4 *Memory and Self-Knowledge*

In 'Individualism and Self-Knowledge' I argued that immediate, authoritative, non-empirically warranted self-knowledge is compatible with anti-individualism about the individuation of propositional attitudes. Paul Boghossian uses my slow-switching cases to argue that such self-knowledge and anti-individualism are incompatible. I will try to show why this argument does not succeed.[1]

I postulated as possible a case like this: An individual grows up in an environment like ours with a normal set of experiences of a particular sort of object or stuff, say aluminum. The individual does not know anything about the micro-structural features of metals. But he has seen aluminum and made use of it. Perhaps he has heard things about it from others. The individual grows to maturity with this learning history. The individual has a normal lay concept of aluminum. It applies to aluminum and nothing else. I presumed that the individual believes that look-alike metals can be different metals. So his conception of aluminum allows that something could be superficially indistinguishable from aluminum and not be aluminum.

I then imagined that the individual is switched unawares to another planet (either forever, or gradually back and forth staying at each planet a substantial amount of time before switching). Given what he knows and can discern perceptually, he cannot distinguish the second environment from his home environment. The second environment contains twaluminum in all the places aluminum occupies in the original environment. Aluminum is lacking altogether. Twaluminum is a different metal that is indistinguishable for the individual from aluminum. I claimed that the individual's original concept does not apply to twaluminum. But I claimed that if the individual had enough experiences with the new metal comparable to those of the old, or if the individual interacted sufficiently with a community that had a term that applied to twaluminum, the individual would normally eventually acquire a concept more appropriate to the new environment than the concept of aluminum.

In 'Individualism and Self-Knowledge' I assumed that the newly acquired concept would be a "twin" concept. So if one had an aluminum concept in the first environment, one would acquire a twaluminum concept in the second. The

[1] Tyler Burge, 'Individualism and Self-Knowledge', *The Journal of Philosophy* 85 (1988), 649–663; Paul Boghossian, 'Content and Self-Knowledge', *Philosophical Topics* 17 (1989), 5–26.

twaluminum concept does not apply to aluminum. In these cases, the individual's word-form "aluminum" expresses on different occasions two concepts whose extensions are disjoint. Let use call such cases "*Disjoint Type* cases".

I will continue to suppose that Disjoint Type cases are possible. But they are not the only sorts of cases that can arise. The new set of concepts may be broadenings of the old. Thus the concept in the second environment may apply both to aluminum and twaluminum. One pressure in this direction is that these metals are indistinguishable to the individual, and instances of both (let us assume) enter into normal, paradigmatic perceptual applications of the individual's word-form "aluminum". If the individual's concept broadened in this way, the individual would acquire a non-natural-kind concept rather like our actual concept of jade.[2] Call these "*Amalgam Type* cases".

Whether in slow switching we have an Amalgam Type case depends partly, probably mainly, on how the individual is committed to the standards of the communities in the two environments. The relevant individual lacks knowledge of the metals' substructures. In the absence of commitments to communal standards and to communal understanding, and given extensive, prolonged new experiences with twaluminum, the relevant individual will lack the cognitive resources for the referent of his word-form "aluminum" to be fixed as a single natural kind. In the absence of other constraints, only normal experience with metals that are in fact instances of a single natural kind would so fix the referent. In the face of normal, constant perceptual application of the term to different natural kinds (where there is no countervailing pressure having, for example, to

[2] This point does not in the least constitute an admission that in the first environment, apart from actual switching, the individual must express this broadened concept in using the word form "aluminum". The concept is fixed partly by actual causal interactions. I take it that the individual's belief that metals may differ while appearing the same helps prevent his concept from applying from the beginning to just anything that looks to him like aluminum. And in the absence of constraining factors, there is no basis for its applying just to aluminum and twaluminum—as opposed to aluminum and all possible look-alike metals in addition to twaluminum. Before the switches, twaluminum played no more role in the individual's concept acquisition than any other possible metal that might look like aluminum.

These matters are complex, however. Some have held that the individual's future constrains his past, so that if he later comes into contact with twaluminum in such a way as to yield a broadened concept, he always has the broader concept—one that applies to both aluminum and twaluminum. There are interesting questions here about how to determine the limits of normal circumstances. But I do not accept this view as a general analysis. I think there is a constraint in the account of concept determination that for most normal human beings requires a concept to be one of a natural kind if one's experience in normal circumstances is overwhelmingly that of instances of a natural kind. And there is a constraint that allows concepts to change extension with sufficient changes in what is normal within the individual's experience. Allowing the whole future to count in determining an individual's concept would not plausibly accommodate conceptual change through change of normal circumstances. It would also underestimate the role of causation in determining an individual's present cognitive abilities, and would pointlessly complicate accounts of communal sharing of concepts. But I need not go into these matters here. For if the amalgam concept is always the individual's concept, there would be no conceptual change. And this would ease rather than threaten the compatibilist position that I am outlining.

do with differential values or purposes to which the metals might be put), the extension will commonly broaden.[3]

The individual's specific intentions may also play a role. The individual might resist such broadening by intentionally and specifically limiting the original concept to the original paradigms in the first environment. Such resistance might yield a Disjoint Type case or even no change at all. But this sort of resistance is certainly not the norm. In the absence of regular application to a single kind, or special intentions by the individual, it appears that something like relations of communal dependence would be needed to prevent the expression "aluminum" from coming, through contact with the second environment, to express a concept that would apply equally to aluminum and twaluminum. How the individual relates to the two communities bears on whether we have a Disjoint Type or an Amalgam Type case. The relevant parameters are complex and contextual, and I will not try to discuss them here.

