

Internalism Entails That Concepts Are Idiosyncratic

Cressida Gaukroger

DRAFT. DO NOT CITE WITHOUT PERMISSION.

March 2017

Abstract

Internalism about mental content states that only those elements internal to the mind can contribute to the content of thought. While accounts of internalism vary, the internalist agrees that two individuals who are molecule-for-molecule identical will have the same mental content, regardless of differences in their environments or histories. This paper argues that, if one accepts the internalist premise, then they must also accept that concepts must be individuated internally, for to do otherwise is inconsistent with the reasons that motivate the internalist position in the first place. However, internal individuation of concepts entails that they will be idiosyncratic – no two people who are not internally identical can have the same concept.

Internalism about mental content has recently been experiencing a quiet resurgence. Among others, Gabriel Segal (2000), Katalin Farkas (2008), Joseph Mendola (2008), and Nicholas Georgalis (2015) have revived the thesis that mental content supervenes on internal, or intrinsic, properties. The internalist position is traditionally both framed in terms of, and criticised in relation to, how it individuates concepts. While internalism is primarily built on the argument that two internally identical individuals must have the same concepts, internalists often also attempt to individuate concepts so that two internally *different* individuals could nonetheless possess the same concept. I will argue

here that the reasons for adopting mental content internalism necessitate that concepts *cannot* be shared between two internally different individuals. This means that if the internalist is correct in their account of mental content, concepts must be idiosyncratic.

The charge of entailing that concepts are idiosyncratic is not limited to externalist criticisms of internalist theories. Many, such as Segal (2000), have laid this charge against the full range of theories of concepts. Functionalism which defines concepts in terms of the behaviour they cause - such that two concepts tokens are of the same concept type if they cause the same behaviour - is unable to give a robust account of behaviour individuation that is itself not idiosyncratic. Conceptual role semantics, on which concepts are individuated relative to the role they play in a mental network - their relationship to other concepts - faces holism, the consequence that for two people to share a concept they must share every other concept they possess. Holism is also a problem for theory theory which argues that concepts are theories, themselves constructed out of more concepts. The problem of not being able to individuate concepts broadly enough such that they generalise across people - that two non-identical people could never share a concept - is a problem at the heart of the debate over what concepts are, and it is one, I will argue, that internalists are simply unable to solve.

A quick note before I begin: being an internalist about mental content does not necessarily entail being an internalist about concepts. However, being the former without being the latter means accepting an extreme form of externalism about concepts. Georges Rey (1985, 1999), for example, understands concepts as abstract entities without mental components. In theory, this Fregean or Platonic account of concepts is compatible with internalism about mental content – as it states that conceptual content is not mental content, since concepts themselves are non-mental. Therefore, the internalist about mental content could consistently hold that concepts, as non-mental entities, are not internal.

However, to be an internalist about mental content, but an externalist about con-

cepts raises the question: why be an internalist in the first place? The central externalist cases that contemporary internalists discuss – Hilary Putnam’s (1975) Twin Earth and beach/elm hypotheticals, and Tyler Burge’s (1979) arthritis case, among others – are specifically focused on concepts: namely, the concepts ‘water’, ‘beech’, ‘elm’, ‘arthritis’. The point of such examples is not just to stir externalist intuitions. They also focus on the possession conditions for concepts, particularly where concept-possessors’ understanding is incomplete. What is being discussed in such cases is concepts (and concepts that are generally presented as having a mental component). So, if an internalist insists that they are talking about mental content, but not concepts or conceptual content, they are not addressing the main externalist arguments.

1. Internalism as a minimal constraint on concept individuation

One way of understanding internalism as it relates to concepts is that it sets a minimal constraint on concept individuation: at the very least, if two people are ‘internally identical’ then their concepts will be identical. Views about what it means for people to be internally identical, and which internal features, components or states are relevant to determining mental content, will differ depending on the kind of internalist theory one endorses.

According to Farkas (2003, 2008), mental content is nothing over and above how a thought appears to the bearer of that thought. Farkas’ account of internalism holds that the relevant quality that doppelgängers share in Twin Earth thought experiments is that their thoughts “would appear the same” (Farkas, 2003: 7). Farkas believes that this, the relevant ‘internal’ feature, can be captured by imagining the thoughts of Putnam’s (1975) Oscar and Twin-Oscar being indistinguishable *to them* – that is, to the individuals having those thoughts. She refers to this criterion for an internalist theory as ‘subjective indistinguishability’.

In contrast, Segal (2000) defines internalism in terms of intrinsic and relational properties. Externalism, he argues, is the theory that cognitive content is a relational property – the content of a mental state is a property that mental state has in virtue of the relationship between a mind and the world. Considered in terms of Putnam’s Twin Earth example, the externalist concludes from such a case that Oscar’s and Twin Oscar’s thoughts are different, not because of any intrinsic bodily differences between them, but because of the relationship they bear to their different environments. Internalism, which Segal defends, is the denial that mental content supervenes on anything but intrinsic properties.

