

Why Broad Content Can't Influence Behaviour

Cressida Gaukroger

PREPRINT. PLEASE CITE PUBLISHED VERSION:

Synthese (2016)

The final publication is available at Springer:

<http://dx.doi.org/10.1007/s11229-016-1085-1>

Abstract

This article examines one argument in favour of the position that the relational properties of mental states do not have causal powers over behaviour. This argument states that we establish that the relational properties of mental states do not have causal powers by considering cases where intrinsic properties remain the same but relational properties vary to see whether, under such circumstances, behaviour would ever vary. The individualist argues that behaviour will not vary with relational properties alone, which means that they don't have causal powers. Four replies are presented which all reject the premise that under such conditions behaviour can never be different, and each of these are refuted. The article concludes by arguing that knowing about the relational properties of mental states gives no predictive advantage over (and, in fact, is predictively worse than) knowing about the intrinsic properties of mental states plus context.

Introduction

Do the relational properties of mental states contribute to their causal powers? The answer to this question has significant consequences for the recently revived debate over individualism.¹ If intrinsic properties are the only properties that can explain the power of mental states to cause behaviour, then this provides an important advantage for the individualist, who individuates mental states on the basis of their intrinsic properties. If the relational properties of mental states do have causal powers then individualism has lost the grounds for many of the arguments in its favour.

There are various ways to investigate or establish the causal powers of the properties of mental states. As the significance of the debate between the individualist and anti-individualist relates specifically to mental states' power to cause behaviour, one suggestion has been to isolate the different types of property – intrinsic and relational – and 'observe' whether behaviour changes. I shall refer to this as the 'behaviour isolation' approach. It is argued that where the intrinsic properties of mental states remain the same, but the relational properties vary, *behaviour remains the same*. However, when the relational properties of mental states remain the same, but the intrinsic properties vary, behaviour can vary. This means that the intrinsic, not the relational properties, are doing the causal work, i.e. only intrinsic properties of mental states have causal power.

The value of this test lies in the fact that it does not rely on arguments relating to physicalism, the causal closure of physical systems, or what one would do in 'ideal'

¹Segal (2000); Farkas (2008); Mendola (2008) and Georgalis (2015), for example, have mounted recent defences of individualism or 'internalism'.

science. It applies a scientific method to assessing which properties have causal powers and does so in a way that appeals to folk intuitions. Insofar as one reply to the individualist focuses on the fact that folk psychology broadly individuates mental states, the ‘behaviour isolation’ approach, if it works, indicates that folk intuitions are nonetheless compatible with only the intrinsic properties of mental states having causal powers. This is because what is doing the work in such a scenario is folk-psychological intuitions – intuitions about how people would act if their histories, environment or experiences etc., had been different.

A variety of responses to such an argument have been offered. Several of them concentrate on the premise that behaviour remains the same when the relational, but not the intrinsic, properties of mental states vary. Such responses, which are variants on what I will call the ‘behaviour is different’ reply, argue that behaviour *does* change as the relational properties of mental states change, even where the intrinsic properties remain the same. Some versions of this argument attempt to undermine the structure of the argument: denying, for example, that it is possible to isolate the relational properties of mental states while keeping context the same. Others accept the premise of the ‘behaviour isolation’ approach – that a correspondence between the change in only one property of a mental state and a change in behaviour indicates the causal power of that property – but argue that changes in relational properties *can* correspond with changes in behaviour.

Here I will show that the four main variations of the ‘behaviour is different’ reply all fail, and will conclude that this confirms the validity of the ‘behaviour isolation’ approach and its conclusion – that the relational properties of mental states do not have causal power. I conclude with a brief discussion of the relationship between individuating behaviour and individuating mental states. I further suggest that if

one is motivated by predictive power, then one should adopt an individualist model for individuating behaviour, as it provides marginally more predictive accuracy than its anti-individualist counterpart.