In any event, I will suppose that after sufficient switching, it is possible for the individual to have a single concept expressed by the word-form "aluminum" that applies to aluminum and twaluminum. These are Amalgam Type cases.

In both types of case the individual has undergone a conceptual change that is unknown to the individual. In Disjoint Type cases, the individual has taken on two concepts for disjointly different sorts of things without knowing it. In Amalgam Type cases, a concept that comes to be expressed by the individual's word-form includes the extension of the original concept before the switching, but is broader. And again the individual is unaware of the change. I leave open whether in Amalgam Type cases, the individual retains the original concept in his repertoire. In both Disjoint Type and Amalgam Type cases the individual's lack of awareness of the change derives from his inability to distinguish aluminum from twaluminum. The individual cannot explain or articulate a distinction between the concepts in the first and second environments. He is unaware that there are two concepts. So there may be times when the individual is unable to determine whether he was thinking, at some earlier time, about aluminum alone or twaluminum or some amalgam.

Nevertheless, I believe that the individual will commonly have immediate, non-empirically warranted self-knowledge of the form *I think (believe, judge) that p*, where *that p* includes a relevant concept (*aluminum, twaluminum*, or the amalgam concept). Before the switches, the individual might know a *cogito*-like thought in thinking it: *I am hereby thinking that aluminum is a light metal*. And the individual can have self-knowledge of this sort—though the content will be

[3] Regular, frequent switching at relatively short intervals—as distinguished from a single switch, or very infrequent switches with long intervals at each place—seem to make Amalgam Type cases more intuitive for some. In effect, such switching might make the conjunction of the two planets seem more plausibly a single "normal" environment. But I think the main issue has to do not with speed or frequency but with whether additional factors, such as communal factors, serve to distinguish the environments in cognitively relevant ways, once the new planet becomes a factor. All of these matters deserve deeper reflection. Needless to say, there will be many don't cares and borderline cases.

different—after the switches as well. In a Disjoint Type case, for example, the individual can think and authoritatively know after a switch: *I am hereby thinking that twaluminum is a light metal.*

There are many complicated issues associated with this view. I will not go into many of these. And I will not repeat the considerations that support the view, except for one reminder. It is not disputed that the individual can on given occasions think *cogito*- or other self-attributional thoughts with a definite concept, say, the aluminum concept or the twaluminum concept. In relevant self-attributions, the individual simultaneously uses and self-attributes concepts (in the reflexive, that-clause way). In these cases the individual cannot get *the content* wrong.[4] For the attributed intentional content is fixed by what the individual thinks. It is not something that he identifies independently.

Much of the literature on this subject deals with problems that arise from the assumption that we need to *identify* the content of our thoughts in such a way as to be able to rule out relevant alternatives to what the content might be. Boghossian, unlike many of those who write on this subject, seems to recognize that this assumption is not acceptable on my view. One's relation to one's content, when one is non-empirically self-attributing in the reflexive, that-clause way is not analogous to a perceptual, identificational relation to which alternatives would be relevant. In present tense self-attributions of the relevant kind, alternatives are irrelevant. Boghossian's strategy is to consider cases of memory and argue that these cases reflect badly on my views about the present tense cases.

Boghossian writes:

> [Burge's claims] amount to saying that, although [the subject] S will not know tomorrow what he is thinking right now, he does know right now what he is thinking right now. For any given moment in the present, say t1, S is in a position to think a self-verifying judgment about what he is thinking at t1. By Burge's criteria, therefore, he counts as having direct and authoritative knowledge at t1 of what he is thinking at that time. But it is quite clear that tomorrow he won't know what he thought at t1. No self-verifying judgment concerning his thought at t1 will be available to him then. Nor, it is perfectly clear, can he know by any other non-inferential means. To know what he thought at t1 he must discover what environment he was in at that time and how long he had been there. But there is a

[4] Compare Tyler Burge, 'Our Entitlement to Self-Knowledge', *Proceedings of the Aristotelian Society* 96 (1996), 91–116. Boghossian rightly points out what I myself had indicated—that self-verifying judgments are just a small sub-class of the self-knowledge to which we have special authoritative, non-empirical entitlements. I think, however, that they provide a paradigm that is suggestive of many of the key features of the larger class. I have discussed aspects of these matters in the above cited article. All I do here is to deal with his objections to my taking anti-individualism about content to be compatible with seeing self-verifying judgments to be cases of non-observational self-knowledge. But the main line of his argument would apply to all cases of ordinary non-empirical self-knowledge, and my reply does also. So neither my examples of *cogito* thoughts nor Boghossian's remarks about self-verification are central to the main issues about memory that I will be discussing. My points above about the individual's inability to get the content wrong in non-empirical present-tense self-attributions (in the that-clause way) apply not only to self-verifying *cogito* thoughts, but to all non-empirical that-clause type self-attributions. I will argue that the same point carries over to certain types of memory of past self-attributions.