The minimal constraints on concepts according to these positions, therefore, are:

F1) If two individuals have subjectively indistinguishable concepts, then they have *the same* concepts.

S1) If two individuals have intrinsically identical concepts, then they have *the same* concepts.

Assuming, for the sake of argument, that an internalist minimal constraint is initially appealing, the challenge for these theories is to provide an account of concept individuation that steps beyond these minimal constraints. What the internalist wants to avoid is committing themselves to one of the following:

F2) If two individuals have subjectively different concepts, then they have *different* concepts.

S2) If two individuals have intrinsically different concepts, then they have *different* concepts.

The problem with individuating concepts in the way set out in F2) and S2) is clear: internal ‘stuff’ always varies between individuals, and indeed in one individual over time. If two concepts have to be internally identical to count as the same concept, then, aside from the examples where internal indistinguishability is specified, such as it is in the Twin Earth scenario, it would almost never be the case that two people could possess the same concept, or even that one individual could have the same concept twice. The question is: does internalism necessitate that concepts must be individuated purely internally (i.e. that they be idiosyncratic)? To reject F2) and S2) the internalist must present a theory of how to individuate concepts such that they are the same when the relevant internal features are the same, but where concepts can also be the same, even when these internal features are different (even if only slightly).

For the internalist to attempt to avoid concluding that all concepts are idiosyncratic, there are three possible ways that they could approach concept individuation. The first is to use the minimal mechanism provided by their particular version of internalism to aid in the task of individuating concepts across persons. The second is to turn to something external, such as reference, to individuate concepts. The third is to identify a feature of concepts that is defined in terms of their narrow properties, but is itself more general than these properties. I will show below that none of these methods are both workable and consistent with the principles that motivate the internalist position.

2. Defining an ‘internalist’ maximal constraint on concept individuation

One way to determine when two concepts are different would be to use the same criteria one used for determining that two concepts are the same, i.e. the minimal criteria outlined in the previous section. There are reasons to want to do this – the reasons that motivated being an internalist in the first place. To individuate concepts using the

minimal criteria is to individuate them *internally*.

Consider Farkas' internalism. Farkas argues that two concepts that are subjectively indistinguishable are the same concept. However, while the principle of subjective indistinguishability can be used for concept individuation, such a principle would rule out cases of two people with subjectively *distinguishable* thoughts sharing the same concept. That is to say – if subjective indistinguishability is a necessary condition for concept sharing, then this would entail that concepts are idiosyncratic. If one wants to individuate concepts on the basis of subjective distinguishability alone then they will get F1), but they will also have to accept F2).

Alternatively, one might argue that subjective indistinguishability is merely a guide to subjective similarity. In the cases where two concepts are subjectively indistinguishable they are clearly similar enough to be the same concept, thereby keeping a commitment to F1). Furthermore, on this account, if Oscar and Twin-Oscar had very slightly different 'watery stuff' experiences, their concepts might be very slightly subjectively distinguishable. However, such concepts could be understood as being subjectively similar enough for them to count as the same concept. This would mean that F2) could be rejected.

The problem with this approach is that the fact of two concepts being subjectively similar cannot on its own account for the identity of any two concepts so long as they are in any way subjectively dissimilar. Consider the following example: you and I are aware that Polly and Penny are identical twins. When I think about Polly it is very subjectively similar to when you think about Polly. There are some differences – we had different experiences of, and interactions with Polly. For example, we have seen Polly in different circumstances, from different angles, wearing different clothing etc. However, our Polly-related thoughts have enough subjective similarities that we feel warranted in saying that we share a 'Polly' concept. Taking Po_M to be my Polly concept and Po_Y to be your Polly concept, it follows that:

a) $Po_M = Po_Y$

However, my thoughts about Polly are also very subjectively similar to your thoughts about Penny. Once again, there are some differences – we have seen Polly and Penny under different circumstances, from different angles, wearing different clothing – but there is as much subjective similarity between my Polly concept and your Penny concept as there is between my Polly concept and your Polly concept.¹ If subjective similarity (but not identity) were enough to say that my ‘Polly’ concept and your ‘Polly’ concept are the same, then it would also be enough to say that my ‘Polly’ concept and your ‘Penny’ concept are the same. With Pe_M as my Penny concept and Pe_Y as your Penny concept, it follows that:

b) $Po_M = Pe_Y$

But, presumably, my Penny concept is also going to be very similar to your Penny concept, meaning:

c) $Pe_M = Pe_Y$

Which entails:

d) $Po_M = Pe_M$

That my Polly and Penny concepts are the same concept is a troubling consequence. Similarity might not be transitive, but identity is, so if you define identity in terms of similarity then you will be faced with a slew of similarly problematic consequences.²

¹I have used the names ‘Polly’ and ‘Penny’ here to differentiate the concepts to make things clearer for the reader, but we could assume that the people we are considering do not know (or perhaps, remember) the names of the different women. This shouldn’t be thought of as a puzzle about names.