A brief note on terminology before beginning: I will take individualism to be the thesis that mental content supervenes on the intrinsic properties of mental states, and anti-individualism to be the thesis 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. For simplicity’s sake, I will take the intrinsic properties of mental states to be whatever it is that the doppelgängers in Putnam’s (1975) “Twin Earth” thought experiment share. For example, Segal (2000) argues that the properties intrinsic to a thought are the microstructural properties that make up the person (or, more likely the person’s brain) having that particular thought. What is important is that these properties do not involve relations to things ‘external’ to the thinker—things located outside their skin. In contrast, relational properties are those properties an individual’s thoughts have in virtue of features such as that individual’s environment, history, location, or socio-linguistic community. Narrow content supervenes on the intrinsic properties of mental states, while broad content supervenes on (some of) the relational properties of mental states.

1. The Significance of Behaviour Being the Same

If there is a difference in the relational properties of mental states, with no difference in intrinsic properties of those mental states, there will be no difference in the

behaviour caused by those mental states. This has been claimed repeatedly, most notably by Fodor (1987), as a reason to hold that the relational properties of mental states do not have causal power.² It is not denied that the environment in which someone finds themselves may well have an effect on how that person behaves, but only insofar as it is mediated by intrinsic mental properties such as states of their brain.

The assumption that in the event of differences in broad content, behaviour remains the same if narrow content remains the same is written into the hypothetical scenarios that have been used to illustrate and defend anti-individualist intuitions. Putnam's (1975) Oscar and Twin-Oscar continue to have identical 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 him³; and, despite the fact that Davidson (1987) argues

²It is important to note that Fodor's (1987) argument combines numerous different components of which this is only one. The 'behaviour isolation' argument is expressed by Fodor as being about individuation practices in the sciences – that science individuates entities in accordance with (or on the basis of) their causal powers. I am not concerned here whether Fodor's arguments about individuation practices in psychology hold (I believe that they don't).

To motivate his argument, Fodor describes the following case: someone with whom I have never come into contact can flip a coin on the other side of the world, the outcome of which I am completely unaware, and this will cause a change in the relational properties of every part of my organism. I (and every particle in my body) will now have either the property of being in a world where that coin displays heads or a world where it displays tails. However, as the coin flipping has no physical interaction with any of the particles that make up me as an organism, having the relational property of being in a world where the coin either displays heads or tails would not be part of the explanatory apparatus of any of my particles. In other words, a change in which side of the coin was displayed would not change the causal powers of any of my particles, and, by extension, any of my brain states (Fodor (1987), p. 34).

Fodor's example combines two different arguments. The first, a 'behaviour isolation' argument, can be seen as a way to test a hypothesis by isolating features of the scenario; the second makes reference to the fact that there are an infinite number of relational properties. The latter suggests (and this has been challenged, e.g. Burge (1989)) that if you allow that one relational property has causal power you must allow that they all do, which is absurd as we can see from the coin case. However, this needn't be part of the 'behaviour isolation' argument.

³While in Burge's example behaviour will change after the doctor's visit, it will do so because the doctor's response will cause a change in the *intrinsic* properties of the subject's mental states.

for *all* content being broad and, therefore, a *complete* change in broad content faced by his 'swamp- man', both the original Davidson and his swampman replica behave in exactly the same way.

Mental states have both intrinsic and relational properties. Assuming that mental states have causal power over behaviour, at least one of these properties must be doing the causal work. While we are unable to test each kind of property in isolation, by keeping one property constant while varying the other, we should be able to identify whether or not there is a change in causal effects, which would indicate that the property we are manipulating is causally responsible for this change in effects.

The claim that the intrinsic properties of mental states have causal powers is undisputed. Using the isolation method to test relational properties will require looking at cases where intrinsic properties remain the same, and relational properties differ. If in some such cases behaviour varies, then this would mean that the relational properties of mental states have causal powers. If in all such cases behaviour remains the same, then we must conclude that the relational properties of mental states don't have causal powers over behaviour. If the anti-individualist were ever in the business of trying to explain behaviour, this might be a problem.

Peacocke (1981) describes this as 'the objection from psychological redundancy'. This position, he argues, does not need to take a side on whether there is a wider way of understanding a psychological state that would result in our being able to say that the mental states of Putnam's doppelgängers differ. Rather it claims that "these wider psychological states are of no significance for the explanation of your or of your doppelgänger's behaviour. Since your actions will be the same, psychological states in this wider sense are explanatorily inert" (Peacocke (1981), p. 198).