mystery here. For the following would appear to be a platitude about memory and knowledge: if S knows that p at t1, and if at (some later time) t2, S remembers everything S knew at t1, then S knows that p at t2. Now, let us ask: *why* does S not know today whether yesterday's thought was a *water* thought or *twater* thought? The platitude insists that there are only two possible explanations: either S has forgotten or he *never* knew. But surely memory failure is not to the point. In discussing the epistemology of relationally individuated content, we ought to be able to exclude memory failure by stipulation. It is not as if thoughts with widely individuated contents might be easily known but difficult to remember. The only explanation, I venture to suggest, for why S will not know tomorrow what he is said to know today is not that he has forgotten but that he never knew. Burge's self-verifying judgments do not constitute genuine knowledge.[5]

Let the thought *that p* be what individual S believes before the environmental switches occur. For example, S may believe *I am thinking that aluminum is a light metal*. Boghossian's argument is as follows:

(1) If S does not forget anything, then whatever S knows at time t1, S knows at time t2.
(2) In the cases at hand S does not forget anything.
(3) S does not know that p at time t2.
(4) So S does not know that p at time t1.

Let us consider Disjoint Type cases first. So assume that at t2, the individual has a new set of concepts disjoint from the ones that he had before the switching began. He need not have lost the old set, however. I do not concede that the individual 'will not know tomorrow what he is thinking right now' (i.e. at t1), at least in the sense of "knowing what" that is relevant to my view. Moving to the other environment and acquiring new concepts will not normally obliterate old concepts or memories that derive from the first environment. If one always lost all past concepts by acquiring new ones after a switch, one would never be able to remember or report accurately what one had said or thought. The old abilities will normally still be there; and there are situations, such as invocation of memory, or reasoning based on memory, or return to the first environment with acts of deference to its communal norms, that can bring these abilities into play.

Boghossian defends (3) by saying 'it is quite clear' that it is true. But it is not clear. In fact, if S has forgotten nothing, I see no reason to think that S will not know (in the relevant sense) at time t2 what he knew at t1. (As I will soon indicate, I think that the phrase "know what he thought" covers two different sorts of "knowing what".) S can at t2 remember his thinking at t1, and his memory can link the content of the earlier thought to that of the memory-induced one, by fixing the memory induced content as that of the remembered one. Merely being

[5] Boghossian, 'Content and Self-Knowledge', 22–23.

in the second environment, with concepts appropriate to that environment, does not prevent him from retaining and thinking thoughts appropriate to the first. Nor does it automatically prevent his retaining knowledge that he had before.

I will concentrate on cases where knowledge is activated through memory. I have maintained that the individual may not know whether yesterday he had an aluminum or twaluminum thought. He does not have discriminative knowledge of this form. But memory need not work by discrimination; it can work through preservation. The memory need not set out to identify or pick out an aluminum rather than a twaluminum thought, trying to find one by working through the obstacles set by the switches. Preservative memory normally retains the content and attitude commitments of earlier thinkings, through causal connections to the past thinkings. That is one of its functions – maintaining and preserving a point of view over time. It need not take a past thought as an object of investigation, in need of discrimination from other thoughts. Memory need not use the form "Yesterday I was thinking a _____ type of thought", where the memory attempts to *identify* the thought content as an object. Again, if it did, the individual might perhaps err by using a thought appropriate to the second environment in making an attribution to a thought event in the first environment.

The memory need not be *about* a past event or content at all. It can simply link the past thought to the present, by preserving it. Such cases involve a particular type and function of memory—preservative memory—which preserves propositional contents and attitudes toward them, rather than *referring* to objects, attitudes, contents, images, or events. The memory content is fixed by the content of the thinking that it recalls. Similarly, the "referent" of the past tense in the memory is fixed not by an independent identification of the past event, but through the memory connection to the event itself. The individual reasons from the past thought, takes it up again, without the memory's taking it or anything else as an object (as, by contrast, the memory does in substantive memory).[6]

There is a broad but qualified analogy between preservative memory and certain aspects of pronomial back-reference. The analogy must be used with caution. I do not model preservative memory on pronomial back-reference. I believe that preservative memory is more basic (both ontogenetically and in explanations of epistemology and rationality) than anaphora in language. In fact, it seems to me that a linguistic theory of anaphora has to be able to account for anaphora supported by preservative memory. Still, the analogy may be helpful in the respects in which it holds.

[6] For discussion of the distinction between the different types of memory, see Tyler Burge, 'Content Preservation', *The Philosophical Review* 102 (1993), 457–488. A psychological analog of my distinction can be found in E. Tulving, 'Episodic and Semantic Memory', in E. Tulving and W. Donaldson (eds.), *Organization of Memory* (New York: Academic Press, 1972), 382–402; and *Elements of Episodic Memory* (Oxford: Oxford University Press, 1983).

In using pronouns, the speaker need not be able to identify the referent of a pronoun, or even its antecedent, in order to secure the antecedent. The speaker might get the antecedent or its referent wrong if he were asked to identify it independently of the pronoun. For to secure an antecedent for the pronoun, it is enough for him to rely on chains inherent in the discourse. The causal chains in preservative memory do a similar job in connecting later thoughts, including later self-attributions, to earlier ones. The same faculty is fundamental to ordinary reasoning, which preserves previous steps of reasoning to make the coherence of reasoning possible. Given appropriate reliance upon preservative memory, and given the existence of causal memory chains back to the states which carried intentional content, preservative memory takes up the "antecedent" content automatically, without having to identify it. As with anaphora, the thinker need not be able to identify the antecedent, much less its referent. He may rely on the mechanisms of memory to do the job.