²Furthermore, for reasons repeatedly argued by Fodor and Lepore (1992), similarity-based accounts of concepts have a difficult time not importing circular identity-based mechanisms to make sense of what

Segal's account of internalism fares no better at providing an account of concept individuation that retains S1) while rejecting S2). By being committed to S1) Segal's theory entails that two concepts are always the same if their corresponding intrinsic properties are the same. To reject S2) Segal must show that two concepts can be the same even in cases where their corresponding intrinsic properties differ.

Segal's account faces the same problems as Farkas' – if the minimal conditions are the only conditions relevant to concept individuation, then one will end up with concepts being idiosyncratic because the intrinsic properties of mental states vary across individuals and within the same individual over time. For the internalist who accepts S1), something like a neural correlate, for example, may serve as both intrinsic and yet realisable across multiple individuals. However, it would be highly surprising if there were the same neural correlates for most or indeed all concepts across people. And, even if all concepts turned out to have neural correlates, to explain why they were neural correlates for the *same* concept would presumably require a non-internal or non-intrinsic explanation.

One might want to argue that, while concept tokens vary between individuals, intrinsic properties can be understood as types that can be instantiated in many individuals. The problem with this approach is that to group together intrinsic properties of mental states into types must be to group them together according to some characteristic that goes beyond mere intrinsic properties. One could do this by using some external mechanism for grouping together concept types (such as causal history, functional role, or extension/reference), but if one wanted to do it *internally* the only option would be to talk in terms of similarity of intrinsic states.

The easiest way for the internalist to provide an account of concept identity would be in purely internalistic terms, using the minimal mechanism provided for them by their respective definitions of internalism. However, to do so would be to end up with an

a robust account of concept similarity could be.

idiosyncratic account of concepts, as their minimal accounts do not provide the tools for defining concepts as types that extend beyond tokens. For a non-idiosyncratic account of concept individuation, therefore, the internalist must look to the external.

3. Reference as a maximal constraint on concept individuation?

If the internalist wants to resist the conclusion that concepts are idiosyncratic, one option would be to reject an exclusively internal model for individuating concepts. This can be done by combining the internal minimal constraint with an external (or relational) maximal constraint on concept individuation. A maximal constraint would be one that defines what it takes for two non-internally-identical concepts to nonetheless count as the same. Any maximal constraint will be better if it is not too narrow – so it does not carve up concepts such that almost no one can share them – and not too broad – it should not carve up concepts such that two people with obviously different concepts are determined to have the same concept. To be compatible with internalism, any maximal constraint on concept individuation must be compatible with three additional basic conditions:

The Minimal Preservation Condition (MPC): Any account of concept individuation must preserve identity of concepts between physical replicas.

The Sharing Restriction Condition (SRC): Any account of concept individuation must not entail that two people have the same concept where this would undermine *MPC*.

The Consistency Condition (CC): Any account of concept individuation must be consistent with the broader goals of internalism.

MPC is necessary given the primary specification that defines the internalist position.

SRC requires that accounts of concept individuation do not entail that those people who have *very* internally-different concepts share those concepts. It must be the case, for example, that non-experts do not always share concepts with experts. Why? Because in all the hypothetical doppelgänger scenarios there are experts with different concepts from each other: e.g. future experts on Earth and Twin Earth; current experts in Burge's (1979) arthritis and 'tharthritis' worlds. If the doppelgängers share concepts with the experts in their respective worlds, and experts across worlds do not share concepts, then the doppelgängers cannot share concepts with each other.

CC is needed to respect the arguments that motivate the internalist beyond intuitions about doppelgänger scenarios. The internalist position is motivated by a range of arguments including the ability to account for concepts of non-existent or impossible entities; the ability to account for Frege Puzzles; the possibility of self-knowledge; and the causal-inefficacy of relational properties. Were a maximal constraint on concept individuation to be inconsistent with the internalist position on these issues, then accepting it would undermine the reasons for being an internalist in the first place.

Providing an external account of concept individuation that respects the above conditions, however, is difficult. Consider reference as a way of individuating concepts. To argue that concepts are ontologically narrow does not entail commitment to the position that concepts do not refer. Could reference as a constraint on concept individuation be made compatible with internalism? There are several serious problems that would face anyone who attempted make it so.