Everything boils down to the question of whether, when the intrinsic properties

of mental states are held constant while the relational properties of those mental states vary, behaviour will also vary. I will devote the large part of the rest of this article to showing that it will not.

2. Variation One: Behaviour is Different Because Behaviour Includes Interaction With the Environment

Consider the case of Oscar and Twin-Oscar – the concepts that they employ when using the word ‘water’ in their respective dialects are intrinsically identical, but relationally different. This makes the Twin Earth scenario a prime case for testing whether the relational properties of mental states have causal powers. The physical movements of Oscar and his doppelgänger are indistinguishable. Does this mean that their behaviours are the same? The first variation of the ‘the behaviour different’ reply argues that, in fact, Oscar and Twin-Oscar exhibit different behaviour. I will refer to this variation as the ‘naïve effects’ reply: quite simply, one is drinking water while the other is drinking XYZ. Depending on how you individuate behaviour, therefore, mental states that are intrinsically identical, but relationally different, may indeed produce different behaviour.

An individual’s context is significant in explaining his or her actions. The question is whether, where two agents are placed in different contexts – for example, if their environments differ – this will change the nature of their behaviour. Peacocke (1981, p. 199) argues that this is exactly what happens.

Let us accept that Peacocke is right in arguing that Oscar and his Twin exhibit different behaviour when they drink the watery stuff on their home planets – does

this show that the differences in their behaviour are caused by the different relational properties of their respective mental states? It is not always illuminating to compare the causal powers of two objects (or the properties of those objects) if those objects are in different contexts. For example, any account of causal powers should be able to accommodate the fact that my toaster has the power to toast slices of bread, even if it has never been plugged in, or has only been tested during a blackout. Put another way, my toaster has the power to make toast, *in the right context*.

The causal powers of an object are best understood across a range of possible contexts.⁴ Take a billiard ball, *B*. In most contexts, when rolled in the direction of another billiard ball the force of *B* will cause any other billiard ball it hits to move; in most contexts, when rolled in the direction of a building the force of *B* will not cause the building to move; and in no contexts will *B* knit a cardigan, make a sandwich, or walk a dog. While we might be able to imagine *B* being rolled towards a building made of cards or jelly, or aimed at a billiard ball that has been nailed to the table, such examples do not mean that the causal powers of *B* have changed. Rather, such alternative world scenarios are included in a set *x*, where *x* in its entirety is the causal powers of *B*. To work out whether *x* applies to *B*, *B* must remain constant between worlds – meaning that its intrinsic properties must remain constant. We will not learn about the causal powers of *B* by doing experiments (whether hypothetical or real) on giant, inflatable billiard balls, though we may get a good idea of *B*'s causal powers through doing experiments on physically very similar billiard balls, and we will learn exactly about *B*'s causal powers by actually observing *B*.

⁴This point is made in Fodor (1987): “[I]dentity of causal powers has to be assessed *across* contexts, not *within* contexts.”

Returning to the example presented above, we can consider the argument that causal powers should be understood across contexts, to analyse the strength of Peacocke's reply. One difference Peacocke highlights is that Oscar drinks water while Twin-Oscar drinks XYZ. However, contra Peacocke, we can see that this difference is not due to a difference in the causal powers of their respective mental states. The difference in their contexts has resulted in a difference in the *effects* of their actions, but as these effects would remain constant in a particular context, independent of which doppelgänger was acting, they cannot be used to illustrate any difference in causal powers of either individual's mental states. If Oscar and Twin-Oscar were suddenly and unknowingly 'switched', Oscar would drink XYZ and Twin-Oscar would drink H₂O and this would be a result of their mental state's intrinsic properties plus their context, as opposed to their relational properties.⁵

No individualist is going to claim that the behaviour of two individuals with intrinsically-identical mental states is always going to be the same. If Oscar and Twin-Oscar had mental states with different relational properties *and* Twin-Oscar was paralysed, behaviour would also vary, but we could hardly put this down to the differences in the relational properties of their mental states. It still matters whether the behaviour of Oscar and his Twin are the same, but it only matters in the case where their context is the same. And when context is controlled for (such as in the switching cases) behaviour does not vary, so the 'naïve effects' reply should be rejected.⁶

⁵See Burge (1988).

