

We Shouldn't Assume Concepts Are Shared

Cressida Gaukroger

DRAFT. DO NOT CITE WITHOUT PERMISSION

October 2018

Abstract

Internalist theories of concepts are often taken to be inadequate because they cannot explain how concepts can be shared. I argue that, while it may appear obvious that people can share concepts, the evidence we have for concept sharing is behavioural and any account of concepts that purports to explain concept sharing cannot explain this behavioural evidence. Therefore, we should rethink whether being able to account for concept sharing is a genuine requirement for a theory of concepts.

One of the most fundamental convictions held about concepts, independent of which theory of concepts one endorses, is that they are the kinds of things that can be shared. As Gottlob Frege expressed it: '[O]ne can hardly deny that mankind has a common store of thoughts which is transmitted from one generation to another.' (Frege, 1948: 212) According to Jerry Fodor concepts '[a]re the sorts of things that lots of people can, and do, share.' (Fodor, 1998: 28) Jesse Prinz similarly includes concept sharing in his account of requirements for theories of concepts: 'Concepts

must be capable of being shared by different individuals and by one individual at different times.’ (Prinz, 2002: 14) These authors and many, if not most others agree: *Concepts are shared*. If true, then theories of concepts that cannot or do not account for how concepts are shared, are at best deficient, and, more likely, just false.

The inability to explain how concepts can be shared has been used as the grounds for rejecting theories of concepts in the ‘internalist’ tradition – those that take concepts to be mental representations the content of which must be determined by intrinsic properties of the mind.¹ Internal lives, dispositions, phenomenal experiences, and physical states are never identical across different individuals, nor do they remain stable in the same individual over time. For this reason it is believed that concepts must be individuated in virtue of their relational properties for concept sharing to be possible. And, as it is widely assumed concepts are shared, this has been a primary argument against internally-based theories of concepts.²

I will argue here in defence of internalist theories of concepts on the grounds that the ability to account for concept sharing is not as important as has traditionally been thought. Concept sharing is believed to be necessary for explaining a range of phenomena: coordination, behavioural success, linguistic communication (Frege (1948); Putnam (1970); Fodor (1998)) and the success of psychological explanations (Rey (1983/1999); Fodor (1998)). We believe concepts to play a role in producing complicated behaviour, such as the ability to generate and understand linguistic ut-

¹There is some debate over whether most of these theories – primarily found in psychology rather than philosophy – are in fact theories of concepts, or whether philosophers and psychologists are just failing to talk about the same thing. In the interests of avoiding a verbal dispute I will, in the tradition of Fodor (1998) and Prinz (2002), focus on the idea that concepts are roughly the basic units of thought, and theories of concepts can be judged on how well they meet a set of conditions or desiderata that are regularly found in accounts of concepts. See also Margolis and Laurence (2011).

²This has also been used as the grounds for rejecting holist theories of meaning (Fodor and Lepore:1992) and internalist theories of mental content (Sawyer:2007).

terances, which could only be possible if concepts were shared. Therefore, this behaviour is taken as evidence that concepts must be shared. However, I will argue, if concepts must be individuated based on their relational properties in order for us to account for their being shared, then the fact that they are shared cannot explain these phenomena because only intrinsic properties of mental states have causal power over behaviour. In other words, being able to explain concept sharing does not give a theory an advantage in explaining communication, or the success of intentional explanations of behaviour.

Section 1 will give a brief account of the different ways in which philosophers have understood 'concept sharing', and look at some of the reasons why accounting for concept sharing has been considered a 'non-negotiable' (Fodor:1998) requirement on a theory of concepts. The following three sections will discuss three different types of theories of concepts and their relation to concept sharing. Section 2 will argue that it is unlikely that internally-based theories of concepts can account for concept sharing. Section 3 discusses one type of 'externalist' approach to concepts that takes them to be mentally independent abstract entities. I will argue here that understanding concepts as abstract entities cannot help to explain how concepts contribute to public communication because those abstract entities need to be psychologically represented in order for us to talk about them, and it was the variation in psychological representations that was thought to be the barrier to such communication in the first place. Section 4 concerns the dominant type of concept theory found in philosophy – that concepts are psychologically real entities, but their content is determined by relational properties. In this section I will briefly outline the case for why the relational properties of mental states don't have causal power over behaviour and then use this to argue that, while such theories

may be able to explain conceptually-relevant behaviour, they cannot do so in terms of shared concepts because, insofar as concepts are shared, they are only shared in virtue of their relational properties. Section 5 will conclude with a discussion of why it can be useful to talk *as if* concepts are shared, even if we accept that they are not in fact shared.

Throughout the history of the concept sharing debate two pieces of data have seemed irreconcilable: internal mental states vary between people, and people (usually) behave as if they don't. But the fact that these are two points of *data* means that they are reconciled – they are both true of the world and we know this because we have observed them to be true. If we want to explain shared behaviour in conceptual terms, we have to go back to their internal properties (the very things that vary) and understand that, counter to common beliefs, all the tools we need for giving a conceptual explanation of this behaviour actually lie there. Any account of concept sharing will be merely taxonomical, and the desire for such a taxonomy is not a strong ground to reject internally-based theories of concepts.

1. Concept Sharing and Behaviour

Many theories do not give a robust account of concept sharing, and it has been argued that they should be rejected on these grounds. However, if all it means for a theory to account for concept sharing is that it merely *specifies* when two concepts are the same, where their shared identity is not a feature which had causal or explanatory power, then the requirement on theories of concepts to account for concept sharing is not as strong as it might seem. This is for two reasons – the first is that

sticking to this requirement would privilege theories with a *bad* account of concept individuation over those with no account. Consider, for example, a theory which individuates concepts on the basis of the astrological sign under which a thought containing those concepts is entertained. It would be a troubling conclusion if the mere fact that such a theory provided an account of ‘concept sharing’ counted as a reason in its favour. The second problem with a mere-sharing requirement is that internally-based theories of concepts could just ‘tack on’ an externalist taxonomy to avoid the charge of not being able to account for concept sharing. An account that argues, for example, that concepts are simple mental images, could adopt an additional component stating that these mental images should be individuated on the basis of what caused them. Such an imagist theory would remain the same in respect to its commitments relating to conceptual ontology and the causal role it takes concepts to play, but it would now have passed the ‘concept sharing’ requirement because of the addition of an inert taxonomy.

If concept sharing is a non-negotiable requirement on theories of concepts, what this requires of theories must be more than the ability to specify mere-sharing. For many who argue that concept sharing is essential to a theory of concepts, the reason concepts *must* be shared is because there is a range of phenomena that could not be explained if there weren’t shared concepts. Or, put another way, there is a range of phenomena that *wouldn’t be possible* without shared concepts – so concept sharing explains these phenomena in that it makes them possible. If we are to take the concept sharing requirement seriously it must be confined to only those theories where the sharing identified can actually play a role in explaining these phenomena (which would rule out the astrology account). A serious concept sharing requirement will also require that the account of sharing be ontological rather than

classificatory because in the case where an internalist theory tacks on an account of concept individuation, the tacked-on account is not doing any of the explanatory work.