One problem with reference serving the maximal constraint role is finding a theory of reference that preserves *MPC* and *SRC*. It is open to the internalist to argue for internalism about *mental content*, while accepting that we may need a different account of reference. Reference itself is a relation that holds between two things – mind and world – and it is not inconsistent to understand the mind part of that relation purely in

terms of its internal (phenomenal/intrinsic) properties. However, if the internalist is to individuate concepts in accordance with reference then they cannot understand reference itself as something outside the remit of an internalist account of mental content. To be consistent with *MPC*, the internalist must reject the idea that it is possible to have the same concept with different referents, if they want to individuate concepts by their referents. Consider the Oscar/Twin-Oscar case. The internalist is committed to the position that Oscar and Twin-Oscar have the same watery-stuff-related concepts in virtue of their mental states being subjectively indistinguishable or intrinsically identical. If the internalist concludes that, while Oscar and Twin-Oscar have indistinguishable thoughts, their concepts have different referents, then reference cannot be used to individuate concepts if *MPC* is to be retained. What the internalist needs is a robust account of reference where, if two concepts are subjectively indistinguishable or intrinsically identical, they *always* refer to the same thing. As such, *MPC* precludes the internalist using any of the traditional or currently dominant theories of reference in individuating concepts.

However, the more troubling problem is that individuating concepts in terms of reference risks violating *CC*. Any attempt to individuate concepts relationally is going to face problems with concepts that do not refer to anything that exists. It is possible to have concepts associated with unicorns, phlogiston, and Thor, and yet no concepts could refer to such animals, substances or individuals, as they do not exist. As these concepts all share a referent - nothing - if they are individuated on the basis of reference, this would entail that they are all the same concept. Furthermore, individuating concepts on the basis of reference also falls foul of the problems illustrated by Frege Puzzles. Any account of concepts that doesn't allow two individuals to have different concepts that have the same referent fails to individuate concepts in a way that can handle Frege Puzzles. Importantly, while such problems have been raised with externalist theories of concept individuation, the internal individuation of concepts was a way of avoiding them. One of the great virtues of internalism is that, if concepts are internal, then it is obvious how

‘unicorn’ concepts and ‘phlogiston’ concepts are different, and how ‘hesperus’ concepts and ‘phosphorus’ concepts are different: they differ internally. To individuate concepts so as to conclude otherwise would, therefore, be to violate *CC*.

4. Concept individuation on the basis of extension conditions

There is a final alternative for the internalist: trying to find an internally-consistent way of generalising across concepts. Segal (2000) proposes such an approach, arguing that internally-defined concepts should be individuated in terms of their extension conditions. Extension conditions will have to be defined in a non-standard way if they are to be compatible with internalism. However, this approach appears initially appealing – extension conditions need not be as coarse-grained or context-specific as reference, nor as fine-grained or idiosyncratic as the narrow properties of concepts. Given the right account, this could save internalism from entailing that concepts are idiosyncratic.

Whatever account of extension conditions the internalist chooses to use in individuating concepts, the first condition it must uphold is *MPC*. In other words it must be the case that, in every instance where two concepts are internally identical, they have identical extension conditions. As Segal (2000:28) argues, this means that extension conditions will be determined depending on the internal properties of any concept: “[E]xtension conditions are themselves narrow... It means that the extension conditions of a thinker’s concepts are determined by intrinsic features of the thinker.”³

³I am going to treat Segal’s position as stating that extension conditions, in determining conceptual content, are the basis for concept individuation: when extension conditions are the same, concepts are the same, and when they are different, concepts are different. At points, however, Segal appears not to hold this exact position: “...we need to think of concepts as organic entities that can persist through changes of extension. Alf takes it that after correction he still deploys the same concept he had earlier. In a sense he is perfectly correct. It is the same concept in the sense that it is the same organic unity that has survived the conversation with the doctor. However, it has undergone a change of cognitive content and even of extension conditions.” (Segal, 2000:77) If this is to be understood as demonstrating that concepts are not, in fact, to be individuated on the basis of extension conditions, but rather individuated on the basis of their intrinsic (in this case physical or organic) properties, then the problem of individuating concepts purely internally returns, as discussed in section 2. It is for this reason that in this section

4.1 *Extension Conditions Determined by Reference*

On Segal's (2000, 2004) account the extension conditions for a concept pick out a set that extends across possible worlds (such as the set that contains all possible watery-stuff). On such an account (which is the account Segal endorses), Oscar and Twin-Oscar would (at the very least) have *both* H₂O and XYZ in their extensions.⁴ In this regard, Segal identifies an account of extension conditions that upholds *MPC*. This account also has the capacity to maintain *SRC*. In the Twin Earth scenario we want to say that future experts on Earth and Twin Earth do not share concepts, but Oscar and Twin-Oscar do share concepts. All that the extension conditions method of concept individuation needs to accommodate this is to make sure that neither Oscar nor Twin-Oscar share concepts with the future experts on their respective worlds. The future-Earth expert's concept's extension conditions only pick out H₂O in all possible worlds because being H₂O is part of her water concept. Oscar's concept's extension conditions, in contrast, pick out much more than H₂O. As different extension conditions mean different concepts, *SRC* is maintained.

