Erkenn (2009) 71:107–121
DOI 10.1007/s10670-009-9170-1

ORIGINAL ARTICLE

Self-Knowledge and the Bounds of Authenticity

Sven Bernecker

Received: 10 March 2009 / Accepted: 10 March 2009 / Published online: 29 April 2009

Abstract This paper criticizes the widespread view whereby a second-order judgment of the form 'I believe that *p*' qualifies as self-knowledge only if the embedded content, *p*, is of the same type as the content of the intentional state reflected upon and the self-ascribed attitude, belief, is of the same type as the attitude the subject takes towards *p*. Rather than requiring identity of contents across levels of cognition self-knowledge requires only that the embedded content of the second-order thought be an entailment of the content of the intentional state reflected upon. And rather than demanding identity of attitudes across levels of cognition self-knowledge demands only that the attitude of the intentional state reflected upon and the attitude the subject self-attributes share certain features such as direction of fit and polarity.

Since propositional attitudes have a propositional as well as an attitudinal component knowing of one's propositional attitudes requires knowing of both components. Self-knowledge of propositional attitudes consists in a two-fold classification. The bulk of the philosophical literature on self-knowledge focuses on the authoritative nature of the access we have to the contents of our propositional attitudes and pays little or no attention to the attitudinal aspect. Moreover, the literature is almost exclusively concerned with knowledge of current propositional attitudes, for presumably it is only our current states that we can have authoritative access to. Counterbalancing the one-sidedness of the literature, this paper addresses the knowledge we have of both the contents and attitudes of both our present and past propositional attitudes.

Knowledge is factive in the sense that an utterance of 'I know that *p*' is true only if *p* is the case. Similarly, knowing that I believe that *p* entails that I believe that *p*.

S. Bernecker (✉)
Department of Philosophy, University of California, Irvine, Irvine, CA 92697-4555, USA
e-mail: s.bernecker@uci.edu

Belief is non-factive. When the first-order propositional attitude reflected upon is non-factive, as in knowing that I believe that *p*, the factivity constraint on knowledge requires only that the embedded content of the self-knowledge—in this case '*p*'—be the same as, or sufficiently similar to, the content of the designated first-order state—the belief that *p*; whether *p* is true is irrelevant to the truth of the knowledge claim.[1] For when I claim to know that I believe that *p* I claim to know the particular attitude I take towards a particular proposition, not whether the proposition is true; I claim something about how things appear to me, not about how things are. When the first-order attitude is factive, however, it is not enough that I faithfully report my first-order content. In the case of self-knowledge of factive attitudes the content embedded in the known content, *p*–must be veridical. I cannot know that I know (remember or see) that *p* unless *p* is the case.[2]

Regardless of whether a first-order propositional attitude is factive, having first-person knowledge about it requires faithfully representing both its content and its attitude. The second-order judgment must be an *authentic* rendering of the first-order state. 'Authenticity,' as I use the term, refers to the accuracy of the representation of a first-order intentional state by means of a second-order judgment. Self-knowledge of factive and non-factive attitudes alike has a mind-at-the-second-order-to-mind-at-the-first-order direction of fit; self-knowledge of factive attitudes, in addition, has a mind-at-the-first-order-to-world direction of fit.[3] Knowledge of one's factive attitudes must be true not only to one's first-order representation of reality but also to reality itself.

Since authenticity allows for degrees the question arises of when a second-order representation of an intentional item is sufficiently authentic to qualify as a piece of self-knowledge. What counts as a faithful representation of the content and attitude of some first-order intentional state? What is the permissible range of aberration between the content and attitude of one's first-order state and the content and attitude one self-ascribes by means of a second-order judgment? The aim of this paper is to specify the bounds of authenticity regarding self-knowledge. Though most of the discussion pertains to diachronic self-knowledge the proposed standard of authenticity applies just as well to synchronic self-knowledge.

[1] Just talking about 'embedded content' is a bit simplistic. Suppose I know that I believe that I see that *p*. 'That I believe that I see that *p*' is embedded; and so is 'that *p*.' So what is the target embedded content? We can't simply say that the target embedded content is expressed by what follows a 'knows that' clause. For consider: I know that I believe that I know *p*. In this case there are two 'knows that' clauses. We may say that the target embedded content is the one following the 'knows that' clause that has the largest scope. To not add unnecessary complications I will prescind from cases where the propositional attitude reflected upon contains further attitudes.

[2] Putting the point in terms of the factivity of first-order attitudes isn't quite right when the propositional attitude reflected upon contains further attitudes. For consider the example from footnote 1: I know that I believe that I know *p*. Here the first-order attitude is factive but it *is* enough that I faithfully report my first-order content; *p* doesn't have to be true when embedded in this way. Yet I will prescind from cases where the propositional attitude reflected upon contains further attitudes.