In the case of anaphora the interpretation or referent of the pronoun is not always the same as that of the antecedent. Anaphora is a syntactic device, whose semantic interpretations may vary, depending on the type of anaphora and the linguistic context of the pronoun. To this degree, anaphora and preservative memory differ. Preservative memory is not primarily a syntactic matter. It is a preservation of content, fundamental to the coherence of rational activity. In the memory case, the content and referent of the remembered material is not distinct from that of the antecedent thought content, which in ordinary that-clause-type self-attributions is both thought and referred to. The point of preservative memory is to fix the content in present mental acts or states as the same as the content of those past ones that are connected by causal-memory chains to the present ones.[7] If the individual relies primarily upon preservative memory, and if the causal-memory chains are intact, the individual's self-attribution is a reactivation of the content of the past one, held in place by a causal memory chain linking present to past attributions.

Memory could preserve the content of a past thought in either of two ways. The individual could remember the past thinking as an event; and only the content of the thinking could be remembered in the preservative way. His

[7] With important qualifications, there is an analogy to pronouns of laziness, pronouns which can be expanded into rewrites of the antecedent. Just as pronouns of laziness are in effect exact reproductions of their syntactic antecedents, preservative memory produces an exact reproduction of its content "antecedent". The difference between syntax and semantics is, however, crucial. Perhaps in thinking about this analogy, it would help to imagine that the antecedent of the pronoun of laziness is a term that is both used and self-referentially mentioned. So the prounoun as rewritten must do the same. *Ordinary* pronouns of laziness can, of course, carry different referents from their antecedents, because of contextual or scope elements of the discourse in which the pronoun is embedded. The function of preservative memory is to preserve content, which is by no means the function of pronouns of laziness. Pronouns are syntactic devices with a variety of semantic interpretations, whereas preservative memory is a feature of thought. In fact, preservative memory is typically a reactivation of earlier material not a pronomial shorthand for it. Purely preservative memory can be expressed in language, of course. But it is not a linguistic device.

memory would then tie current conceptual use to the concepts of that past thinking. Here substantive event memory and preservative memory would work together—the former identifying an event, the latter preserving the content and attitude-modality of the event. Or, second, memory could just preserve and bring up the attitudinal-commitment and the content of the earlier thought, making use of a file that tracks tenses and indexicals within it, without referring to the past thinking event. Here the memory is not referentially *of* the earlier event. It simply carries forward the content and force of the earlier thought for later use. In neither case need the individual *identify* the *content*. Memory functions to allow him to employ it again. It preserves the content regardless of what the individual thinks about it or knows about the world.

When activated, the remembered thought content will no longer be associated with a cognitive state that is indexed with present tense. And any new thought occurrences produced by the memory will not themselves be self-verifying. Thus instead of *I am hereby thinking that aluminum is a light metal*, one would be remembering what can be approximately expressed as *I was thereby thinking that aluminum is a light metal*. But—this is important—the "was" and "thereby" have the special preservative character involved in preservative memory. They relate to elements in the original thought preservatively rather than referentially. There is no independent reference, from the time-perspective of the present memory, to the time of the past thinking. Nor is there independent reference to the act originally expressed in "hereby". The preserved thought content will be preservatively linked to the tense and self-verification of the original thought event.[8]

Despite these differences between the memory and the remembered thought, I have no objections to thinking of the thought content of the memory as being the same content as that of the original self-attribution. At any rate, the memory can remember the self-verification and reflexivity of the original thought. The linkages are made by the cognitive system to the original applications of concepts. They do not depend for their operation—or, I will argue, for their being epistemically warranted—on any new cognitive relations to the environment. The causal-preservative linkages simply allow for redeployment of old concepts.[9]

[8] There is some analogy between these anaphorical uses of tense and Castenada's "I*". Compare H. N. Castaneda, 'Indicators and Quasi-Indicators', *American Philosophical Quarterly* 4 (1967), 85–100, and Castaneda, 'On the Logic of Attributions of Self-Knowledge to Others', *The Journal of Philosophy* 65 (1968), 439–456. Thus although the preserved thought content no longer picks out the present in its tense, it keeps track of the present-tense character as well as the time of the original thought. In this respect "was" is misleading. Similarly for "thereby" and "hereby". The self-verification of the original thought is preserved, even though the remembering may not itself be self-verifying.

[9] Boghossian appears to place weight on the supposed later unavailability of the self-verification of a self-verifying thought. As noted (note 8), I think the past self-verifying thought is available through preservative memory, even though the activated memory is not itself self-verifying. But self-verification is not a necessary feature of our non-empirical entitlement to self-knowledge. I used *cogito* cases primarily for expositional purposes in the original article; I think that they carry clues to understanding the wider array of non-empirically known self-attributions. In any case, the respects in which self-verification is not re-enacted in preservative memory are not respects in which the

Boghossian writes:

> Now, let us ask: *why* does S not know today whether yesterday's thought was an [*aluminum*] thought or a [*twaluminum*] thought? The platitude insists that there are only two possible explanations: either S has forgotten or he never knew . . .