Similarly, in Burge's (1979) arthritis case, Alf's concept picks out ailments of the joints and ailments of the thigh, so he does not share his concept with the experts in his world whose concepts extend only to ailments of the joints. In contrast, as the experts in Twin-Alf's world do have a concept which extends to ailments of the thigh as well as ailments of the joints in just the way that Twin-Alf's and, indeed, Alf's concepts do, Alf, Twin-Alf and the Twin-experts all share a 'tharthritis' concept. Once again, this account is able to avoid violating *SRC*.

When it comes to *CC*, however, Segal's account of extension conditions faces several

I will consider only the argument that concepts should be individuated in accordance with extension conditions, using Segal as a guide to what such a position would look like. Additionally, I will take as my target a position that is exclusively internalist, avoiding discussion of the fact Segal (2004) actually argues that the extension conditions for concepts which refer to individuals are externalist.

⁴Segal (2000:19): "[M]y view is that both Zowies' diamond concepts apply to both diamonds and twin diamonds, so contrary to what some might initially think, if Zowie pointed to a twin diamond and said "That's a diamond," she would be saying something true in her idiolect."

difficulties. Such an account should have no bearing on the arguments regarding the causal inefficacy of relational content, nor undermine the principle that we can know the content of our own concepts. However, the account Segal presents will not have any internalist advantage in explaining Frege Puzzles, as he argues that the extension conditions for ‘single’ concepts (those that refer to individuals as opposed to sets, kinds, or categories) are externalist. In this regard, his particular account of extension conditions violates *CC* insofar as it individuates ‘single’ concepts too broadly to be able to account for how people could have different concepts that nonetheless refer to the same individual. Furthermore, if extension conditions supervene on actual extension (even cross-world extension) then the internalist loses the ease they had in explaining concepts with empty referents (or empty extensions). One argument in favour of adopting internalism was that it is good at handling concepts of non-existent entities. If part of my ‘Father Christmas’ concept is that he doesn’t exist, and part of my ‘unicorn’ concept is that unicorns don’t exist, then, if extension conditions are determined internally, the extension conditions will presumably be the same for both concepts, for they must both pick out nothing in all possible worlds. But if the extension conditions for concepts with empty referents are the same (at the very least in cases where the concept includes non-existence as one of its internal or intrinsic features) then, on the extension condition account, the concepts are the same, violating *CC*.

Possibly the greatest problem with Segal’s account, however, is how this account ties extension conditions to actual extension. If the extension conditions are determined by what a thought actually extends to or ‘picks out’ then why talk in terms of extension conditions at all, as opposed to mere extension? If extension conditions reduce to (or supervene on) extension, then no two concepts could have the same extension conditions and yet also have different extension. Segal’s account, therefore, faces the same problems as any account that tries to individuate concepts on the basis of reference, as discussed in section 3.

4.2 A Modal Account of Extension Conditions

Given the limitations of Segal's account, it is worth considering an alternative internalist account of extension conditions, which may be used to individuate concepts. A modal account could be presented where concepts include or contain an (internally-determined) formula that picks out their referents, dependent on the context of the concept-possessor. On such an account, the extension conditions for concepts work much like the reference-fixing conditions for indexicals – what concepts pick out may vary across worlds, while their extension conditions remain the same. To give an example, Oscar's concept would pick out the 'watery stuff' in the world he is in, meaning that it actually only picks out H_2O . However, if Oscar were transported to Twin Earth, his concept would pick out XYZ, because in that world the 'watery stuff' is XYZ. This account satisfies *MPC* – Oscar and Twin Oscar have the same 'watery stuff' concepts in spite of the fact that their concepts have different substances in their actual extensions, because they would refer to all and only the same substances in any world inhabited by both doppelgängers (i.e. even where extension varies, extension conditions remain the same). The account differs from that of Segal because extension conditions are not determined by what a concept actually refers to (meaning, for example, that Oscar's concept needn't refer to all 'watery stuff' in all possible worlds), but rather that reference is determined by extension conditions, which take into account modal context.

The modal account has no trouble dealing with *SRC* as, while Oscar and the future experts in his world have 'water' concepts that actually share an extension, were they to be transported to Twin Earth, the expert's 'water' concept would continue only to have H_2O in its extension, while Oscar's corresponding concept would now have XYZ in its extension. This means that extension conditions of the two concepts, and therefore the two concepts themselves, are different.