[3] There is an important contemporary strand of thinking according to which taking myself to have a certain intention plays a constitutive role in determining the intention I have (cf. Moran 2001; Wright 1987). Such an 'enactivist' or 'constitutionist' position complicates the issue of direction of fit. Yet I will prescind from discussing the constitutionist position in this paper.

Section 1 is a critical discussion of the widespread view whereby a second-order judgment of the form 'I believe that *p*' qualifies as self-knowledge only if the embedded content, *p*, is of the same type as the content of the intentional state reflected upon and the self-ascribed attitude, belief, is of the same type as the attitude the subject takes towards *p*. Section 2 offers a preliminary analysis of diachronic self-knowledge. Section 3 explains and defends a novel criterion of authenticity for content representation and Sect. 4 does the same for attitude representation. Section 5 offers some concluding remarks.

1 The Inclusion Account of Self-Knowledge

There is widespread agreement that, at least sometimes, when reflecting on an occurrent thought, one knows what one is thinking in a way no one else could regarding one's thought. Knowledge of certain of one's mental states is epistemically direct or immediate in some sense (e.g., in being non-inferential or not being based on evidence), and thus authoritative, perhaps in being incorrigible, or infallible, or transparent to oneself. The most plausible examples of authoritative self-knowledge are what Tyler Burge (1988) calls *cogito-like judgments*, that is, judgments about one's conscious and current first-order thoughts. He offers as examples of cogito-like judgments those such as 'I am now thinking that writing requires concentration' and 'I hereby judge that examples need elaboration.'

Many philosophers claim that the authoritativeness of second-order thoughts is due to the distinctive causal (or functional) relations between them and the states they are about. The idea is that second-order thoughts are justified by the causal process from which they result. Since proponents of the causal account tend to embrace epistemic reliabilism, a subject is said to be able to enjoy privileged access to his intentional states while lacking internalist justification. The problem with the causal account, pointed out by Brie Gertler (2002), is that while it can account for highly reliable self-knowledge it cannot show that there is a principled epistemic disparity between self-knowledge and other-knowledge. The highest epistemic privilege for self-knowledge which the causal account can accommodate is one which, on that account, other-knowledge can also meet.

An alternative account of self-knowledge developed by Burge (1988, 1996) and Heil (1988, 1992, Chap. 5) holds promise as a way to capture the principled disparity between self-knowledge and other-knowledge. On this account a second-order thought is authoritative because its content is contextually self-verifying—it contains as a constituent the first-order state it is about. In the case of cogito-like judgments, Burge maintains,

> One simultaneously thinks through a first-order thought… and thinks about it as one's own…. [B]y its reflexive, self-referential character, the content of the second-order judgment is logically locked (self-referentially) onto the first-order content which it both contains and takes as its subject matter (1988, pp. 659–660).

Since the first-order content is included in, or contained by, the reflexive content of the second-order thought first- and second-order contents cannot come apart and no errors are possible in cogito-like judgments (ibid. p. 658). In my (1995) I have dubbed this account of privileged access the *inclusion theory of self-knowledge*.[4]

Given the inclusion account, when one reflects upon a current, conscious intentional state and consequently forms a second-order thought, the second-order thought is sufficiently authentic only if it contains the same content-type as that tokened in the intentional state it is about. The content of the reflective thought must have a constituent of the same type as the content of the state reflected upon. The inclusion theory has been developed to account for the authoritativeness of synchronic self-knowledge but it has also been extended to diachronic self-knowledge (cf. Burge 1993, 1998; Peacocke 1996).

The paper argues that while the inclusion account is successful in explaining the epistemic disparity between certain kinds of self-knowledge and other-knowledge it is a mistake to infer from this that the authenticity criterion assumed by the inclusion account holds for self-knowledge in general. If self-knowledge in general required that the embedded content of the second-order state is type-identical with the relevant first-order content token it would be a very rare commodity indeed. This is particularly perspicuous in the case of diachronic self-knowledge but applies equally to synchronic self-knowledge. The reason it is rare to find a perfect match between the embedded content of a second-order judgment and the content of the corresponding past intentional state is that most of our diachronic self-knowledge rests on memory and that memory is not only a passive device for reproducing contents but also an active device for processing stored contents. The psychologist Susan Engel explains:

> Research has now shown that… retrieval is almost always more a process of construction than one of simple retrieval. One creates the memory at the moment one needs it, rather than merely pulling out an intact item, image, or story. This suggests that each time we say or imagine something from our past we are putting it together from bits and pieces that may have, until now, been stored separately. Herein lies the reason why it is the rule rather than the exception for people to change, add, and delete things from a remembered event (1999, p. 6).

There is, of course, a difference between saying, as I do, that memory need not amount to the exact reproduction of some previously recorded content and saying, as Engel does, that, as a matter of principle, memory constructs rather then reproduces previously recorded contents. Engel seems o lose sight of the factivity and authenticity constraints on memory.

