the very idea of such logic is highly problematic, as "(...) there are knowers and knowledge constituting counterexamples to every extant theorem (...) of every epistemic logic" (Hocutt, 1972, p. 435). (The same might be said, of course, of believers and

¹² Experimental philosophers might want to test this (I personally do not think it is worth testing). Probably, the scenario in which the summarizing question is: "Would the reply 'Yes, I knew it' be the honest answer to your friend's comment?" is going to be the most empirically valuable.

¹³ Let me add that neither Rose and Schaffer, nor Buckwalter and Rose note the first- and third-person asymmetry problem, they are also not interested in agents capable of having higher order attitudes. Both issues, I think, may happen to be of some relevance to the study of differences in perspective and "projective use of language" in attitude attribution studies (the relevance of perspective and projectiveness is stressed by Rose and Schaffer).

doxastic logic.)

Reply: Actually, the argument remains correct no matter what stance towards epistemic or doxastic logic one takes (the author, for instance, is very sympathetic towards Hocutt's skepticism). The only assumption one has to endorse is that of the minimal rationality of cognitive agents. By “minimal rationality” I mean the following constraint: if a believes that p and if q follows from p , then if a had been informed about the structure and content of the correct argument from p to q and if s/he grasps the argument (which requires, among other things, that s/he is capable of grasping q), s/he would have added q to her/his beliefs (e.g. if s/he had believed not- q before, she would have given up this belief)¹⁴.

For the sake of illustration, consider the principle [6] ($\Box \forall x (B_x p \Leftrightarrow B_x B_x p)$). One of the obvious cases against it (and similar rules) appeals to iteration and the finiteness of cognitive agents (cf. Rescher, 2005, p. 22): each cognitive agent seems to have a limited attentiveness that excludes the possibility of being disposed to sufficiently complex meta-attitudes (exceeding the ability to have tenth-order thoughts, let us say). This, however, might be represented just as an additional individual constraint imposed on an agent's capacity to grasp certain propositions¹⁵. The constraint is applicable both to cases in which you consider a certain embedding of attitudes as possible object of direct (i.e. non-inferential) awareness, and to cases where you consider it a possible consequence of your first-order belief that p . One may additionally note that our considerations appealed to the embedding of attitudes of minimal complexity only.

Problem 3: The argument appeals to reflective agents who have the disposition to form meta-attitudes. One may still claim that compatibilism is true of agents that are not in that group (young children, some or all animals, etc.)

Reply: Yes, the argument does not apply to such unreflective agents. However (although this remains a matter of empirical controversy), there are reasons to believe (*vide* false-belief task and its interpretations) that agents who do not have such abilities project their own internal states onto

¹⁴ The idea is, of course, taken from Hintikka who writes (explaining the notions of defensibility and indefensibility):

“(…) suppose that a man says to you, 'I know that p but I don't know whether q ,' and suppose that p can be shown to entail q logically by means of some argument he would be willing to accept. Then you can point out to him that what he says he does not know is already implicit in what he claims he knows. If your argument is valid, it is irrational for our man to persist in saying that he does not know whether q is the case. If he is reasonable, you can thus persuade him to retract one of his statements without imparting to him any fresh information beyond certain logical relationships (the rules of which he is assumed to have mastered from the beginning). You have done this by pointing out to him that he would have come to know that q all by himself if he had followed the consequences of what he already knew far enough” (Hintikka, 1962, p. 31).

¹⁵ In epistemic logic the constraint might be imposed purely syntactically: 'If p is a formula, then $B_x p$ is a formula only if p contains no more than n iterations of B.' n might be different for distinct cognitive agents. If n is zero, then one may say that the 'logic' applies to a non-reflective agent.

others. The question now is what argument proponents of incompatibilism are going to provide here in support of their view. They probably cannot follow the empirical route of Myers-Schulz and Schwitzgebel because, as we noted above, there is an asymmetry between the first- and third-person scenarios here. Of course, I do not want to claim that this proves the correctness of (KB) for unreflective agents (that would be a case of *argumentum ad ignorantiam*). Nonetheless, since compatibilism is far from being an intuitive and obvious claim (both for reflective and unreflective agents), it is clearly in need of additional rationale.

Problem 4: By defending the (KB) thesis you are committed to attempting to give sufficient and necessary conditions for the application of the concept of knowledge. Such attempts to define knowledge are known to be theoretical dead-ends.

Reply: I am not committed to such definitional endeavors. What is defended is a claim to the effect that *had such definitions been possible, they would have had to contain the (KB) claim* (or a clause that entails it). If they are not possible, I am very happy to remain with the claim that believing is a necessary condition of knowing.

I conclude that there are reasons to think that the traditional picture is correct. Besides, the reasons to think otherwise are far from convincing. It seems, therefore, that we should remain skeptical about the philosophical prospects of compatibilism and incompatibilism.

References

- Abelson, R. (1968). Knowledge and belief. *Journal of Philosophy*, 65(22), 725–732.
- Black, C. (1971). Knowledge without belief. *Analysis*, 31(5), 152–156.
- Buckwalter, W. & Rose, D. (2015) Belief through thick and thin. *Noûs* 49 (4): 748-775.
- DeRose, K. (1992). Contextualism and knowledge attributions. *Philosophy and Phenomenological Research*, 52, 913–929.
- Feit, N., & Cullison, A. (2011). When does falsehood preclude knowledge? *Pacific Philosophical Quarterly*, 92, 283–304.
- Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 24, 121–123.
- Hemp, D. (ONLINE) The KK (knowing that one knows) principle. In *Internet Encyclopedia of Philosophy*. <<http://www.iep.utm.edu/kk-princ/>>
- Hintikka, J. (1962). *Knowledge and belief. An introduction to the logic of two notions*. Ithaca. New York: Cornell University Press.

- Hintikka, J. (1970). "Knowing that one knows" reviewed. *Synthese*, 21, 141-162.
- Hocutt, M. O. (1972). Is epistemic logic possible? *Notre Dame Journal of Formal Logic*, 13(4), 433–453.
- Lehrer, K. (1990). *Theory of knowledge*. Boulder: Westview Press.
- Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, 77, 549–567.
- Mannison, D. S. (1976). 'Inexplicable knowledge' does not require belief. *Philosophical Quarterly*, 26(103), 139–148.
- Myers-Schulz B., & Schwitzgebel, E. (2013). Knowing that p without believing that p. *Noûs*, 47(2), 371–384.
- Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Harvard University Press.
- Rescher, N. (2005). *Epistemic logic. A survey of the logic of knowledge*. Pittsburgh: University of Pittsburgh Press.
- Radford, C. (1966). Knowledge – by examples. *Analysis*, 27(1), 1–11.
- Rose, D. & Schaffer J. (2013) Knowledge entails dispositional belief. *Philosophical Studies* 166: 19-50.
- Russell, B. (1992). *The analysis of mind* (11th ed.). London: Routledge.
- Tokarz, M. (1990). On the logic of conscious belief. *Studia Logica*, 49(3), 321-332.
- Williamson, T. (2000). *Knowledge and its limits*. New York: Oxford University Press.