Furthermore, the modal account has some advantages over Segal's account of extension conditions in dealing with *CC*. In response to Frege Puzzles, it can say that the

concepts of those people who did not equate the morning star with the evening star were different because they had different extension conditions. Their ‘morning star’ concept had in its extension only the star that appeared in the sky in the morning. In the world that they were in, this happened to be the star that also appeared in the evening. But in a counterfactual world where there were two stars which appeared in an identical way to our one star, their concepts would only pick out the one that appeared in the morning. In the world where the star only showed itself in the evening, their ‘morning star’ concept would pick out nothing. As our present-day concepts pick out the morning star and evening star as the same star, we do not share a ‘morning star’ concept with those people in the past because our concepts have different extension conditions. A modal account of internally-determined extension conditions, therefore, looks like it could satisfy *CC* when it comes to being able to explain Frege Puzzles.

One potential challenge to the modal account’s ability to explain Frege Puzzles comes from the fact that ‘morning star’ and ‘evening star’ concepts *didn’t* pick out a star. People believed that what they were looking at and thinking about – what their concepts picked out – was a star, but they were wrong, it was in fact a planet, Venus. Can we say that Venus fell under the extension of the ‘morning star’ concept, if part of that concept included it being about a star, not a planet? The ‘morning star’ concept could extend to Venus, only if the people who possessed it had been wrong about their own concept (wrong about it being about a star, or applying to a star). However, there is no basis on which the internalist can argue that people were wrong in believing that this concept applied to a star, if extension conditions are determined internally. To work out the extension (or extension conditions) of a concept in an internalist-compatible way, we have to *first* look to the internal features of the corresponding concept, because the extension conditions themselves will be determined by the internal qualities of the concept-possessor. So, in the case of the ‘morning star’ concept, the modal internalist about extension conditions would have to conclude that it did not have anything in its

extension in the actual world as there was nothing that was both a star, and observed in the mornings in the position of Venus etc. As, for the same reasons, the ‘evening star’ concept also didn’t have anything in its extension, the modal account might appear as if it has to conclude that these are (or were) the same concept after all. The modal account, however, does have a reply to this challenge. The fact that the ‘morning star’ concept did not have anything in its extension does not mean that it did not have extension conditions that would distinguish it from the similarly non-referring ‘evening star’ concept. It is not part of the ‘morning star’ concept that it does not refer, but rather that it picks out a star that has certain other qualities such as appearing in the sky in the morning, looking a certain way etc. There will, therefore, be possible worlds in which such a star exists, and, were the concept possessor in one of these worlds, then their concept would have that star in its extension. The modal account is, therefore, able to adequately explain Frege Puzzles.

This account will still struggle to explain concepts which the concept possessors themselves do not believe refer – such as the case of ‘Father Christmas’ given above. The internalist who does not rely on extension conditions for concept individuation has a good way of explaining how my ‘Father Christmas’ and ‘unicorn’ concepts are different, even if I believe both of Father Christmas and of unicorns that they (necessarily) don’t exist. But, if I believe that neither of these beings exist then nothing can fall in the extension of either of my concepts, even in other possible worlds, which, if one individuates concepts in accordance with extension conditions, would mean that these are actually the same concept. In this way, even the modal account – the most promising internalist account of concept individuation provided so far – still cannot help but violate *CC* insofar as it comes to (at least some) non-referring concepts. It is possible, however, that this is a price the internalist is willing to pay – being weaker than traditional internalist theories of concepts on just this one aspect may not be reason enough to reject a theory that otherwise saves internalism from the conclusion that concepts are idiosyncratic. In

the following section, however, I will show that there are deeper problems with theories that attempt to fix the extension conditions of concepts in a way that is compatible with internalism.

5. Fixing extension conditions: the trouble with multiple realisability

Leaving aside the question of consistency with *CC*, it is worth noting that consistency with *MPC* and *SRC* is, on its own, no reason to adopt an extension condition model of internalist concept individuation. After all, individuating concepts idiosyncratically satisfies *MPC*, *SRC*, and indeed *CC*. If the extension condition model of concept individuation is meant to do more than idiosyncratic internalist individuation, it must also be *robustly* non-idiosyncratic. Segal believes that his model is robustly non-idiosyncratic due to the fact that extension conditions are multiply realisable: (some) concepts that differ internally will nonetheless have the same extension conditions. If this is correct, then, while internal states determine extension conditions, the conditions themselves, individuated more broadly than idiosyncratic internal states, could avoid the problem of concept individuation that is too fine-grained. But, of course, to say only this much is to say too little.

It is not enough to argue that two concepts are the same in virtue of their sharing the same extension conditions. It must be shown how we can work out when extension conditions are the same. Internalism itself is not committed to any position on when two internally different individuals share the same concepts. It is ad hoc to argue that those cases where we intuitively take two internally different people to be sharing concepts are cases where those people's concepts share extension conditions, if it is not explained *why* this is the case, or what unifies such instances.