The fact that our memory not only stores but also processes the incoming information should not be regarded as an abnormal lapse of an otherwise reliable cognitive faculty, but as part of the very function of memory. Since much of our diachronic self-knowledge rests on memory it would be implausible to demand that

[4] Advocates of the inclusion theory of self-knowledge, besides Burge and Heil, are Davidson (1989), Gertler (2001), Macdonald (2007), Sawyer (2002), and with reservations Peacocke (1996).

a second-order judgment concerning a past intentional state must contain the same content-type as that tokened in the past intentional state. But, of course, when the reconstructive nature of memory gains the upper hand the distinction between diachronic self-knowledge, on the one hand, and confabulation, on the other, becomes blurred.

Focusing on the diachronic rather than the synchronic case helps to see that type-identity of contents across levels of cognition is not a necessary condition of self-knowledge in general. For this reason I will concentrate on knowledge of one's past intentional states. Section 2 offers a preliminary analysis of diachronic self-knowledge. Sections 3 and 4 examine the question in what respect and to what extent two diachronic propositional attitude tokens may differ from one another and one of them still count as an authentic representation and knowledge of the other.

2 The Analysis of Diachronic Self-Knowledge

A preliminary analysis of diachronic self-knowledge of a non-factive attitude looks like this: *S* knows at t_2 that he represented (at t_1) that *p*, where 'to represent' stands for some non-factive attitude only if

(1) *S* justifiably believes at t_2 that he represented (at t_1) that *p*,
(2) *S* represented at t_1 that p^*,
(3) is identical with, or sufficiently similar to, p^*,
(4) the attitude that *S* believes at t_2 himself to have taken (at t_1) towards *p* is the same as, or sufficiently similar to, the attitude that *S* took at t_1 towards p^*.[5]

The four conditions may be labeled, respectively, the *justified belief condition* (1), the *past representation condition* (2), the *content condition* (3), and the *attitude condition* (4).

The *justified belief condition* (1) claims that knowing that you represented that *p* implies believing that you represented that *p* where this belief is evidentially supported. To believe something a person need neither to actively reflect on it nor to be absolutely certain that it is true. All that is required for belief is some kind of acceptance in the interest of obtaining truth.[6] The justification component of (1) is there to prevent lucky guesses from counting as knowledge when the guesser is sufficiently confident to believe his own guess. It is one of the vexing problems of epistemology to specify the kind of justification that transforms true belief into knowledge and to account for the epistemic privilege that self-knowledge is

[5] Two terminological notes: First, the value of the index in the subscript to '*t*' determines whether the time referred is in the past or the present: the relatively biggest number indicates the present. So here 't_2' is the present and 't_1' is the past. When there is more than one past time involved, 't_1' indicates the distant past, 't_2' the close past and 't_3' the present. Second, the addendum 'at t_1' is put in parenthesis because we frequently know about our past intentional states without knowing the exact time at which we entertained these states.

[6] Williamson (2000, Chaps. 1–3) argues that knowledge is a simple and irreducible mental state, a mental state that cannot be explained in terms of belief plus certain other conditions. Limitations of space prevent me from examining Williamson's thesis.

supposed to have vis-à-vis knowledge of things other than one's own mind. This paper, however, is not concerned with either the nature of justification or the authoritativeness of self-knowledge.

The motivation behind the *past representation condition* (2) is that knowledge implies truth. I can only know that I believed (at t_1) that p if it is indeed the case that I believed at t_1 that p*. The verb 'to represent' in condition (2) is meant to cover all sorts of non-factive attitudes towards a proposition; believing is among these attitudes but is not the only one.

The *content condition* (3) states that the content of diachronic self-knowledge must be the same as, or sufficiently similar to, the thought content one has entertained previously. By allowing that the contents of the past and the present attitudes are not identical but only sufficiently similar condition (3) contradicts the standard view. As was explained in Sect. 1, the standard view has it that for a second-order judgment of the form 'I believed (at t_1) that p' to qualify as a piece of self-knowledge the embedded content, p, must be of the same type as the content of the belief one had at t_1, p*. The standard view comes in two flavors: either the embedded content is conceived of as the reproduction of the past content—the causal model—or it is construed as the activation of the past thought content token—the inclusion model. Either the embedded content of self-knowledge must be type-identical to the past content (the causal model) or it must be token-identical (the inclusion model).[7] What presumably motivates either version of the standard view is the worry that if the embedded content of self-knowledge were allowed to differ from the original content this would fly in the face of the undeniable doctrine that knowledge implies truth. Yet as will be shown in Sect. 3, a liberalization of the authenticity standard for self-knowledge need not conflict with the factivity of knowledge.

For the same reason that diachronic self-knowledge does not require that the embedded content be of the same type as the content of the original representation, it does not require that the psychological attitude ascribed to one's former self be of the very same type as the psychological attitude one did occupy in the past. It is sufficient that past and present psychological attitude tokens merely be similar; or at least so the *attitude condition* (4) asserts.