As the basic constituents of thought, concepts are employed in explanations of phenomena from understanding and reasoning, to communication and categorisation. Take, for example, the case of language. Insofar as the meanings of words can be understood by many people, argues Putnam (1970), they are constituted by shared concepts. What it is then to hold a concept, as with what it is to understand the meaning of a word, cannot be unique to its holder, but must be both simple and general enough that its sharability can be explained. If concepts are not shared, what could explain the success of linguistic communication (Fodor and Lepore (1992))?

Similarly, shared concepts explain the success of intentional explanations of behaviour. This is true both for folk explanations and explanations in psychology (Rey (1985); Burge (1986; 1989); Fodor (1987); Sawyer (2007)).³ The explanations that we give of each other's behaviour invoke talk of mental states either explicitly, eg. 'Peter ran away because he has a fear of snakes', or implicitly eg. 'Lara opened the fridge to get a beer' (which involves assumptions about Lara's beliefs and desires concerning the contents of the fridge). Such explanations require that, for example, Peter shares a SNAKE concept with others, and with those formulating the explanation. If concepts are not shared, then to refer accurately to Peter's concept we need to introduce a neologism SNAKE_P which picks out a concept that only he possesses (and so on for every one of his and everyone else's concepts). In fact, if a theory cannot account for concept sharing between individuals it is unlikely to be able to

³"[W]hat makes such explanations intentional is that they appeal to covering generalizations about people..." Fodor (1998:29).

explain how a concept retains its identity over time even in the same individual, so neologisms will be required not just to accommodate the individual who possesses the concept, but the time at which they possess (or employ) it. This would prevent the generalisations necessary for identifying or describing psychological laws, and it would prevent our being able to understand or explain any of Peter's behaviour.

We can understand this argument in favour of concept sharing as including the following commitments:

1) MENTAL SHARING AXIOM: Concepts are mental things, which are common to many minds.

OR

1)* PUBLIC ACCESSIBILITY AXIOM: Concepts are non-mental things, which are accessible to many minds.

2) SOPHISTICATED BEHAVIOUR AXIOM: Concepts explain an important set of complex behaviour primarily found in humans.

3) SHARING EXPLAINS BEHAVIOUR AXIOM: This sophisticated behaviour could not occur if concepts were not shared.

In this paper I will take 'concept sharing' to cover both the MENTAL SHARING AXIOM and the PUBLIC ACCESSIBILITY AXIOM. The dominant view of concepts understands them to be mental representations. If this is the case, then what it means for two

people to share a representation is that they have distinct token representations, that are of the same type. My ELEPHANT concept may be uniquely represented by my mind/brain, but it is a concept I have in common with many others because it was caused by contact with elephants, or because I am a member of a linguistic community where the term 'elephants' refers to elephants, or because it functions to detect elephants, etc. There are many competing theories regarding how concepts should be individuated, but what they have in common is that they agree that the correct taxonomy of concepts will identify concept types that are multiply realised across different individuals. If you believe concepts are mental representations, you are likely to accept the MENTAL SHARING AXIOM.

In contrast, an alternative view of concepts understands them as non-mental abstract objects. On this view what it means for concepts to be shared is that they are *publicly accessible*, meaning that they can be represented by multiple individuals, even if these representations themselves vary from person to person. While the MENTAL SHARING AXIOM can be understood as stating that concepts are shared in the way that a book is shared if multiple people own different copies of *the same* book, the PUBLIC ACCESSIBILITY AXIOM can be understood as stating that concepts are shared in the way that a book is shared if multiple people have access to, or co-own, *one and the same* book.

This paper will challenge axioms 1), 1)* and 3), but it will not challenge the SOPHISTICATED BEHAVIOUR AXIOM. As stated above, concepts play a central role in explanations of a range of behavioural phenomena, and intentional explanations of behaviour involve the attribution of concepts to others. Such explanations that employ talk of concepts are very successful. Concepts are also central to explanations of different phenomena of thought such as linguistic understanding and rational

inference making. The main evidence we have for such phenomena existing is the behaviour such thoughts or thought patterns generate.⁴ A range of behavioural phenomena, then, are our primary window into, and evidence for, the nature of our concepts – in particular the complex behaviour that is primarily witnessed in humans. I will herein refer to these types of behaviour as ‘Sophisticated Behaviour’ (SB).

The SHARING EXPLAINS BEHAVIOUR AXIOM is crucial for the case in favour of concepts being shared. It can be understood as saying that either 1) or 1)* explains 2). This axiom captures the belief, expressed above, that it is not the mere existence of concepts that explains SB, but rather the fact that concepts are shared that explains SB. In the case of language, for example, it seems impossible to explain how understanding or successful linguistic communication could be possible if concepts weren’t shared by communicating individuals. If concepts explain SB, it must be in part because they are shared.

It clear that not *just any* account of concept sharing could play a role in explaining SB. For the ability to account for concept sharing to be a *virtue* of a theory of concepts, the theory must first explain the explanatory significance of its account of concept sharing. In other words, it must explain *how* its account of concept sharing explains SB. It is the SHARING EXPLAINS BEHAVIOUR AXIOM that answers the

⁴It is true that for some of these phenomena we have introspective, not just behavioural evidence for their presence and nature. I’m intentionally avoiding a discussion of introspection here as it is not clear what introspection reveals about concept sharing, particularly as it relates to the internalism/externalism debate. In defense of externalism Burge (1979) has argued that we must trust individuals’ reports of their own concepts, and these reports support the case for conceptual consistency over time even in the face of changes in our conceptual representations. For example, Burge argues, when someone learns that she has been using the term ‘arthritis’ incorrectly she does not just insist that she was using a different, idiosyncratic concept, but rather acknowledges that she just got something wrong about arthritis. However, as externalists posit that our concepts and thoughts can contain features of which we are unaware, and may not contain the features we take them to contain, self-knowledge generated from introspection notoriously poses serious problems for externalism. See, for example, Boghossian (1989).

question of *why* concepts must be shared.

As vital as the SHARING EXPLAINS BEHAVIOUR AXIOM is, however, it is rarely explicitly stated. Instead it is assumed that if one has an account of concepts according to which they are shared or publicly accessible, then the account automatically explains language and intentional explanations of behaviour. However, concepts being shared is not the same as concepts explaining the success of language. A step is missing. An account of concepts must not only be able to explain SB, but, if the SHARING EXPLAINS BEHAVIOUR AXIOM is correct, it must be able to answer the following question: What is it about the fact that (or the way in which) concepts are shared, that explains SB?

Rather than assuming that being able to account for concept sharing is a virtue of a theory of concepts, I will take it to be a virtue only if a theory that accounts for concept sharing also satisfies the SHARING EXPLAINS BEHAVIOUR AXIOM. However, I will argue that none of the accounts that satisfy the MENTAL SHARING AXIOM or the PUBLIC ACCESSIBILITY AXIOM satisfy the SHARING EXPLAINS BEHAVIOUR AXIOM. In fact, concept sharing *cannot* explain SB. I conclude that being able to account for concept sharing is not a necessary feature of a theory of concepts after all.

2. Internalists Can't Account for Concept Sharing

This section will explain why internalist theories of concepts do not appear to have the resources required to provide a robust account of concept sharing. It is such theories that are at stake in this debate, for if it turns out that theories of concepts do not in fact need to account for concept sharing then one of the greatest barriers

to their being accepted will be removed.