Boghossian is asking the wrong question here. So doing yields a misleading application of the "platitude". In preserving knowledge, S (or S's memory) need not be in the third-person position of solving the problem of whether yesterday's knowledge had one content rather than another. That would be to take the past thought as an object of identification. The knowledge to be preserved did not have the form that Boghossian's question implies, something like: *I am thinking that aluminum is a light metal; and the thought just thought is an aluminum thought as distinct from a twaluminum thought*. S did not think a thought that looked on a content from the outside and opened itself to questions of comparative discrimination.

The right question is whether knowledge first expressed by *I am thinking that aluminum is a light metal* can be preserved. I have always maintained that thinking back to yesterday, S might be unable to discriminate the aluminum thought from a twaluminum thought. My view has been that S need not make such discriminations, except insofar as they are made by what he actually thinks—and by what is preserved in memory from those thoughts. If the individual tries to *identify* in memory a past thought as an object of investigation, of course, he may misattribute thoughts appropriate to the second environment to uses in the first.

The form of the question is important. The question to be asked on behalf of the individual's memory is not whether the original thought contained a concept of aluminum (or twaluminum). Such a question takes past thoughts as objects of investigation. The question is whether the original thought or knowledge can be preserved. To connect with my position, Boghossian should have asked why S does not know today that he thought[10] that aluminum is a light metal. I think

environment bears on the individual's concepts. Thus the issue of self-verification is irrelevant to the nature of the individual's concept *aluminum*. (Compare note 4.) That concept is redeployed in preservative memory.

In switching cases, especially if there are multiple switches back and forth, the individual's memory may have difficulty separating the causal-anaphoric files from one another. Many past individual thinkings may be lost to memory for all practical purposes. But insofar as one can remember an individual thinking event through preservative memory, one is in a position to retain the knowledge. Moreover, the beliefs (and other standing attitudes) associated with the thinkings are there to be accessed, and will cause fewer practical problems than remembering individual thinkings in the overgrown past. Here again I am assuming that the individual has not lost the concepts appropriate to the original environment. If the individual has lost those concepts, or if there are problems accessing past thoughts, we cannot assume that the individual has forgotten nothing.

[10] The past tense here refers back, anaphorically, to the time of yesterday's thought. But "yesterday" is not specified as such in the content of the preservative memory. If it were, the individual would go beyond preservation to identification, and would be vulnerable to an error of identification.

that S *does* know this insofar as the knowledge derives from preservative memory rather than from a third-person perspective on his past thoughts. If he tried to access the knowledge in the ways Boghossian's question suggests, he might fail to know by failing to identify the right thought. He might lack any special entitlement to his conclusions. When the individual thinks *I thought that aluminum is a light metal*, where a relevant thought is preserved from the original environment, preservative memory, working properly, automatically links the present thought with the remembered thought, with its aluminum concept, in the original environment.[11] It does so by relying on causal-preservative relations to the past thought event.

The differences in the form of the question about the individual's knowledge of his past correspond to two ways of understanding the question of whether the individual "knows what he was thinking yesterday". If one has identification in mind, perhaps it is natural to infer that in the slow switching cases, the individual does not know what he was thinking yesterday, since he cannot discriminate between two seemingly relevant possibilities. One is inclined to think that at least in many relevant switching cases, the existence of alternative contents accessible to the individual will make the alternative contents relevant to whether the individual knows, in the sense of "knows how to pick out", what—that is, which—thought he was thinking.

But identification is not at issue in preservation of self-knowledge, as distinct from third-person identificatory thinking back on it. If the "what" indicates preservation, then where the individual does not forget past thoughts, he will continue to know what he thought. He will continue to know the same content that he knowledgeably thought before.

Is the individual epistemically entitled to the products of preservative memory? One is, I think, entitled to rely upon such products, as long as it is in fact working properly, except perhaps in certain cases where one has reason to think one's memory is slipping and not maintaining the causal memory chains. To be entitled to rely on such memory, one need not supplement it with discriminatory identifications. One need not be able to defend it against potentially confusing challenges that would require one to distinguish the belief one actually calls up from beliefs that are similar to it. In fact, I think that as long as the causal memory links are in place, preservative memory is authoritative in something like the way much immediate present-tense self-knowledge is.[12]

[11] It is easy to confuse the situation being discussed with situations in which the individual knows about the switches but does not know when they occurred. This is easy because we who are thinking about the case know about the switches. Where the individual knows about the switches, there are special opportunities for confusion for that individual. He may confusedly despair of relying on memory, for example. But I think that where he does rely upon preservative memory, he will remain epistemically warranted. Since the argument I am considering does not rest on examples in which the individual knows about the switches, I have not gone into them.

[12] I think that since we are not dealing with identification, issues about relevant alternatives do not arise in anything like same way that they arise with perception or other forms of identification. As long as preservative memory is working properly, there is no possibility of error. There are no issues about "look alikes", since the memory is not "looking". Several issues here need detailed discussion, but since the argument we are considering does not raise them, I will not go into them on this occasion.

Perhaps Boghossian thinks that one is not entitled to rely on memory because of one's inability to distinguish original thoughts from the new twin thoughts, or inability to distinguish a thought actually remembered from a twin thought. If so, he needs to argue for this view. For assuming it in effect begs the question. I began with the claim that one need not be able to distinguish aluminum thoughts from twaluminum thoughts to have knowledge in certain self-ascriptions, such as the *cogito*-like judgment of *I am hereby thinking that aluminum is a light metal*. This same view applies to memory, at least purely preservative memory—which works by simply preserving thoughts already thought and making them available for reactivation and reasoning.