Quine tells us that "there is nothing more basic to thought and language than our sense of similarity; our sorting of things into kinds" (1969, p. 116). One function of similarity is, e.g., to allow us to make educated guesses in face of limited knowledge. Notwithstanding the importance of similarity judgements, '*A* is similar to *B*' is a meaningless statement unless one can say in what respect *A* and *B* are similar. And, as Nelson Goodman has shown, similarity in terms of respects makes similarity superfluous. "[T]o say that two things are similar in having a specified

[7] Locke seems to endorse the inclusion model of diachronic self-knowledge in the second edition of his *Essay*: "But our *Ideas* being nothing, but actual Perceptions in the Mind, which cease to be any thing, when there is no perception of them, this *laying up* of our *Ideas* in the Repository of Memory, signifies no more but this, that the Mind has a Power, in many cases, to revive Perceptions, which it has once had…. And in this Sense it is, that our *Ideas* are said to be in our Memories, when indeed, they are actually no where, but only there is an ability in the Mind, when it will, to revive them again; and as it were paint them anew on itself" (1694, p. 150).

property in common is to say nothing more than that they have that property in common" (1972, p. 445). The meaning of similarity is conveyed by the specific respects, not by the general notion of similarity. That is why Goodman concludes that similarity "is insidious," that it is "a pretender, an imposter, a quack," that it is a "false friend" (1972, p. 437).

Granted these strictures on similarity, the notion of similarity employed in the content and the attitude conditions (3) and (4) must be replaced by notions whose explanatory value is unproblematic. In Sect. 3 the notion of content similarity will be explicated in terms of the entailment relation. And in Sect. 4 the notion of attitude-similarity will be defined in terms of sameness of direction of fit and polarity.

Diachronic self-knowledge of *factive* attitudes is governed by the same conditions as diachronic self-knowledge of non-factive attitudes and by two additional truth conditions:

(5) p is true at t_2
(6) p^* is true at t_1

Though I can know that I *believed* that the moon is made of green cheese I cannot know that I *knew* or *remembered* that the moon is made of green cheese. The reason is that knowledge, like memory and unlike belief requires truth. In the case of diachronic self-knowledge of factive attitudes both the embedded content, p, and the content of the past attitude, p^*, must be veridical. In other words, diachronic self-knowledge of factive attitudes demands not only that the proposition emerging from the memory process be veridical but also that the proposition fed into the memory process be veridical. (In the case of timeless truths it is, of course, sufficient to state only one truth condition.)

Though the ability to know one's past attitudes frequently rests on the ability to remember them, it is worth noting that the conditions of memory are not the same as the conditions of diachronic self-knowledge. For example, it is widely thought that memory is governed by a *causal condition* which states that *S*'s belief at t_2 to the effect that he represented (at t_1) that p is suitably causally connected to *S*'s representation at t_1 that p^* (cf. Bernecker 2008, Chaps. 2–4). The goal of the causal condition of memory is to exclude re-learning from the ranks of remembering and to establish that the representation had in remembering is a retained representation. Though an indispensable component of the analysis of memory, the causal condition does *not* apply to diachronic self-knowledge. Knowing one's past attitudes does not require that the past attitude be the cause of one's present state of knowing—unless, of course, one subscribes to a crude causal theory of knowledge of the kind Alvin Goldman (1967) proposed four decades ago. And just as the causal condition is a necessary component of the analysis of memory but not of knowledge, the justification condition is a necessary component of the analysis of knowledge but not of memory. For unlike knowledge, memory does not imply justification. Not only is it possible to remember something one did not justifiably believe in the past but also one might acquire between t_1 and t_2 some misleading but reasonable evidence that destroys the status as justified belief of the once-genuine justified belief which one still remembers. What enters the memory process and what leaves

it may be merely a belief, not knowledge. Knowledge supervenes on some but not all cases of remembering (cf. Bernecker 2007).

3 Knowing One's Contents

For two diachronic content tokens to be suitably similar in the sense of the content condition (3) the present embedded content must be entailed by the past content. Diachronic self-knowledge requires that the content ascribed to one's earlier self be an entailment of the original content. Entailment is the criterion of authenticity. I shall label this the *entailment thesis*:

> *Entailment Thesis*: A second-order judgment at t_2 qualifies as an instance of self-knowledge of a first-order attitude at t_1 only if the embedded content of the second-order judgment is entailed by the content of the first-order attitude.