Assuming for the present that concepts are psychologically real entities or states, such entities will be comprised of certain internal or intrinsic psychological components. We can think of these as Fregean (1948) ‘conceptions’.⁵ What these internal components are may vary depending on your theory of concepts, but at the very least they are whatever is in the head, whether that is mental images; basic ‘switches’ in the mind that indicate or detect features of the external world; mental feature-lists; exemplars; theories; or any other number of internally-defined mental components. As every person has different psychological states in virtue of their dispositions and experiences, no two people have concepts that share qualitatively identical psychological properties.

For theories that take concepts to be mental representations of some kind or another, what it means to account for concept sharing is to provide a taxonomy which individuates concepts such that different individuals possess concepts of the same type. This is what is required to satisfy the MENTAL SHARING AXIOM. There are two ways in which one could approach the question of concept individuation if concepts are mental representations. The first way of individuating concepts – of providing a theory that tells us what it takes for any two concept tokens to be the same concept type – does so on the basis of their internal characteristics alone. Such an approach would entail that where two concepts share the same internal features, they are the same concept, but where their internal features differ, they are different concepts. This way of individuating concepts can be understood as being ‘internalist’ as it honours one of the fundamental internalist litmus tests, namely that it entails

⁵While what Frege had in mind was closer to unstructured feelings and images, the word ‘conceptions’ has been used more recently by some to indicate a class of theories primarily found in psychology that identify concepts with structured mental mechanisms. (Eg. Rey:1985; 2010).

that the ‘Oscar-doppelgängers’ of Putnam’s (1975) “Twin Earth” thought experiment will have the same WATER concept.

The second way of individuating concepts can, in contrast, be understood as ‘externalist’, though here it may be easier to think of it as ‘relational’ as, insofar as we accept that concepts are psychological entities, it individuates such entities based on one or more of their relational properties.⁶ These relational properties can be determined, for example, by the concept possessor’s social and/or linguistic community (e.g. Putnam (1975); Burge (1979)), their teleological role (e.g. Millikan (1984)), or their history of interaction with natural kinds (e.g. Kripke (1981); Putnam (1975)). A common argument in favour of theories of concepts that individuate them externally, is that this the *only way* that a theory can satisfy the MENTAL SHARING AXIOM.

To understand this challenge to ‘internalist’ theories of concepts, take prototype theory. The best known version of prototype theory explains concepts in terms of stored lists of the prototypical features of categories (Rosch (1973), (1975)). The features that make up one person’s concept DOG, for example, may include having a tail, barking, having a snout etc. According to prototype theory, the important properties of concepts – those properties that explain the role concepts play in guiding behaviour, and the role they play in thought (i.e. their relationship to other concepts and non-conceptual thoughts) – are their internal properties, namely the structure and contents of these feature lists. If we want to understand concepts, which in-

⁶Segal (2000) defines the distinction between internalism and externalism about mental content in terms of intrinsic and relational properties. According to Segal, externalism is the theory that cognitive content is a relational property, while internalism is the denial that mental content supervenes on anything but intrinsic properties. Note that by ‘relational’ properties Segal and I both mean relational between mind and (external) world. Relations between an individual’s concepts are ‘internal features’ for the purposes of this distinction.

cludes understanding what makes two concepts the same or different, we need to look no further than how they are represented (as opposed to looking at what they refer to, or what the corresponding word means in their linguistic community).

Prototype theory has a number of virtues. It is, for example, able to capture and explain certain kinds of categorising behaviour. As it originated in psychology, much of this behaviour is at a fine level of detail, performed by test subjects. For example, prototype theory predicts that if an individual has only (or mostly) been exposed to small, European birds, then a bird resembling a robin is likely to be the prototype she uses to distinguish her concept BIRD (Barsalou (1987)). This is measured by timing how long it takes subjects to classify something as a bird – individuals are quicker at identifying and classifying birds that are closer to their prototype. This theory predicts that prototypes will vary depending on the birds one has been exposed to, but even with variation in prototypes there is usually enough similarity that there is a very strong overlap between what different individuals classify as a ‘bird’, even if their classification times vary.

Prototype theory is not without its problems, and there are several other competing theories which can be found in psychology.⁷ However, a primary criticism from philosophy has focused on the inability for this theory to satisfy the MENTAL SHARING AXIOM.

⁷Further, Edouard Machery (2009) highlights the great deal of evidence that supports all three main theories of concepts found in psychology: prototype theory (which states that concepts are representations of the most typical features of category members), exemplar theory (which states that concepts are a set of singular representations of particulars) and theory theory (which states that concepts are theories which store causal or functional information about category members). The fact that there is evidence to support each of these theories, argues Machery, indicates that none of these are the ‘correct’ theory of concepts. Rather, we have many heterogeneous types of concepts which may co-refer, but have very little in common. Machery’s conclusion is that there is no correct or unified theory of concepts in psychology, and we should recognise that these different theories may not be in competition, but rather serve different ends.

According to prototype theory, possessing a complete feature list is sufficient for concept possession, although no one feature may be necessary for something to count as a member of the concept category. One person's DOG concept feature list may lack 'barks' as a feature, and another person's DOG concept feature list may lack 'has a tail' as a feature yet both could contain enough of the features of dogs that they could still be understood as being DOG concepts. If feature lists needed to be identical across people to count as being concepts of the same type, then it would be very rare for people to share concepts because feature lists are rarely identical across people.

There is evidence to suggest that the features considered to be prototypical of a particular category will differ depending on what tokens of that category-type individuals have been exposed to. In fact, it is even more complicated than this. According to prototype theory, our feature lists are not like a list on a page. Rather, the features that make them up will differ in importance or weight. Both 'has a snout' and 'is brown' might be on my DOG concept feature list, but 'has a snout' is a far more important feature than 'is brown'. It is, for example, more likely to come to mind when I think of dogs, and I will give it more weight when identifying dogs. This is referred to as 'graded structure'. The graded structure of concepts governs how likely you are to identify something as a typical instance of a category member. 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. Indeed, this graded structure is thought to be at work even in obscure or novel concepts, 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)).

What would it take for two prototypes to be the same based solely on their internal properties? They would not only need to be comprised of identical feature lists as regards their contents, but the contents of these lists would need to be identically-weighted. However, it is unlikely that any two people's concepts will ever be the same in this way. 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 one individual over time.

Insofar as it individuates concepts internally, prototype theory is unable to give an account of how normal people, under normal circumstances could have *the same* concepts. However, this is not something that it tries to do. Advocates of this theory reject the claim that a theory of concepts needs to satisfy the MENTAL SHARING AXIOM. Instead prototype theorists argue that it is sufficient to account for *similarity* between concepts. Even if feature lists are not shared, they can resemble one another. Rosch and Mervis (1975) describe the relationship between concepts in terms of the Wittgensteinian idea of 'family resemblance'. In other words there is nothing that we could define as being the real features of a particular concept, but rather we can identify a family of concepts which share some features with different members of other concepts in this family, even if there is no defining feature common to them all.