Preservative memory is necessary to any reasoning that takes place over time, hence any reasoning. We are as fundamentally entitled to rely upon it as we are entitled to rely upon reasoning. In fact, if we were not entitled to rely upon preservative memory, we would not be entitled to rely upon reasoning.

In the case of preservative memory, as in the case of direct self-knowledge, entitlement depends only on ordinary understanding and on the normal working of one's cognitive faculties. In neither case is discursive defense needed to safeguard the entitlement.

Consider the role of preservative memory in deductive reasoning. Such memory is needed just to carry the argument along over time. Such memory does not constitute or enhance the justificational force of the individual's justification for believing a deduced theorem. But the individual must be entitled to rely on preservative memory to be entitled to rely upon the deductive reasoning. One is entitled to rely on one's memory in such reasoning if it is working properly—if the thoughts are preserved in the course of the reasoning. Even if there are alternative thoughts that one cannot distinguish from those one is in fact thinking, one is entitled to the reasoning one is actually carrying out, as long as the reasoning is understood and deductively sound. And one is entitled to rely on the preservative memory on which the reasoning depends—as long as it preserves the thoughts thought earlier in the argument. It does so not by identifying the earlier thoughts or discriminating them from similar thoughts, but simply by preserving them for later employment.

Similarly, the individual may not be very good at distinguishing the content of past events, taken as objects for identification. But purely preservative memory of the past contents remains; and that is sufficient for knowledge through memory. Where self-attributions rely upon preservative memory that is working properly, the content will be knowledgeably preserved. Entitlement to memory-dependent thought or reasoning depends on the proper working of preservative memory, not on the individual's ability to specify what the memory is doing or retrieving, or on the individual's checking to verify that the memory is in good working order.

Considerations about preservative memory help undermine another of Boghossian's criticisms of anti-individualism. This criticism concerns reasoning, not self-knowledge. Boghossian considers some switching cases in which a

person has two concepts expressed by the same word-form.[13] He claims that externalism (which we shall not distinguish from anti-individualism) opens an unattractive possibility of undetectable equivocations in reasoning. He draws the stronger conclusion that "externalism undermines our ability to tell apriori whether any particular inference of ours satisfies one of [the forms of valid inference]".[14]

Boghossian seems to assume that apriori warrants must be "internally detectable". He seems to assume further that it follows from a certain argument's being vulnerable to possible undetectable equivocation that we can never tell apriori, even in normal cases, that they are justified. I think that both assumptions are

[13] Paul Boghossian, 'Externalism and Inference', *Philosophical Issues* 2 (1992), 11–28; Boghossian, 'Reply to Schiffer', *Philosophical Issues* 2 (1992), 39–42. Boghossian claims ('Reply', p. 39) that it has been widely overlooked that according to anti-individualism (or "externalism") a thinker might have Earthly and Twin Earthly contents cohabit in his system without there being any internal indication to him that this is so. He even claims that "no one seems to have noticed" the point ('Externalism and Inference', p. 17). In fact this was the central construal of the switching case I explain in 'Individualism and Self-Knowledge (p. 652; 58 in this volume). Much of the interest and challenge of the issues about self-knowledge derive from the assumption of cohabitation. In setting up the thought experiment, I wrote about acquiring new concepts in the new situation, and said nothing about losing old ones. Displacement was never part of the switching cases, at least in my understanding of them. Cohabitation was always the assumed case. I did not and do not consider the displacement model (as a general model for switching cases) a plausible account. I did not discuss and criticize the displacement model in 'Individualism and Self Knowledge', but largely because I thought it clearly implausible. I did consider it, and I included elements in my account that were meant to suggest its implausibility. Displacement raises obvious problems about memory in many standard cases. And it has no basis, as far as I know, in our ordinary understanding of how concept acquisition works in ordinary "moving" cases, especially among relatively mature language users with good memories and recurrent uses for the old concepts. Merely being switched stealthily from one place to another and gaining new concepts will not in general cause one to lose conceptual abilities. Part of the idea behind my postulation of switching was to indicate that the individual would retain a prospective use for direct applications of the old concepts (in addition to being able, I think, to remember objects, events, and thoughts from the departed environment). Concepts mark abilities; *just* moving around and acquiring new concepts will not in general obliterate such abilities, especially given that one still has uses for the old concepts and a perfectly good memory.

I wrote 'Now suppose that, after decades of such switches, one is told about them and asked to identify when the switches take place. The idea is that one could not, by making comparisons, pick out the twin periods from the "home" periods.' ('Individualism and Self-Knowledge', p. 653; 58 in this volume). One person has thought that only if the concepts are displacing each other will the question *what concepts a person has* track when the switches took place. My supposition was not as fully characterized as it should have been. But this interpretation misses my reasoning. What I think tracks when the switches take place is not what concepts the individual has at given times, but what uses the person makes of the concepts. I assumed that one would mostly use earth concepts on earth and mostly use twin earth concepts on twin earth—once one had acquired each set. So if one knew when the various thinkings took place and if one had some internal sign (which I think we in fact lack) that would enable one to *compare* and distinguish the earth and twin-earth thinkings (concept-uses) in retrospect, one could tell (more or less) when the switches took place. Although there may be some uses on twin-earth of earth concepts, they would not normally be dominant. If one could distinguish in memory ("by comparison") between uses on twin-earth that were governed by memory of earthean uses, on one hand, and twin-earth uses that were more tied to the immediate circumstances (and similarly for uses on earth), one could sort out the differences of location *through* memories of those uses even more exactly. Of course, I think that we cannot do these things. The relevant memory works through preservation, not comparison.