Suppose you came to believe at t_1 that John F. Kennedy (JFK, for short) was assassinated. At t_2, all you can remember is that JFK died of unnatural causes—you have forgotten the circumstances of his death. You form a second-order judgment to the effect that you believed (at t_1) that JFK died of unnatural causes. Notwithstanding the fact that *JFK was assassinated* and *JFK died of unnatural causes* are different propositions, it is natural to suppose that the second-order judgment qualifies as self-knowledge—provided the justified belief condition (1) is met. The reason the difference between the original belief content and the self-ascribed belief content does not and should not prevent us from granting diachronic self-knowledge, I suggest, is that the proposition *JFK died of unnatural causes* is entailed by the proposition *JFK was assassinated*.

Insofar as knowledge of one's past attitudes is non-inferential and rests on memory the entailment thesis states that while memory allows for the retrieved content being informationally impoverished in comparison to the original content, it does not tolerate an increase or an enrichment of information. If your original belief had been that JFK died and if, later on, you had claimed to have believed that JFK was assassinated, this would be neither a case of non-inferential memory nor a case of non-inferential self-knowledge.

Since the entailment relation preserves truth, analyzing the notion of content similarity in terms of the entailment relation is perfectly compatible with the authenticity constraint on self-knowledge. If q is entailed by p, and if p is true, so is q. Thus if *JFK was assassinated* is true, so is *JFK died of unnatural causes*. Provided that the contents fed into the memory process are veridical and that there are no external circumstances changing the truth values of the stored contents, the entailment thesis ensures that the remembered contents are veridical as well. (To not add unnecessary complications I will prescind from cases where the truth value of a proposition changes while it is retained in memory.) And since each proposition entails itself the entailment thesis is also compatible with the inclusion account of self-knowledge.

If entailment is understood in terms of the material implication, the entailment thesis is too liberal to provide a plausible account of the authenticity of diachronic

self-knowledge.[8] The truth-table for material implication tells us that any conditional with a false antecedent is true and so is any conditional with a true consequent. But it is an intolerable result that a second-order judgment of the form 'I believed that *p*' qualifies as self-knowledge of *any* past belief in a false proposition. A further problem for the entailment thesis is that each contradictory (or impossible) proposition entails every proposition and that each necessary proposition is entailed by every proposition.

The entailment thesis is too liberal even when we prescind from necessary and impossible propositional contents as well as from false propositional contents. Since every proposition entails infinitely many propositions the entailment thesis allows for some far-fetched entailments of one's past thoughts to count as instances of diachronic self-knowledge. Take the proposition *JFK was assassinated* and the proposition *air molecules moved*. Given the physical conditions of Earth, the proposition *JFK was assassinated* entails the proposition *air molecules moved*. Yet the thought that air molecules moved can hardly count as an authentic representation of one's past thought that JFK was assassinated. And so we need to impose a limit on how far removed a propositional content may be from its implicandum for it to qualify as an authentic representation of the implicandum. We need to ensure that the antecedent and the consequent are thematically related, are on the same topic.

I suggest that the notion of entailment employed by the entailment thesis should be interpreted along the lines of Anderson and Belnap (1975) relevance logic. According to Anderson and Belnap, p is relevant to q iff q can be inferred from (not just under) p; that is, iff p could be used in a proof of q from p. For instance, they reject $p \rightarrow (q \rightarrow q)$ because p may be irrelevant to $(q \rightarrow q)$ in the sense that p is not used in arriving to $(q \rightarrow q)$. In order to infer q from p it is necessary that p and q have some common meaning content. Since Anderson and Belnap think that in propositional logic commonality of meaning is carried by commonality of propositional variables, they conclude that p and q should share at least one propositional variable. And when p and q share a propositional variable, then they are thematically related.

Even when the notion of entailment employed by the entailment thesis is read along the lines of relevance logic, a critic might worry that the entailment thesis is still too liberal. Consider the belief that JFK was assassinated in 1963 and the belief that Jacqueline Kennedy became a widow in 1963. Despite these beliefs sharing at least one propositional variable it might be argued that the latter cannot qualify as an authentic representation of the former.

Frequently the thought contents we ascribe to our past selves contain some additional information or background knowledge. When you believed at t_1 that JFK was assassinated in 1963 and when you now claim to know that you believed at t_1 that Jacqueline Kennedy became a widow in 1963, you are using the verb 'to know' in an elliptical sense. When made aware of the fact that the embedded content of your present second-order judgment differs from your past belief content, you would presumably substitute for the statement 'I believed at t_1 Jacqueline Kennedy became a widow in 1963' something like the following conjunctive statement: 'I

[8] The terms 'to entail' and 'to imply' are used interchangeably here.

believed at t_1 that JFK was assassinated in 1963 and since JFK was married to Jacqueline Kennedy it follows that Jacqueline Kennedy became a widow in 1963.' The statement 'I believed at t_1 Jacqueline Kennedy became a widow in 1963' is an ellipsis, the meaning of which is given by the conjunctive statement. The first conjunct expresses what you in fact believed and the second conjunct expresses some additional information.[9] Rather than being a weakness, it is a strength of the entailment thesis that, unlike the inclusion theory, it can account not only for non-inferential but also for inferential self-knowledge.