The approach of talking in terms of similarity as an alternative to concept identity has faced serious challenges. Jerry Fodor and Ernest Lepore (Fodor and Lepore (1992); and Fodor (1998)) reject the idea that there can be a meaningful account of concept similarity. They argue that for two things to be similar they must share

some identical features. But, they continue, if identity is necessary for similarity, and one has a theory that cannot account for identity, they cannot simply avoid this problem by claiming that all they are aiming for is similarity. They use prototype theory to illustrate this point: It is easy to claim that two concepts are sufficiently similar, they say, if they have enough overlap of feature-lists, but what is it for a feature on one list to overlap with a feature on a different list? If my concept DOG and your concept DOG both include ‘has a snout’ as an item on our feature-list then we are still going to have to explain what it means for both of us to mean the same thing by ‘has a snout’. Such a feature minimally requires the concepts POSSESSION and SNOUT, but this assumes that these concepts are *identical* between our two concept possessors. Fodor and Lepore argue that, while different theories of concepts may construct their accounts of concept similarity in different ways, they all face the same problem: “[A]ll the theories of content that offer a robust construal of conceptual similarity do so by presupposing a correspondingly robust notion of concept identity.” (Fodor (1998): 34)⁸

The problem prototype theory faces, that it entails concepts are idiosyncratic, is equally true of other internally-based theories of concept individuation. Consider theories that individuate concepts on the basis of inter-conceptual relationships. If what it is to be a particular concept relies on other concepts (or areas of mental information) which themselves rely on yet other concepts, then we need to explain the whole network to say what a single concept is. As Fodor (1995b: 6) points out: “If what you’re thinking depends on *all* of what you believe, then nobody ever thinks the same thing twice, and no intentional laws ever get satisfied more than once...”

⁸There have been attempts to answer Fodor and Lepore, see Churchland (1993); Laakso and Cottrell (2000); and Prinz (2002).

Without factoring in some relational properties it seems that no account of concept individuation can classify concepts as types as opposed to individuating them so narrowly that only token-identical concepts can be said to be the same concept. Consider the book analogy: What it is to share concepts for any theory that takes concepts to be mental states or representations is like different people having different copies of the same book. However, if what makes two books count as ‘the same’ is that they share all the same words, have the same cover, and are in the same condition, then if everyone owns similar books that have different covers, contain a few different words, or range in condition, they cannot be said to have different copies of the same book. Any model that individuated concepts *solely* in terms of their psychological properties cannot explain how two individuals whose concepts differed internally, *even slightly*, could have the same concept. Indeed, if we can no longer talk in terms of two concept tokens being members of the same type then we cannot even give an account how one person can retain the same concept over time in the face of changing beliefs and experiences. Theories of concepts found in psychology – prototype theory, exemplar theory, and theory theory, for example – all face this problem. This is also a problem for internalist theories of mental content.⁹

We can understand, therefore, the significance of the belief that theories of concepts must either satisfy the MENTAL SHARING AXIOM or the PUBLIC ACCESSIBILITY

⁹Joseph Mendola rejects this critique of internalism as he denies that internal properties are as private and idiosyncratic as we assume they are (Mendola (2008): 227). This follows from his unusual definition of internalism, which individuates the contents of thoughts as being the same if they lead to the same behaviour (defined a-contextually such that behaviour is determined exclusively by bodily movements) (Mendola (2008): 229; for some problems with this account see Ebbs (2013); Richard (2013)). However, it is not as if behaviour can avoid problems of vagueness that would undermine attempts at robust individuation, particularly if we cannot individuate it in terms of the relationship between a body and an environment. If behaviour is defined merely as the position of limbs, for example, what does it mean for limbs to have the ‘same position’ either in two people or in one person over time?

AXIOM. If not being able to account for concept sharing is a reason to reject a theory of concepts, then this will lead us to reject internally-based theories of concepts and internalist theories of mental content. That such theories cannot account for concept sharing has often obscured their other theoretical advantages. If concepts are mental representations which get their content from what is inside the head there is no difficulty in accounting for concepts of non-existent entities, while theories of concepts that individuate them on the basis of reference to sets of classes of real things in the world struggle to explain concepts like ‘unicorn’ for which there is no set which it picks out (Segal (2000)). If content is determined by the way the concept is represented rather than what it picks out in the world, then Frege Puzzles stop being so puzzling – having separate ‘morning star’ and ‘evening star’ concepts which have been caused by the same planet doesn’t present any puzzles on internalist accounts (Jackson (2003)). If, as the externalist argues, what is outside the head contributes to conceptual content, then we face difficulties explaining how we can have first-person knowledge of our own minds (Loar (1988); Boghossian (1989); Farkas (2008)). As we have seen (see also section 4) internalist theories are at an advantage in explaining the practical aspects of concept possession – how we acquire concepts (Prinz (2002)), and how concepts can cause our behaviour (Fodor (1987); Crane (1991); Gaukroger (2017)).

The inability to account for the MENTAL SHARING AXIOM or the PUBLIC ACCESSIBILITY AXIOM is not the only charge levelled against internalist theories of concepts, but it is certainly a formidable barrier to their being more widely accepted. However, remember that these two axioms are, on their own, not particularly important. Rather, what makes them important is their conjunction with SHARING EXPLAINS BEHAVIOUR AXIOM. The mere fact that internalist theories cannot ac-

count for concept sharing shouldn't be so bad, if it turns out that, counter to the way it is normally presented, concept sharing does not in fact explain sophisticated behaviour.

3. Concepts as Abstract Objects

Conceptions differ between people – this is the problem. As stated in the introduction, one way of accounting for how concepts are shared is to identify concepts with conceptions, but to individuate them according to their relational properties (as opposed to the features of these conceptions themselves). An alternative is to deny that conceptions are concepts.

In accounting for how concepts play a role in explaining the success of language, Frege (1948), for example, was concerned with how to reconcile the following: to be able to communicate linguistically, communicators need to have some kind of shared understanding or knowledge of the words they use. However, we cannot know everything about the referents of our words and often have very different mental representations associated with them from the representations held by others. For Frege, the units of which there must be shared knowledge to enable linguistic communication cannot be the referents of our words, because it is possible for one person to associate different concepts with the same referent. Conceptions, Frege argued, cannot explain the publicity of language – in fact they threaten to undermine it. Frege, therefore, introduced the idea of a 'mode of presentation'. As abstract objects, modes of presentation (herein, MOPs) do not vary from person to person (although people can have varying conceptions of the same MOP), but they

are also not synonymous with reference, since more than one MOP could apply to the same referent.