[14] Boghossian, 'Externalism and Inference', 22n.

mistaken.[15] But my discussion here challenges even the view that according to anti-individualism we must be subject to possible undetectable equivocations producing invalid reasoning, because of the possibility of switching cases.

An example of Boghossian's argument adapted to our case goes as follows. Assume anti-individualism to be true. Then Alice might remember an event of picking up, and feeling the light weight of, some aluminum on earth, before she was switched; and in remembering the event, she might think correctly that she picked up some aluminum at that time. Then remembering a sample of twaluminum on twin earth that she saw yesterday, she might think that yesterday there was some twaluminum beside her. (Both of these are uses of substantive, object-oriented memory.) She might reason from these premises, fallaciously, to the conclusion that she once picked up the same sort of thing that was beside her yesterday. The word form "aluminum" undetectably expresses for Alice two different concepts. The concepts used in the reasoning are supposed to be different because whereas the earth concept is evoked by the memory of the long past event of picking up aluminum, the twin earth concept is supposed to be evoked by the memory of what is in fact twaluminum on twin earth. The inference appears to Alice to be valid; but because of the switch in concepts, it is invalid.

The weakness of examples of this sort is that they overlook the centrality of preservative memory in reasoning, and fail to note the particular character of the relation between the different concepts within the individual's cognitive system. As Boghossian insists, there is little ground to think that Alice made a mistake in reasoning. But contrary to his claims, there is no necessity on anti-individualistic grounds to attribute one to her.[16]

Alice's argument is carried out in thought. I think it natural to agree with Boghossian's account of the first premise. In determining what Alice is thinking in the second premise, one must remember that both the substantive memory of yesterday's experience, and the preservative memory of the use of the concept in the first premise, are operating in Alice's thinking. What it is to carry out valid arguments in thought is to connect premises, holding them together in a way that

[15] For a discussion of apriority that entails the falsity of both assumptions, see Burge, 'Content Preservation'.

[16] Although I do not agree with all of his remarks, Stephen Schiffer seems to me to make essentially the right points in his reply (Schiffer, 'Boghossian on Externalism and Inference', *Philosophical Issues* 2 (1992), 29–38). I simply bring out a generalization of Schiffer's points, and invoke the notion of preservative memory. Boghossian claims regarding the second premise of an analog of the example about Alice that it is "entirely independent" of the first premise. But in actual reasoning we typically tie key terms in premises together through preservative memory, as Schiffer in effect points out. Schiffer also shows that various rephrasings of arguments, using relative clauses instead of separate premises, would elicit the non-independence of relevant components of the two premises, given the reasoner's intentions. One might make the same sort of point with respect to other examples by appeal to pronouns. These tyings together are particularly strong in these cases, in view of the fact that the reasoner would, if the opportunity arose, identify the same objects using either concept. But even if an individual does not tie premises together in this way, Boghossian's argument fails, as I shall show.

supports the conclusion. And preservative memory—even in short arguments that we idealize as occurring in a specious present—is essential to this enterprise. Insofar as we think intuitively that Alice is not making a mistake in reasoning, it is natural, and in most cases I think correct, to take her to be holding constant, through preservative memory within the argument, the concept used in the first premise in her thinking the second premise. The role of the concept aluminum in the reasoning is primary in her thinking, and preservative memory takes the occurrence of the concept in the first premise as a basis for its reuse in the second premise.

Anti-individualism does not say that every thought's content is fixed by the type of object that occasions the thought. Although free-standing memories normally evoke the concepts utilized in or appropriate to the remembered context, the exigencies of reasoning will often take precedence. One commonly utilizes concepts used earlier in an argument to identify objects in memories invoked in later steps.

Given this usage by Alice, the second premise is false: She is mistaken in applying her concept aluminum, preserved from the first premise and originally evoked by her experience on earth, to her experience yesterday on twin earth. She is using the concept obtained from the first premise to *identify* the metal she remembers seeing, as expressed in the second premise. The mistake is a mistake of memory identification, not one of reasoning. Variations on this point apply to all of Boghossian's examples.[17]

Perhaps there are cases where the reasoner does not tie the parts of an argument together in this way. That is, the reasoner's intended reasoning does not close the question of whether the concepts expressed by the same word-sound in an argument that is otherwise syntactically valid are the same. Insofar as the reasoner's intentions in reasoning are not dominant in requiring "anaphorically" that the same concept be used through the reasoning, and insofar as we think that there is a gap between the premises that the reasoner has not made explicit, it would seem obvious that the reasoner tacitly and mistakenly presupposes that the concepts apply to the same objects. This presupposition is not present in cases of equivocation that occasion invalid reasoning. In such cases the individual overlooks the difference between concepts expressed by the same term, but has no tendency to treat the concepts as interchangeable in general beliefs, and no tendency to apply them to the same objects. So to fully capture the reasoner's