The modal account provides a basic framework for fixing extension conditions by arguing that extension conditions are a formula that picks out referents based on the con-

text of the concept possessor. However, without knowing the formula itself (and indeed, how contexts are defined) we can say neither whether the use of extension conditions to individuate concepts is consistent with the fundamental conditions of internalism, nor whether, even if it were internalism-compatible, using such extension conditions would individuate concepts any more broadly than idiosyncratic individuation. After all, there need only be one world in the infinite number of possible worlds where the concepts of two people who are identical in every way except for one tiny internal difference would pick out different things (even *slightly* different things), for this to mean that these concepts are different. Just as it was impossible to think of a way to individuate concepts internally such that they would not end up being idiosyncratic, it may be impossible to individuate extension conditions internally such that they could be shared where there were any internal differences between concept-possessors. Something more still has to be brought in to determine when two internally different concepts nonetheless share extension conditions.

I will end this paper by considering one attempt to reconcile the demands of individuating extension conditions in a way that is determined by the internal features of concepts, while itself not being idiosyncratic. This is the case of extension conditions being determined by the actual or potential behaviour of the concept possessor.⁵ One of the arguments for internalism focuses on the causal inertness of relational properties – if it is the narrow properties of mental states that have causal power over behaviour, then we would expect conceptually-relevant behaviour to vary with the internal properties of concepts. Furthermore, behaviour is often used to work out concept extension: behaviours that use concepts (such as categorisation behaviour, and using language) can indicate how those concepts actually extend. For these reasons, behaviour is the most plausible contender for the ‘something more’ that needs to be brought in to determine when internally-determined extension conditions remain the same, even in cases where

⁵Segal (2004) himself draws on behaviour to argue that the extension conditions of natural kind concepts are not as Putnam describes them.

the internal properties of concepts vary.

However, behaviour will not work as a way of individuating extension conditions in an internally-consistent way. Consider, for example, the position that states that extension conditions are fixed by a disposition to identify. When asked to identify something as water, so long as they were in the same immediate context Oscar and Twin-Oscar would always pick out the same things. Extension conditions understood in terms of disposition to identify would explain how the doppelgängers' concepts were different from the experts in their respective worlds. In some contexts (i.e. on Earth) Oscar and future Earth experts would identify the same substance as water, but in other contexts (i.e. on Twin Earth) Oscar and the experts would diverge in their identification practices. This approach works nicely insofar as it is both consistent with, and potentially motivated by the principles that underpin the adoption of internalism – the argument that it is exclusively the intrinsic properties of mental states that have causal powers and that this explains the fact that people who have the same concepts will behave in the same ways. However, to know whether or not someone has classified a substance under their 'water' concept we already need to know what concept is in play in their classificatory practice. To know this we must already have settled the question of how to individuate concepts – so what we are left with is a regress. Disposition to identify just cannot do what we need it to – provide an independent means to individuate concepts.

A similar problem will befall an alternative account that states that extension conditions are determined by public language behaviour. On this approach it is the disposition to say "this is water" which determines the extension conditions of your water concept. However, to say something you need more than just to utter a particular sound – to determine whether the word Oscar is using is actually "water" and not merely a homonym we must establish what concept was associated with his speech utterances. Even if you don't accept the pragmatic picture that speakers' intentions determine speakers' meaning, it is difficult to make sense of an account of how disposition to say particular things

could be used as a way of individuating concepts if the speakers did not need to have the concepts we associate with the meanings of the words they are disposed to use.

Another concern for the internalist is that they must avoid the conclusion that any likely or future behaviour of concept-possessors is going to indicate the extension of their concepts. Taking future behaviour as a guide to current concept extension conditions (and, therefore, the nature of our current concepts) risks violating *MPC*. If Oscar found out that the watery stuff on his planet was H_2O and Twin Oscar found out that the watery stuff on his planet was XYZ, they would both be inclined to classify H_2O and XYZ as different substances, on the basis of which (future) behaviour we could conclude that they had had different concepts all along. In fact, looking to future behaviour is the kind of thing the externalist was trying to argue for. Burge (1979:94-5) believes this intuition to be so fundamental that it is written into the language we use to discuss concepts: “The patient does not say (or think) that he had thought he had some-category-of-disease-like-arthritis-and-including-arthritis-but-also-capable-of-occurring-outside-of-joints in the thigh instead of the error commonly attributed. This sort of response would be disingenuous. Whatever other beliefs he had, the subject thought that he had arthritis in the thigh.”

A further challenge, if the extension condition model of concept individuation is to be based on behaviour, is to present an account of behaviour individuation that is not itself idiosyncratic. Concept-driven behaviour is often varied, vague, and inconsistent, and therefore hard to individuate in a clear-cut way. Consider, for example, the fact that for pretty much any category people will ‘rank’ items as being more or less typical members of that category (Rosch (1973, 1999)). Penguins are considered less typical birds than sparrows, for example. Having an anvil dropped on your head is considered a less typical way of being killed by the mafia than being shot in an Italian restaurant (Barsalou (1987)). This is referred to as concepts having ‘graded structure’. Graded structure is rejected by some externalists as not being about the metaphysics of concepts themselves,

but rather the epistemology or behaviour of individual concept possessors (Rey (1999)). However, the internalist does not have the same motivation to reject the idea that concepts themselves are graded. In fact, it can be seen as a strength of internalism that it can account for this feature of how concepts are actually used by those who possess them. There is good reason to believe that our typicality judgements relating to category membership are represented in the internal structure of our concepts, not least because categorising behaviour is understood to be a typical example of conceptually-driven behaviour.