⁶Williamson (2000) argues that at least one relational property of mental states – the property of a mental state being *knowledge* – makes a causal difference over the mere combination of narrow properties of a mental state and context. However, as pointed out by Molyneux (2007), the case presented by Williamson to support this argument (involving two burglars, one of which knows that a diamond is hidden in a house, the other only having a true belief that this is the case) does not involve internally identical individuals, making it ineligible as a 'behaviour isolation' argument.

3. Variation Two: Behaviour is Different Where Relational Properties are Recognised

The first variation of the ‘behaviour is different’ reply prompts the individualist to recognise that the behaviour of Oscar and Twin-Oscar is different when they are in their original contexts (original worlds). As argued above, this challenge does not stand because Oscar’s drinking *water* and Twin-Oscar’s drinking *twater* is not the product of the relational properties of their respective ‘watery-stuff’ concepts – rather this is an accident of their respective circumstances. The second variation of the ‘behaviour is different’ reply, the ‘sophisticated effects’ reply, focuses on this aspect of the individualist’s case. This reply argues that we can conceive of a case where the relational properties of the doppelgängers’ mental states *are* causally responsible for the fact that the substances that they are drinking are different.

Burge (1989) believes that he has found such a case. An individual, Chloe, has followed the histories of the two Oscars so that she knows of each man which planet they were raised on. Both men are now in the same environment (a spaceship between the two Earths where there are equal quantities of H₂O and XYZ) and Chloe is tasked with providing drinks for them. Because she knows the histories of the two men, when Oscar asks for ‘water’ she believes him to be asking for H₂O and so brings him H₂O, and when Twin-Oscar asks for ‘water’ she believes him to be asking for XYZ and so she brings him XYZ. Oscar and Twin-Oscar are in the same context, argues Burge (i.e. they are on the same spaceship), and yet their behaviours (their uttering the sounds that correspond with the sentence ‘I would like some t/water’) have different effects (Oscar’s utterance causes Chloe to bring him water, while Twin-Oscar’s utterance causes her to bring him XYZ). So we have a case where

the intrinsic properties of the Oscars' mental states are the same, but the relational properties of their mental states are different, the context is the same, and the behaviour is different. Accepting the premise on which the individualist's argument is based – that isolating the respective properties of mental states will reveal their causal powers – this case appears to show that the relational properties of mental states *do* have causal powers.⁷

It is, however, incorrect to say that there is no variation in context in this case. The Oscars' environment (in terms of their location and the substances contained within it) may remain constant, but Chloe has different beliefs regarding the two Oscars – if her beliefs had been the same, they would both have been given the same substance to drink. What makes this example more sophisticated than its predecessor is that it appears to show that Chloe's different beliefs (and subsequent actions) are actually caused by the relational properties of the mental states of the doppelgängers. This would mean that the relational properties of our mental states *do* have causal powers: they have the power to cause others to treat us differently depending on our individual histories. The point of dispute in this case, therefore, is whether Chloe's beliefs really are caused by the relational properties of the doppelgängers' mental states.

To say that Chloe's mental states are caused by the relational properties of the doppelgängers' mental states, *is to assume that the question of the causal power of relational properties of mental states has already been solved*. As this is the very point being disputed, Burge must give a convincing account of how the relational properties of the doppelgängers' mental states caused Chloe's different beliefs. Notice that,

⁷“Since the individuals have different physical effects in the same context, by Fodor's own test they have different causal powers” (Burge (1989), p. 325).

understood this way, the important question here has nothing to do with whether Chloe's behaviour caused different behaviour in the doppelgängers, or whether her beliefs caused her behaviour. The problem for Burge is that it is not intuitive that it was the relational properties of the Oscars' mental states that caused the difference in Chloe's beliefs. Chloe's beliefs about the doppelgängers have been developed on the basis of the observations that she has made of them, and while she can observe which environment the doppelgängers are in, she cannot directly observe the relational properties of their mental states. Indeed, if Chloe's observations were of the locations of the doppelgängers, then her beliefs developed through observing something that itself *was not caused by the relational properties of their mental states*, upon which she inferred information about the relational properties of their mental states. But in this case relational properties of the doppelgängers' mental states did not in fact *cause* Chloe's mental states, which means that we cannot take 'causing Chloe to have such-and-such mental states' to be one of their causal powers.