According to Frege, the meaning of a sentence is a proposition – an abstract object with truth bearing properties – so sentence meanings exist independently of the mind. Propositions are what one grasps as one comes to understand the meaning of a sentence. When two people have the same belief – such as the belief that ‘elephants are large’ – what makes it the case that they have the same belief? According to the Fregean tradition, ‘elephants are large’ is a proposition to which many people can all bear a belief-relation (it can be the object of multiple beliefs). Applied to concepts, one can understand propositions as being made up of more basic units – similar to sentences that are made up of words. As the proposition itself is abstract and mind-independent, this entails that the units that make it up will similarly be abstract objects. It is these basic units that make up Fregean propositions that are understood by some to be concepts.¹⁰

Georges Rey (1985; 1999), defends the view that concepts are abstract objects and rejects the idea that concept ontology requires a psychological component. According to Rey, concepts are like numbers – entities that can be grasped mentally, and used in rational thought, but which are neither mental, nor natural artefacts. This theory understands concepts in a Platonistic or Fregean way (Rey (1999): 296). Rey refers to his particular account of concepts as the ‘Hypothesis of External Definitions’: “the correct definition of a concept is provided by the optimal account of it, which need not be known by the concept’s competent users.” (Rey, 1999: 293)

Rey’s theory of concepts, sometimes known as definitionism, appears to easily

¹⁰Peacocke (1992) and Zalta (2001) hold such a belief about the ontological nature of concepts. See Margolis and Laurence (2007) for a more detailed account of the view of concepts as abstract objects.

satisfy the PUBLIC ACCESSIBILITY AXIOM. For example, the fact that being composed of H_2O is necessary and sufficient for a substance to be water is part of the WATER concept on this account, independent of whether anyone knows this. What it takes for an individual to 'possess' a WATER concept is for that individual to have the appropriate relationship with the definition of water – a relationship that may well stand short of their actually (or perhaps fully) knowing this definition. For two individuals to 'possess' the same concept, they do not need to represent the same definition in the same way. It is sufficient that both concepts refer to a category that is individuated by its definition conditions, even if the concept-possessors are not entirely aware of these conditions. There is, therefore, no problem posed by the fact that there is variation in individual conceptions according to Rey.

It appears then that defining concepts in entirely non-psychological terms avoids the problem of people with different psychological lives still being able to share concepts. However, this is not the case. Arguing that there must be an externally located conceptual component that accounts for how psychologically or subjectively different individuals can possess the same concept does not avoid the problem that arises when we consider how these external components are represented internally. After all, the problem for which concepts as external entities were posited to answer is how we get from differing internal mental content to public communication. It seemed as if the problem was that internal ways of representing different referents would always differ. Rey and Frege claim that we do not internally represent the referent of a concept, but the concept itself. How does this solve the problem? These new external entities – abstract objects – must themselves be represented internally. The varying nature of internal representations, if it is a problem for theories that do not involve MOPs, cannot be solved via the introduction of MOPs.

Our having two different conceptions of the same abstract object may satisfy the PUBLIC ACCESSIBILITY AXIOM, but it cannot explain SB any more than our having two different conceptions of the same concrete object can. This problem with the failure of abstract object theories of concepts to satisfy the SHARING EXPLAINS BEHAVIOUR AXIOM is mirrored in criticisms others have made of such theories. For example, Fodor (1998) makes the following observation: Frege brings in the idea of MOPs, as reference alone is not enough to differentiate concepts. However, Frege was concerned to make MOPs external – otherwise they would fail to account for their public accessibility; for, if they were internal then they would differ from person to person. If concepts are conceptions, thinks Frege, then this would mean that there are, potentially, an infinite number of concepts relating to the one referent and, therefore, we cannot explain communication. But, Fodor points out, there are many ways that modes of presentation can entertained:

[I]f MOPs *aren't* mental, what kind of thing *could* they be such that *necessarily* for each MOP there is only one way in which a mind can entertain it? (And/or: what kind of mental state could entertaining a MOP be such that *necessarily* there is only one way to entertain each MOP?)
(Fodor (1998): 20)

Consider Rey's definitionism. If concepts are externally located definitions, a question arises over how these definitions are manifested mentally (or recognised, or learned). Even if we were to agree that only certain definitions need to be mastered for us to share the same concept, we must nevertheless be provided with an account of what it takes to master a definition. If mastering a definition requires forming a particular mental representation, then we are left needing to explain how these

mental representations come to be identical.

Concepts in the Fregean sense were brought in to solve the problem that conceptions, in virtue of the fact that they vary from person to person, could not explain sophisticated behaviour. It was not just that an explanation in terms of abstract objects was given for SB. Rather it was believed that such an explanation was *needed* to explain SB precisely because explanations that looked to conceptions were fundamentally unable to account for such behaviour. Defining concepts in entirely non-psychological terms was meant to avoid the problem of explaining how people with different psychological lives were still able to share concepts. However, unless we are puppets unconsciously controlled by the realm of abstract objects, the presence of such abstract objects means nothing for explaining human behaviour if they are not represented some way or another in actual human minds. If the question is how we get from differing internal mental content to public communication, the answer cannot come from bypassing a discussion of the role of conceptions in causing public communication.

4. Conceptions as Externally-Individuated Mental Representations

Abstract object theories of concepts are unable to satisfy the SHARING EXPLAINS BEHAVIOUR AXIOM, without which the PUBLIC ACCESSIBILITY AXIOM is worthless. The problem that the abstract object theories of concepts face is that they claim to be able to explain SB, but fail to account for the mental mechanisms that actually cause SB. It is unclear what explanatory work abstract objects could do when

it comes to accounts of SB. An alternative to the abstract object approach that has been more widely adopted is to argue that, while concepts are mental representations, the variation between them is irrelevant because their content is determined externally. We can understand this line of argument by using the analogy of words. A word, say ‘canal’, can be represented in many ways that vary widely – it can be written or spoken, pronunciation and writing style can differ etc. However, these variations are irrelevant because what is significant about the word ‘canal’ – what explains how it is used, how it relates to other words, how it influences thoughts and behaviour – is what it refers to. The content of the word is not determined by any of the intrinsic features of its token presentations, but by a relational property, namely the reference relation that it bears to canals. Similarly to words, concepts pick out things in the world, and so if they can be individuated in terms of these relational properties, we can have an explanation of SB which is not undermined by the fact that there is variation in the intrinsic properties of individual conceptual representations. Or, at least, so the argument goes.¹¹

This approach to concept sharing is common to pretty much any ‘externalist’ theory of mental content which argues that mental content supervenes on the relational properties of mental states. For an example of someone who is very explicit about how this works in the case of concepts, consider Fodor. In “Concepts: A Pot-boiler”, Fodor (1995a) distinguishes between the two different types of properties of concepts – internal properties (which he calls ‘causal’ properties) and relational properties (which he calls ‘semantic’ properties). Consider, for example, a SPOON

¹¹This is a slight simplification – concepts, like words, do not just pick out things in the world, but also refer to (or aim to refer to) abstract objects, theories, and non-existent entities. The fact that causal and referential theories of conceptual content have a harder time accounting for such concepts is another factor that counts against such theories. See, for example, Segal (2000).

concept. How my SPOON concept appears to me; what I associate with it; what prototypes I employ to identify spoons; even what patterns of neurons firing realise my concept – all of these are internal properties of my concept. Consistent with the arguments presented in section 2, Fodor argues that these properties must be irrelevant to concept sharing. According to Fodor, the SPOON concept held by different people will be realised by symbols with different syntactic features, or that play different roles in relation to their other concepts – this concept will differ in its internal properties from person to person. However, just because people differ internally, does not mean that they cannot share a SPOON concept. If individuals do share concepts, it must be in virtue of the relational properties of their concepts.

Fodor (1995a) argues that I share my SPOON concept with anyone who has the right relationship with spoons: anyone who, like me, has *some kind of symbol* that is reliably activated when and only when spoons are present. For Fodor (1995a, see also Fodor and Pylyshyn (2014)), when we individuate concepts – when we say what it is for my SPOON concept to be a *spoon* concept, or when we identify what it is that accounts for psychologically different people being able to share concepts – this must be done on the basis of relational properties.