Next consider the following putative counterexample to the entailment thesis. At t_1 you came to believe that if p, then q and non-q. At the time you did not realize that what you believe entails non-p. At t_2 you put two and two together and claim to have believed at t_1 that non-p. According to the entailment thesis, your second-order judgment at t_2 may qualify as an instance of self-knowledge. A critic might object, however, that you cannot know at t_2 that you believed at t_1 that non-p. The reason you cannot know this is that it is not true. You did not believe non-p, though, of course, you could have believed it had you not failed to draw a deduction from what you in fact believed.

Though the second-order judgment that I believed that non-p cannot qualify as non-inferential knowledge of my past belief that if p, then q and non-q, it can qualify as inferential knowledge. The statement 'I know that I believed at t_1 non-p' is an ellipsis, the meaning of which is given by the conjunctive statement 'I believed at t_1 that if p, then q and non-q *and* now (at t_2) I know that this entail non-p.'

Relevant entailment licences the move from conjunctions to conjuncts ($(p \wedge q) \rightarrow p$) and from disjuncts to disjunctions ($p \rightarrow (p \vee q)$). The move from conjunctions to conjuncts is in accord with our intuitions regarding the authenticity of self-knowledge. Suppose you came to believe at t_1 the proposition *JFK was assassinated and Lee Harvey Oswald is the culprit*. At t_2 you claim to have believed (at t_1) that JFK was assassinated but you can't recall whom you took to be the culprit. I reckon we all agree that your second-order judgment at t_2 to the effect that you believed (at t_1) that JFK was assassinated counts as knowledge—provided, of course, the justified belief condition (1) is met.

Next consider disjunction introduction. Suppose you believed at t_1 that JFK was assassinated. At t_2 you are not quite sure what you took to be JFK's cause of death. You claim having believed at t_1 that JFK was assassinated or died of natural causes. Does the second-order judgment meet the intuitive criterion for self-knowledge, provided the justified belief condition (1) is met? The answer, I take it, is 'yes'. Though the second-order judgment in question may not be sufficiently specific to be of much use it is nonetheless true and contains *some* information about your past belief. Just as usefulness is not a condition on knowledge in general it isn't a condition on self-knowledge. The move from disjuncts to disjunctions yields inferential self-knowledge because there is information added to the original thought content.

Since it is generally assumed that belief reports should be understood *de dicto* non-inferential self-knowledge does not allow for the substitution of coreferential expressions. Suppose you came to believe at t_1 that JFK was assassinated. At t_2 you

[9] For an analogous analysis of elliptical memory see my (2007, pp. 152–154).

claim to have believed at t_1 that the 35th President of the United States was assassinated. Your second-order judgment may qualifies only as an instance of inferential self-knowledge.

In the case of non-inferential self-knowledge, it is not enough to demand that the embedded content of the second-order judgement be entailed by the content of the first-order attitude; it must also be the case that no additional premise is needed and used by the agent to derive the embedded content of the second-order judgment from the content of the first-order attitude.

Let's take stock. I have rejected the claim whereby diachronic self-knowledge requires the embedded content of the second-order state to be type-identical with the relevant past content on psychological reality grounds. Rather than demanding identity of content, diachronic self-knowledge demands only that the present embedded content be an entailment of the past one. Thus I propose replacing the content condition for diachronic self-knowledge (3) by the following revised content condition (3′):

(3′) p is entailed by $p*$,

where 'entailed' is understood along the lines of relevance logic. In the case of non-inferential self-knowledge there is the further requirement that no additional premise is needed and is used by the agent to derive p from p*.

4 Knowing One's Attitudes

Given that diachronic self-knowledge does not require that the present embedded content be of the same kind as the past content, it is reasonable to suppose that diachronic self-knowledge does not require that the psychological attitude ascribed to one's former self be the same as the attitude one took towards the proposition in question. Both attitudes need to be only sufficiently similar. This raises the question of which features diachronic attitude tokens must share to count as sufficiently similar. What is the permissible range of aberration between the attitude identified in a self-knowledge report and the original attitude? In preparation for an answer to this question we must identify the criteria which allow us to classify different attitudes towards a proposition—polarity, direction of fit, factivity, and factorability.

Propositional attitudes can be categorized according to their polarity as positive or negative. My attitude towards something is positive (a pro-attitude) if I favor the thing or am favorably disposed towards it, negative (a con-attitude) if I view it unfavorably. Believing, hoping, suggesting, and desiring are examples of positive attitudes; denying, doubting, being afraid, and disclaiming are examples of negative attitudes. Just as one can have a positive attitude towards a proposition that expresses a state of affairs one dislikes (e.g., I believe that I face a threat) one can have a negative attitude towards a proposition which expresses a state of affairs one places a value on (e.g., I doubt that my lottery ticket will win).