Where there is a difference in graded structure reflected in behaviour, should the behaviour be understood to be different? If so then, if extension conditions are determined by behaviour (such that different behaviour means different extension conditions), it becomes very unlikely that two instances of conceptually-relevant behaviour will ever be the same, entailing that extension conditions, and therefore concepts, are idiosyncratic. If behaviours reflecting different graded structures of concepts are not understood as being different, then it must be for reasons that are not reflected in the internal properties of concepts. As was shown by Barsalou (1987), people's graded categories change with context; there is a lot of variation between individuals (something that is not shown when results are only given as averages across groups); and the way people structure and apply their concepts does not even remain the same in the same individual over time.

Indeed, even when people seemingly agree on the definition of a concept, they will vary in the way they actually apply it. Barbara Malt (1994), for example, showed that subjects were likely to classify some substances such as pool water as being water, and others such as tea as not being water even though pool water contains less H₂O than tea. Even when subjects said they believed that water was H₂O, this was not borne out by their behaviour. While two internally-identical people may behave in exactly the same way under all the same conditions, those who differ internally, even if only slightly,

are likely to differ in their behaviour, if behaviour is individuated internally. Which leaves us with a further conclusion the internalist has to try to get around, very much like F2) and S2), which is that while two people who behave the same have the same concepts, two people that have (even slightly) different behaviour will also have different concepts.⁶

If we cannot individuate behaviour in a way that is both non-idiosyncratic and compatible with internalism, this rules out the last plausible way of trying to determine extension conditions such that they could form the basis for an internalist account of concept individuation that did not entail that concepts were idiosyncratic.

Conclusion

Internalists are not keen to embrace the idiosyncratic nature of concepts.⁷ However, as we have seen, there is no robust means of individuating concepts internally, or in accordance with internalist principles, that does not entail that concepts are idiosyncratic. No matter how you divide concepts up, if you hold something like F1) or S1), you are going to be committed to something like F2) or S2). Having run through all the possible internalist-compatible accounts of concept individuation, we can see that the internalist has no option but to accept that their position entails that concepts are idiosyncratic.

⁶It must be the case that, when two people exhibit the same behaviour they have the same concepts if concepts are individuated on the basis of behaviour and *MPC* is to be retained.

⁷Segal, for example, often brings up examples to support his position that assume both that concepts can retain their identity even after undergoing changes over time (2000:77-8), and that cognitive content is the same in cases where psychology would usefully generalise it as being so (“So, for example, the “water” concepts of twinned Bedouins on Dry Earth, XYZ Twin Earth and Earth share a cognitive content because a good psychology would subsume all of them under the same generalizations.” (Segal, 2000:85)).

References

- Barsalou LW (1987) The instability of graded structure. In: Neisser U (ed) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, Cambridge University Press, Cambridge, pp 101–140.
- Burge T (1979) Individualism and the mental. *Midwest Studies in Philosophy* 4(1):73–121.
- Farkas K (2003) What is externalism? *Philosophical Studies* 112(3):187–208.
- Farkas K (2008) *The Subject's Point of View*. Oxford University Press, Oxford
- Fodor J, Lepore E (1992) *Holism: A Shopper's Guide*. Blackwell, Cambridge, Mass.
- Georgalis N (2015) *Mind, Language and Subjectivity: Minimal Content and the Theory of Thought*. Routledge, Cambridge, Mass.
- Malt BC (1994) Water is not h₂o. *Cognitive Psychology* 27:41–70.
- Mendola J (2008) *Anti-Externalism*. Oxford University Press, Oxford
- Putnam H (1975) The meaning of meaning. *Minnesota Studies in the Philosophy of Science* 7:131–193.
- Rey G (1985) Concepts and conceptions. *Cognition* 19:297–303.
- Rey G (1999) Concepts and stereotypes. In: Margolis E, Laurence S (eds) *Concepts: Core Readings*, MIT Press, Cambridge, Mass., pp 279–300.
- Rosch E (1973) Natural categories. *Cognitive Psychology* 4:328–350.
- Rosch E (1999) Principles of categorization. pp 189–206.
- Segal GMA (2000) *A Slim Book About Narrow Content*. MIT Press, Cambridge, Mass.

Segal GMA (2004) Reference, causal powers, externalist intuitions and unicorns. In:
Schantz R (ed) *The Externalist Challenge: New Studies on Cognition and Intention-*
ality, De Gruyter