The problem with this view lies in the following thesis:

RELATIONAL INERTNESS THESIS: The relational properties of mental states do not have causal power over behaviour.

The RELATIONAL INERTNESS THESIS is a widely-believed thesis, even amongst those who believe that mental states must be individuated relationally. Stich (1978), Burge (1982), McGinn (1982), Fodor (1987), and Jacob (1992), for example, all argue that,

regardless of one's theory of content, causation is local. A person's context may well have an effect on how they behave, but only insofar as it is mediated by intrinsic mental properties such as states of their brain. This point is made in arguments in favour of internalism on the basis that externalism fails to account for the causal aspects of mental content.¹²

A change purely in the relational properties of mental states, but with no difference in intrinsic (or as we are calling them here, 'internal') properties of mental states, will see no difference in the causal powers of those mental states. This thesis is, in fact, so widely accepted that it is assumed in many of the scenarios used to illustrate the case *for* relationally individuating concepts (i.e. for externalism about concepts). Oscar and Twin Oscar continue to have very similar interactions with the watery stuff around their respective selves; the subject in Burge's (1979) arthritis example goes to the doctor and complains of the pain in his thigh in all possible worlds Burge imagines for him; and, despite the fact that he argues for all content being broad and, therefore, a complete change in broad content faced by Donald Davidson's (1987) swampman, both the original Davidson and his swampman replica exhibit the exact same behaviour.

The RELATIONAL INERTNESS THESIS has received a range of different replies arguing that the relational properties of mental states actually do have causal powers. Such replies mostly focus on the question of whether people with internally identical but relationally different mental states, such as Oscar and Twin Oscar, actually behave the same on their respective planets, or whether they behave differently.¹³ In the simplest version of this objection it is pointed out that the doppelgängers are, af-

¹²Loar (1988); Crane (1991)

¹³See Peacocke (1981), Evans (1982) and Hornsby (1986). See Fodor (1987); Jacob (1992); Gaukroger (2017) for replies.

ter all, going to be drinking different substances, swimming in different substances, washing in different substances, and so on. It is argued that while their physical actions might remain the same, their *behaviour* differs between Twin Worlds.

However, this line of argument misses the point. When Oscar and Twin Oscar behave differently (if you believe that they do), is this behaviour not caused by the internal properties of their mental states? The difference in their behaviour is explained by the difference in their environments, and the difference in their respective relations to their environments, *not* the difference in the *relational properties* of their respective mental states. It is only if the *relational properties* of mental states are doing the work that we can feel confident that the relational properties of mental states have causal power. Were the two Oscars ‘switched’ so that Oscar was now on Twin Earth but still, presumably for a while at least, retained his WATER concept, Oscar would behave exactly as Twin Oscar did on Twin Earth in spite of their different concepts – Oscar would drink XYZ just as Twin Oscar did, would categorise XYZ under the label ‘water’, and would be able to successfully communicate about the watery stuff around him. This suggests that, in this case the difference in behaviour would not be explained by the difference in the relational properties of their respective concepts, but rather that the difference in their behaviour *and the difference in the relational properties of their mental states* are jointly explained by a third factor, the differences in their contexts. Accepting the role of context in determining behaviour does not require believing that the relational properties of mental states have causal power.¹⁴

¹⁴I will use ‘explaining behaviour’ below to mean ‘causally explaining behaviour’. While the relational properties of a concept could be used in an explanation of behaviour as a way of picking out a particular concept, its use would be *non-explanatory*. Consider someone in the past who, believing whales are fish (Sainsbury (2014)), utters: “The boat is damaged because it was hit by a fish.” The classification as ‘fish’ helps the listener identify the subject of the description. This classification

To reject the RELATIONAL INERTNESS THESIS in a way that is consistent with satisfying the SHARING EXPLAINS BEHAVIOUR AXIOM, one would need to present an account where it is not just *any* relational properties of mental states that cause public behaviour, but specifically those properties that form the basis of a theory of concept individuation which can account for concept sharing. It is easy to understand what form an explanation of SB could take if given in internal terms as it is possible for the explanation to be entirely mechanistic. In contrast, individuating concepts relationally is an exercise in taxonomy or classification, not in identifying alternative causal mechanisms. Introducing an alternative classificatory system is not going to be enough to explain SB if the internal mechanism does not exist to support it. Classification cannot explain what a failure in the mechanism is unable to explain.¹⁵

Some, for example, Loar (1988, 2003), Block (1986), Chalmers (2002), and Prinz (2002), argue that there is a significant role for narrow content (content that supervenes on the internal properties of mental states), but it should be combined with broad content (that which supervenes on the relational properties of mental states). Such two-factor theories, however, are at no advantage when it comes to satisfying the SHARING EXPLAINS BEHAVIOUR AXIOM. Two-factor theories have no resources in explaining SB that go beyond those of internalist and externalist one-factor theo-

does not contribute to a causal explanation of what happened to the boat. See section 5.

¹⁵There are additional challenges for externalists who believe their theories of concepts can explain behaviour such as communication. Pollock (2015) argues that social externalism is inconsistent with the standard account of communicative success. Mendola (2008) argues that vagueness or indeterminacy of reference causes problems for theories of semantic content based on reference. This Quinean point can extend to non-referential theories of concept individuation – it is not enough to say that DOG concepts are shared because they are caused by dogs, one needs to give an account of why it was something about the category *dog* that was doing the causal work as opposed to particular breeds of dog, a set of the specific dogs an individual has come into contact with, or animals that have dog-like properties etc.

ries. If they individuate concepts on the basis of their intrinsic properties they will face the same criticism as regards accounting for concept sharing that those solely-internal theories do, namely they fail to satisfy the MENTAL SHARING AXIOM. If, alternatively, they individuate concepts relationally their theories may inherit other virtues, which I will discuss briefly below, but, for the reasons outlined above, they will still fail to show how *concept sharing* explains SB. The tools you need to account for concept sharing just aren't the kinds of tools that can explain SB.

5. Roles for and Alternatives to Concept Sharing

I have argued here that those theories of concepts that can account for either the MENTAL SHARING AXIOM or PUBLIC ACCESSIBILITY AXIOM have been unable to account for the SHARING EXPLAINS BEHAVIOUR AXIOM. This is a problem if you take SB as evidence for there being concept sharing which, I have argued, is precisely what advocates of concept sharing do. One could still maintain an account of concepts that individuates them externally in spite of this problem, but the fact that concepts are shared on such an account would not contribute to our understanding of what concepts do in terms of the role they play in human behaviour. It has been argued that concept sharing (and by extension an externalist account of concepts) is needed for explaining language, but without being able to explain acts of linguistic communication in any causal terms, we might question the value of such an explanation.

A further problem for externalist theories of concepts that follows from my arguments in section 4 is that it is not clear what could count as evidence for concepts

being shared. It is also not clear what could count as evidence for one externalist account of concept individuation over another. These facts should all be of serious concern for those who defend externalist theories of concepts, but they should also offer some relief for any defenders of internalist theories of concepts who were concerned by the inability of their theory to provide a robust account of concept sharing. Whatever behaviour it is that we took concept sharing to explain must in fact be explainable *in spite* of the fact that the features of concepts causally responsible for such behaviour are not shared.