In cognitive attitudes (such as belief), a proposition is grasped as patterned after the world; whereas in conative attitudes (such as desire), the proposition is grasped as a pattern for the world to follow. Beliefs aim to track the world and have a mind-

to-world direction of fit. Desires, on the other hand, aim to impose themselves onto the world and have a world-to-mind direction of fit.

Some attitudes with a mind-to-world direction of fit imply truth. Knowledge, for example, is factive in the sense that an utterance of '*S* knows that *p*' is true only if *p* is the case. If not-*p*, then *S* may think he knows that *p*, but cannot actually know that *p*. Other factive attitudes with a mind-to-world direction of fit are remembering, learning, and seeing. Among non-factive attitudes with a mind-to-world direction of fit are believing, suggesting and considering.[10]

There are simple attitudes and there are complex attitudes that are composed of simple ones. An example of a complex attitude with a mind-to-world direction of fit is knowledge which has belief as one of its components. Intending is an example of a complex attitude with a world-to-mind direction of fit. Intending a certain state of affairs *A* by performing an action of type *B* is usually broken down to having a desire for *A* and believing that in performing *B* one will bring about *A*. By contrast, believing and desiring cannot be analyzed into more basic attitudes.

Let us return to the question of which features tokens of different attitudes need to share to count as sufficiently similar for the purpose of self-knowledge. Since attitudes do not entail each another we cannot use the entailment thesis to spell out the notion of attitude similarity. Yet the idea guiding the entailment approach *can* be applied to the issue of attitude similarity. The idea is that non-inferential self-knowledge allows for the dilution of informational content but not for its enrichment. In the process of coming to non-inferentially know one's past attitudes one may lose information, but may not add information.

To know one's past propositional attitude one must correctly identify its polarity. If one doubted at t_1 that *p* and if one takes oneself at t_2 to have believed at t_1 that *p*, then one clearly fails to know about one's intentional state at t_1. Similarly, if one wanted at t_1 that *p* and if, at t_2, one claims to know that one disliked at t_1 that *p*, one cannot be said to have knowledge about one's past attitude. Polarity allows for degrees of strength. The polarity of the past attitude may be reported as being less intense than it in fact was. Non-inferential self-knowledge allows representing, say, a state of hatred as one of dislike, a state of deep love as one of affection, and a conviction as a belief.

A second-order judgment can qualify as a piece of knowledge only if it represent the past propositional attitude as having the direction of fit it did in fact have. Self-knowledge requires identity of direction of fit across levels of cognition. Consider the following example. At t_1 I hoped that *p*. At t_2, I report having preferred at t_1 that *p*. Hoping and preferring are different attitudes but since they share the mind-to-world direction of fit my second-order judgment may indeed qualify as a piece of self-knowledge.

Since the requirements on factive attitudes are stricter than those on non-factive ones a factive attitude may be represented as a non-factive one. Suppose I saw at t_1

[10] Hazlett (2009) challenges the orthodox view among philosophers that 'knows,' 'remembers,' 'learns' etc. are factive. He argues that if the orthodox view is true, then we should expect the claim that all known propositions are true to be obvious to anyone who knows the meaning of 'knows.' However, the fact that 'knows' or 'remembers' might not be obviously factive for some competent language users is fully compatible with their being indeed factive all the same.

that p. At t_2 I can have non-inferential knowledge whereby it seemed to me at t_1 that p. While non-inferential self-knowledge only allows for factive attitudes to be represented as non-factive ones inferential self-knowledge also allows for non-factive attitudes to be represented as factive ones. Suppose it seemed to me at t_1 that p. When I subsequently learn that the sensory experience met the conditions for sense-perception then I am epistemically entitled to represent at t_2 the ostensible perception of t_1 as an instance of veridical perception.

A complex attitude may be represented as being one of the simple attitudes of which it is composed. Suppose I knew at t_1 that p. At t_2 I claim to have believed at t_1 that p. If I know that knowledge implies belief, there is no reason not to grant me knowledge of my past attitude. After all, I am not violating the factivity constraint on knowledge.[11] There may even be cases of inferential self-knowledge where a simple attitude is represented as a complex attitude of which it is a component. Suppose I believed at t_1 that p. If I subsequently learn that the past belief was justified and true, I am in a position to inferentially know that I knew at t_1 that p.

In light of the sketched analysis of attitude similarity, I propose replacing the attitude condition for diachronic self-knowledge (4) by the following revised attitude condition (4′):

> (4′) The attitude that S represents at t_2 himself having taken (at t_1) towards p and the attitude that S took at t_1 towards p^* are the same or the two attitudes share the direction of fit and polarity.

In the case of non-inferential self-knowledge there are further requirements. In addition to the attitudes having to share the direction of fit and polarity the following conjunction must be met: (1) the polarity of the attitude at t_2 is not more intense than that of the attitude at t_1, (2) if the attitude at t_2 is factive so is the attitude at t_1, and (3) the attitude at t_1 is not a component of the attitude at t_2, nor vice versa.