[17] As I have noted if someone who has undergone switches relies upon memory to *identify* a past object or event—including a past thought—he is subject to error. But supposing that Alice thought yesterday that twaluminum is beside her, she is in a position, relying on *preservative* memory, to remember what she thought then. And if this memory were to generate reasoning, then it would normally be held constant through an argument, and it might generate misidentifications of other past thoughts in later premises. Then although the reasoner knows what those past thoughts were, and could call them up if she relied purely on preservative memory, the reasoner might make mistakes about those past thoughts through reliance on identificatory substantive memory. These cases bring out again that accessing past knowledge in switching cases involves more parameters than in normal cases.

cognitive state in a case where the reasoner does presuppose (mistakenly) that the concepts apply to the same objects, one would have to supply for the reasoner the mistaken presupposition that twaluminum is aluminum. Again, there is no mistake in reasoning, only a mistake in presupposition.

I am doubtful that there are any clear cases of invalid equivocation deriving from switching cases. But if there are, they are marginal. [PS, 2011: I have given up this doubt.] And one can avoid any such cases by firmly and intentionally relying upon preservative memory in maintaining the same concepts throughout the course of one's arguments. In view of the fact that switching cases leave one inclined to apply different concepts to the same objects, this intention will accord with both *de re* and general memories, beliefs, and desires involving the different concepts. Applications of concepts governed by connection to other premises may then express misidentifications, or other mistaken beliefs, that the switched individual is already inclined to.

Let us return to self-knowledge. So far I have focused criticism on step (3) of Boghossian's argument: The assumption that in slow switching situations, assuming anti-individualism, we cannot know at the later time, without empirical investigation, what was known at the earlier time. All of my discussion of self-knowledge has simply granted step (2)—that nothing is forgotten. I have granted that the individual has, without realizing it, both the original concept and a new concept after slow-switching. And I have assumed that this situation does not (in general) result in the individual's actually forgetting anything. So the original beliefs are not forgotten. At worst, they can fail to be accessed in certain situations.

But some cases may run contrary to the assumption that forgetting is irrelevant. It may be that in certain Disjoint Type cases, in which an individual does not reactivate the concepts from the original environment (say, if he forgets specific events from the earlier environment), the individual eventually loses the original concept. He loses the ability to think specifically about aluminum. I see no reason to think that in such cases the individual's knowledge of his thoughts before he lost the concept is threatened. The argument Boghossian gives will simply be short-circuited at a different point, step (2).

Let us finally turn to Amalgam Type cases. It will be recalled that these are cases where, in addition to having the original concept, the individual acquires a jade-like concept that applies equally to aluminum and twaluminum. So assume that S has a broadened concept which includes the extension of the original one. If S has not lost the original concept, then the response I gave for Disjoint Type cases reapplies. I see no reason to think that S will ordinarily lose the original concept, even in Amalgam Type cases. Normally S will be able to access through memory the old conceptual ability to think about aluminum, about past-aluminum experiences, or about past aluminum thoughts. But if S does lose the original concept, then a different reply is appropriate. If losing a concept is a form of forgetting, then premise (2) of the argument is mistaken. Clearly if one loses a concept when it is replaced by a new one, and for that reason one has no access to

beliefs one once had, one may lose knowledge one once had.[18] Whether (1), (2), or (3) is at fault, I see no reason here to doubt that one's authoritative, non-empirically warranted self-knowledge is compatible with the truth of anti-individualism.[19]

[18] There is reason to think that step (1) is false in any case. There appear to be cases in which one knows something at one time but loses the knowledge at a later time, not because one forgets anything, but because one's original warrant is (misleadingly but reasonably) outweighed by prima facie defeating alternatives that emerge at a later time. Suppose that one has true warranted belief amounting to knowledge at a given time. For example, suppose one sees a lectern and thereby knows that the lectern is in the room. Suppose at some later time one acquires good reason to believe that one's apparent experience of the lectern was or may well have been the product of an illusion-causing hologram. Then even though one had in fact seen the lectern, one could at the later time perhaps lose the knowledge that the lectern was in the room, despite the fact that one forgets nothing. Such cases might seem to be relevant to some of the switching cases at hand. One might think that after the switching, one's original warrant is undermined by relevant alternatives, even though one's original self-knowledge was in place. So one could be prevented from having knowledge of one's past (originally known) thoughts, even though one forgets nothing.

I will not discuss here whether there are any such cases in which step (1) fails. I am not inclined to press this point because I believe that in neither the present-tense self-knowledge case nor the case where preservative memory preserves self-knowledge, is the relevant-alternatives epistemology applicable. And whenever preservative memory is relied upon, the original self-knowledge can be retained regardless of what switching has occurred. I am inclined to think that in certain basic epistemic functions, knowledge is immune to threat from alternatives as long as the basic processes are working properly (and perhaps additionally, those processes' proper working is not called specifically and reasonably into question), and as long as those processes are not subverted by irrational interferences from within the cognitive system itself. Authoritative self-knowledge and preservative memory are among these functions. I have not defended this view in appropriate generality. But it has informed my picture of self-knowledge from the beginning, and has so far not been explicitly attacked head-on. In fact, as I have noted, Boghossian's criticism in effect acknowledges the view and tries to defeat my position without questioning it.

[19] Thanks to Glenn Branch and Peter Ludlow for comments on a draft, and to Youichi Matsusaka whose seminar paper on this subject helped clarify a key point.