But, one might ask, how can we explain behaviour in terms of internal properties if these properties are not shared? Wouldn't we have to introduce neologisms to pick out each concept token which would undermine the fact that explanations do not work unless they can be generalised? There is an ontological question relating to concept sharing: Is there a way of individuating concepts according to their real nature such that they are shared? I have argued that the answer to this question is 'no'. And then there is a pragmatic question: Whether or not concepts really are robustly shared, might it not be useful to talk *as if* they were? I think the answer here is 'yes'. The fact that it is useful in facilitating general explanations to talk as if concepts are shared does not mean that this talk accurately represents their underlying nature. However, we have a reason not to give up on such talk altogether.

Being able to individuate concepts such that we can generalise across them is extremely *useful* – making generalisations on the basis of individual instances is the basis of science, it is the basis of folk psychology, it is what allows us to meaningfully talk of concepts in the first place. The fact that it is useful to be able to generalise across concepts, is not a reason to reject internally-based theories of concepts which

do not have written into them an account of how to broadly individuate concepts.¹⁶ We know that psychological explanation works – as I have said above we are very good at predicting the behaviour of others and of ourselves on the basis of psychological explanations that employ generalisations over concepts.

Recall in section 1 I suggested that an internalist theory of concepts could just ‘tack on’ an externalist theory of concept individuation to meet the concept sharing requirement. In such a case internalist theories could still be judged on their own merits which would not place an emphasis on how they individuated concepts. However, we may need a separate set of requirements to judge the usefulness of the account of concept individuation they were paired with. If generalising across concepts is needed for psychological explanation, but classifying concepts in a way that generalises them as if they were shared is not going to be grounded in shared properties that are essential to their nature, on what grounds can we distinguish between competing theories of concept individuation? One solution is to individuate concepts relative to our explanatory goals. Indeed, as the reasons for choosing between competing theories of concept individuation would be their usefulness in facilitating successful explanations of behaviour, different models of concept individuation may be appropriate under different circumstances.

Frances Egan (2014b,a) argues that grounds for determining a good account of cognitive content are pragmatic, not ontological. According to Egan the ascription of content should take into account context, our explanatory aims, and the ease with which we can understand and employ different levels or types of generalisations.

¹⁶Burge (1986) argues that psychology itself type-distinguishes mental states on the basis of relational properties, and this extends to folk psychology; after all, we use intentional language to describe behaviour. However, the fact that broad language is used in psychology does not entail that psychologists are (or should be) committing themselves to the position that the relational properties of concepts determine their content.

For this reasons there can be multiple content ascriptions that “serve different explanatory purposes. The idea that computational cognitive science is looking for, or fixing on, *the* content of mental states finds no support from actual theorizing.” (Egan (2014b): 131) The aim of Egan’s argument is slightly different to mine here, but its spirit is very similar.

Talking in terms of usefulness as a criterion of concept individuation not only treats the task of individuating theories appropriately (as non-ontological), but it has the advantage that it is, to some extent, testable: once we have fixed our aims we have a basis for comparing competing individuation models relative to how well they satisfy those aims. Such a criterion has the level of flexibility required given the wide range of contexts and aims in which psychological explanation would be useful. Furthermore, it does not tell us which aims we should have, which means it avoids smuggling in assumptions about concepts into an account of concept individuation. In other words, we needn’t fall back on mere intuitions to choose between competing models of concept individuation, even if those models were originally shaped to fit our intuitions.

If concepts should be individuated differently depending on the context of explanation then there is no independently ‘correct’ account of concept individuation. Consider another case where competing models, introduced to allow generalisations, are judged more or less appropriate based on their usefulness in relation to a specific context where it doesn’t make sense to say that one model is more or less ‘correct’ than another: statistical probability. Statistical probability is a function over a population – its value depends on the interests of the user for whom it is a tool. For an individual to have a particular probability of, say, developing diabetes, means that she is a member of a population x where n -number of members of x have

historically developed diabetes. The probability of developing a disease for any one person is not an instantiated trait. Any one individual's likelihood of developing the disease is merely a function over a particular group of which she is a member. Her actually developing the disease, if she does, has a local explanation. Statistics can be very useful for making predictions, but they are not causes within themselves.

The parallel with concept individuation can be understood as follows: We have established that individual instances of behaviour have a local explanation – that is to say insofar as they are partly explained by the presence of certain concepts, they are explained by the intrinsic (and so individuation-neutral) properties of these concepts. However, we might want to say something that generalises over individual instances so we can explain and predict behaviour more broadly. In which case we can talk as if individuals share concepts just as we talk as if statistical probability is a real property that individuals possess.

Any one individual will have a range of probabilities of developing a certain disease in virtue of the fact that they are a member of a variety of population groups. Jemima may, for example, have a particular probability of getting diabetes based on her family history, say, a 3% likelihood. However, when she is seen as a British citizen, she may have a 7% probability of getting diabetes. As she lives in Nauru, however, the country with the highest rate of type-2 diabetes, her probability in virtue of being a resident there rises to 30%. All these things are true of Jemima – she simultaneously has a 3%, 7% and 30% probability of getting diabetes. Probability in this case is a tool that allows us to do many of the things that an individuation model should allow us to do – aid in explanations and predictions.¹⁷ What features

¹⁷There are cases of useful generalisations across groups that do not pick out the features of *any* one member of that group. The idea of the 'learning curve', for example – a curve which measures the rate of progress in acquiring new knowledge or skills – has proved to be a useful tool, particularly

we take to be relevant to a particular measure of probability will be determined by what we are using probability to help us explain.

This is the same in the concept individuation case. As pointed out by Loar (1988) if we want to understand the behaviour of the Twin-Earth style doppelgängers we do not need to know whether they live in a world with H₂O or XYZ. Classifying the twins as having the same WATERY-STUFF concept would be useful for many contexts in which we wanted to explain their actions. However, as we rarely need to make psychological predictions which apply equally to real people on earth and people on structurally different twin planets, the more cumbersome classification of ‘WATERY-STUFF concept’ may actually be less appropriate than the usual WATER concept when looking at individuals all in the same context (ie the context where the main watery stuff around is water).¹⁸ At other times, it will be useful to individuate concepts such that individuals who have all had experience of H₂O nonetheless do not share concepts – in some contexts, it is appropriate to distinguish between the concepts of experts and lay people, for example, in order to enable more fine-grained predictions in behaviour. In the same way that it can be true that Jemima has many different probabilities of getting diabetes, it can also be true that a single concept possessed by Oscar can simultaneously be a WATERY-STUFF concept, a WATER concept, and a NON-EXPERTS WATER concept. None of the competing concept individuation models in this case are more accurate in the sense that they more accurately represent the

when making predictions about the rate and structure of learning for a population. However, no one actually learns in a curve. Rather, graphs that show the rate of learning in single subjects will consist in steps and spikes – the smooth ‘curve’ comes from averaging over trials and/or participants. No one individual might have concepts in the way we describe them when individuating them in accordance with relational properties, and still the relational individuation could be useful so long as we were talking in terms of populations rather than individuals.