5 Conclusion

I have argued that while the inclusion account provides a plausible explanation of the epistemic disparity between certain kinds of self-knowledge and other-knowledge it is a mistake to infer from this that self-knowledge in general requires the embedded content of the second-order thought to be type-identical with the content of the intentional state reflected upon. Likewise it is not a necessary condition of self-knowledge in general that the psychological attitude self-ascribed at the second-order be type-identical with the attitude one takes towards the proposition in question. Self-knowledge implies truth but it does not imply identity of contents and attitudes across levels of cognition. Rather than requiring identity of contents self-knowledge requires only that the embedded content of the second-order thought be an entailment of the content of the intentional state reflected upon.

[11] Representing a complex attitude as one of the simple attitudes of which it is composed violates the factivity constraint when the simple attitude is factive and the complex attitude is non-factive. Yet it is doubtful whether there are any non-factive attitudes which have factive attitudes as their components.

And rather than demanding identity of attitudes self-knowledge demands only that the attitude of the intentional state reflected upon and the attitude the subject self-attributes share certain features such as direction of fit and polarity. Though the discussion has focused on diachronic self-knowledge the proposed liberalization of the authenticity criterion applies also to synchronic self-knowledge. For there is no reason to suppose that the constraints on synchronic self-knowledge are stricter than those on diachronic self-knowledge.

Acknowledgments Previous versions of this paper were presented at the University of Manchester (March 2006), the University of Hertfordshire (May 2006), and a conference on first-person authority at the University of Duisburg-Essen (September 2007). I am grateful to the audiences for helpful questions. The paper has been improved by valuable advice from Brendan Larvor, David Makinson, Philip Nickel, Oliver Petersen, Thomas Spitzley, and an anonymous reviewer for this journal.

References

Anderson, A. R., & Belnap, N. D. (1975). *Entailment: The logic of relevance and necessity*. Princeton: Princeton University Press.

Bernecker, S. (1995). Externalism and the attitudinal component of self-knowledge. *Noûs, 30*, 262–275.

Bernecker, S. (2007). Remembering without knowing. *Australasian Journal of Philosophy, 85*, 137–156.

Bernecker, S. (2008). *The metaphysics of memory*. Dordrecht: Springer.

Burge, T. (1988). Individualism and self-knowledge. *Journal of Philosophy, 85*, 649–663.

Burge, T. (1993). Content preservation. *Philosophical Review, 102*, 457–488.

Burge, T. (1996). Our entitlement to self-knowledge. *Proceedings of the Aristotelian Society, 96*, 91–116.

Burge, T. (1998). Memory and self-knowledge. In P. Ludlow & N. Martin (Eds.), *Externalism and self-knowledge* (pp. 351–370). Stanford: CSLI Publications.

Davidson, D. (1989). What is present to the mind? *Grazer philosophische Studien, 36*, 3–18.

Engel, S. (1999). *Context is everything: The nature of memory*. New York: W.H. Freeman.

Gertler, B. (2001). Introspecting phenomenal states. *Philosophy and Phenomenological Research, 63*, 305–328.

Gertler, B. (2002). The mechanics of self-knowledge. *Philosophical Topics, 28*, 125–146.

Goldman, A. (1967). A causal theory of knowing. *Journal of Philosophy, 64*, 357–372.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects* (pp. 437–446). Indianapolis: Bobb-Merrill.

Hazlett, A. (2009). The Myth of Factive Verbs. *Philosophy and Phenomenological Research*. (in press).

Heil, J. (1988). Privileged access. *Mind, 47*, 238–251.

Heil, J. (1992). *The nature of true minds*. Cambridge: Cambridge University Press.

Locke, J. (1694). *An essay concerning human understanding* (2nd ed.). In P. H. Nidditch (Ed.), 1979. Oxford: Clarendon Press.

Macdonald, C. (2007). Introspection and authoritative self-knowledge. *Erkenntnis, 67*, 355–372.

Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton: Princeton University Press.

Peacocke, C. (1996). Entitlement, self-knowledge and conceptual redeployment. *Proceedings of the Aristotelian Society, 96*, 117–158.

Quine, W. V. O. (1969). Natural kinds. In W. V. O. Quine (Ed.), *Ontological relativity and other essays* (pp. 114–138). New York: Columbia University Press.

Sawyer, S. (2002). In defence of Burge's thesis. *Philosophical Studies, 107*, 109–128.

Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
Wright, C. (1987). On making up one's mind: Wittgenstein on intention. In P. Weingartner & G. Schurz (Eds.), *Logic, philosophy of science and epistemology: Proceedings of the 11th international Wittgenstein symposium* (pp. 391–404). Vienna: Hölder-Pichler-Tempsky.