¹⁸See Egan (2010, 2014b) for a more detailed example used to illustrate the importance of accounting for ‘ease of explanation’ when choosing to assign mental content.

essential natures of the relevant concepts. Rather, each account will be better or worse relative to our explanatory goals.

I have argued here that SB cannot be causally explained by concepts in virtue of their relational properties. This means that if we want conceptual explanations of SB, we must look to their intrinsic properties. However, this does not mean that we cannot incorporate accounts of concept individuation to allow us to generalise across findings. Rather we should understand the ability to generalise as a tool available to, as opposed to a feature of, these accounts of concepts. To understand generalisations across concepts in this way would be to recognise that the property of being shared is not intrinsic to concepts, nor is it essential to them, but rather it is a function of the way we classify them.

Theories of concept individuation, therefore, should be taken as serving a different purpose to theories of concepts themselves. Talking of concepts as being shared is useful, but it is not strictly correct: concept sharing does not consist in anything over and above our classification systems, which should be assessed not in terms of accuracy but relative to pragmatic criteria. Therefore, being able to account for concept sharing is not a requirement (or even a desiderata) of a theory of concepts – it should not be used as the basis for rejecting one theory of concepts or preferring another.

References

Barsalou LW (1987) The instability of graded structure. In: Neisser U (ed) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categoriza-*

- tion, Cambridge University Press, Cambridge, pp 101–140.
- Block N (1986) Advertisement for a semantics for psychology. *Midwest Studies in Philosophy* 10(1):615–678.
- Boghossian P (1989) Content and self-knowledge. *Philosophical Topics* 17(1):5–26.
- Burge T (1979) Individualism and the mental. *Midwest Studies in Philosophy* 4(1):73–121.
- Burge T (1982) Other bodies. In: Woodfield A (ed) *Thought and Object*, Oxford University Press
- Burge T (1986) Individualism and psychology. *The Philosophical Review* 95(1):3–45.
- Burge T (1989) Individuation and causation in psychology. *Pacific Philosophical Quarterly* 70:303–322.
- Chalmers DJ (ed) (2002) *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press, Oxford
- Churchland P (1993) State-space semantics and meaning holism. *Philosophy and Phenomenological Research* 53(3):667–672.
- Crane T (1991) All the difference in the world. *The Philosophical Quarterly* 41(162):1–25.
- Davidson D (1987) Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association* 60(3):441–458.
- Ebbs G (2013) Mendola's internalism. *Analytic Philosophy* 54(2):248–257.

- Egan F (2010) Computational models: A modest role for content. *Studies in History and Philosophy of Science* 41:253–259.
- Egan F (2014a) Explaining representation: a reply to matthen. *Philosophical Studies* 170(1):137–142.
- Egan F (2014b) How to think about mental content. *Philosophical Studies* 170(1):115–135.
- Evans G (1982) *The Varieties of Reference*. Clarendon Press, Oxford
- 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.
- Fodor JA (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press, Cambridge, Mass
- Fodor JA (1995a) Concepts: A potboiler. *Philosophical Issues* 6:1–24.
- Fodor JA (1995b) *The Elm and the Expert*. MIT Press, Cambridge, Mass.
- Fodor JA (1998) *Concepts: Where Cognitive Science Went Wrong*. Oxford University Press, Oxford
- Fodor JA, Pylyshyn ZW (2014) *Minds without Meanings: An Essay on the Content of Concepts*. MIT Press, Cambridge, Mass.
- Frege G (1948) Sense and reference. *The Philosophical Review* 57(3):209–230.
- Gaukroger C (2017) Why broad content can't influence behaviour. *Synthese* 194(8):3005–3020

- Hornsby J (1986) Physicalist thinking and conceptions of behaviour. In: Pettit, McDowell (eds) *Subject, Thought and Context*, Clarendon Press, pp 95–116.
- Jackson F (2003) Narrow content and representation, or twin earth revisited. *Proceedings and Addresses of the American Philosophical Association* 77(2):55–70.
- Jacob P (1992) Externalism and mental causation. *Proceedings of the Aristotelian Society* 92:203–219.
- Kripke S (1981) *Naming and Necessity*. Blackwell, Maldon, Mass.
- Laakso A, Cottrell GW (2000) Content and cluster analysis: Assessing representational similarity in neural systems. *Philosophical Psychology* 13(1):47–76.
- Loar B (1988) Social content and psychological content. In: Merrill RHGDD (ed) *Contents of Thought*, University of Arizona Press, Tucson, pp 99–139.
- Loar B (2003) Phenomenal intentionality as the basis for mental content. In: *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, MIT Press, Cambridge, Mass.
- Machery E (2009) *Doing Without Concepts*. Oxford University Press, Oxford
- Margolis E, Laurence S (2007) The ontology of concepts – abstract objects or mental representations? *Noûs* 41(4):561–593.
- Margolis E, Laurence S (2011) Concepts. *The Stanford Encyclopedia of Philosophy*
- McGinn C (1982) The structure of content. In: Woodfield A (ed) *Thought and Object: Essays on Intentionality*, Clarendon Press, Oxford, pp 207–258.
- Mendola J (2008) *Anti-Externalism*. Oxford University Press, Oxford

- Millikan RG (1984) *Language, Thought and Other Biological Categories*. MIT Press, Cambridge, Mass.
- Peacocke C (1981) Demonstrative thought and psychological explanation. *Synthese* 49(2):187–217.
- Peacocke C (1992) *A Study of Concepts*. MIT Press, Cambridge, Mass.
- Pollock J (2015) Social externalism and the problem of communication. *Philosophical Studies* 172:3229–3251
- Prinz J (2002) *Furnishing the Mind: Concepts and their Perceptual Basis*. MIT Press, Cambridge
- Putnam H (1970) Is semantics possible? *Metaphilosophy* 1(4):187–201.
- Putnam H (1975) The meaning of meaning. *Minnesota Studies in the Philosophy of Science* 7:131–193.
- Rey G (1983/1999) Concepts and stereotypes. In: Margolis E, Laurence S (eds) *Concepts: Core Readings*, MIT Press, Cambridge, Mass., pp 279–300.
- 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.
- Rey G (2010) Concepts versus conceptions (again). *Behavioural and Brain Sciences* 33(2-3):221–222.
- Richard M (2013) Content inside out. *Analytic Philosophy* 54(2):258–267.

- Rosch E (1973) Natural categories. *Cognitive Psychology* 4:328–350.
- Rosch E (1999) Principles of categorization. pp 189–206.
- Rosch E, Mervis C (1975) Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7:573–605.
- Sainsbury M (2014) Fishy business. *Analysis* 74(1):3–5.
- Sawyer S (2007) There is no viable notion of narrow content. In: McLaughlin JD, Brian P; Cohen (ed) *Contemporary Debates in Philosophy of Mind*, Blackwell, Malden, MA, pp 20–34.
- Segal GMA (2000) *A Slim Book About Narrow Content*. MIT Press, Cambridge, Mass.
- Stich SP (1978) Autonomous psychology and the belief-desire thesis. *The Monist* 61(4):573–591.
- Zalta EN (2001) Fregean senses, modes of presentation and concepts. *Philosophical Perspectives* 15(335–359.)