There is a further question about whether the differences in the contexts of the doppelgängers described in this case are relevant to judging whether or not their behaviour differs as a result of the different properties of their mental states. Chloe knows which world each doppelgänger has lived on, but if, as she went to get them their drinks the doppelgängers switched their positions on the spaceship without Chloe knowing it, she would now give Oscar XYZ and Twin-Oscar H₂O. When context is controlled, therefore, behaviour remains the same. If one were to argue that, in presenting such a switching case, we removed a *relevant* difference in context, they would have to show not only why the difference in Chloe's beliefs was relevant, but also why the behaviour they cause her to perform is relevant. Imagine the case where Chloe has a strong prejudice against people from Twin Earth (peo-

ple who therefore had beliefs about XYZ). When Twin-Oscar asks for 'water' she is incensed by being asked for something by a Twin-Earthian, using this as grounds for punching Twin-Oscar in the face. Under the same circumstances she would have given Oscar a glass of water. If the initial claim about the causal powers of the relational properties of Twin-Oscar's mental states had worked, one of the causal powers of having 'twater' thoughts would be anger in others, and being punched in the face, but this would be absurd.

4. Variation Three: Behaviour is Different in the Case of Indexical Beliefs

The third type of reply, the 'indexical reply', focuses on the claim that we must test the causal powers of relational properties by isolating context. The objection states that, as soon as context is changed, certain relational properties change, and so this cannot be used as a method for testing the causal powers of relational properties. Evans (1982, p. 203) describes a case where two men with intrinsically-identical mental states are looking for their cat, and both have the thought 'my cat is in this room'. In looking for his cat, the first man is not going to search the second man's room and vice versa. In this case the room to which their thought refers (and the room that they would search when looking for their cat, etc.) will vary depending on context of the person having the belief – though the intrinsic (and, in this case, syntactic) properties of their thoughts remained the same across contexts.

Evans (1982) argues that it is not just the case that an agent's relationship to context is relevant to individuating behaviour, but that our behavioural expectations

are different in cases where we know that two agents have intrinsically identical mental states with different relational properties. The difference in the two men's contexts has not only led to differences in behaviour, but differences relevant for prediction. Peacocke (1981, p. 199) similarly points out that identical indexical beliefs that 'this is the time to do x' will result in actions at different times, if the time when this is thought differs. For the case proposed by Evans (1982), where two men both have the thought 'my cat is in this room', if we merely knew the intrinsic properties of each man's mental states, we would not be able to predict where he was going to look – we would not know that the two men were going to look in different places, for example. This would mean that knowledge of the relational properties of mental states contributes to predictive power, which further suggests that knowing about such relational properties is equivalent to knowing about causally efficacious properties.

In responding to the 'indexical reply', it is important to distinguish between *indexically-fixed* relational properties, and *historically-fixed* relational properties.⁸ Indexically-fixed relational properties are those that will vary with context, such as in Evans' example. Switching cases cannot be used to assess indexically-fixed relational properties as two individuals with the same intrinsic mental states in the same contexts could not have mental states that were different in terms of their indexically-fixed relational properties.

The problem is that no anti-individualist theory of mental state individuation individuates concepts indexically. In fact, anti-individualists focus on historically-fixed relational properties – those that depend on something in the history of the

⁸This distinction has been made, for example, by Jacob (2002) who talks in terms of 'historical extrinsic properties'.

mental-state-possessor, as opposed to her current context. For the anti-individualist who believes that conceptual content depends on real natural categories in the world, it is not just the fact that Oscar is in a world with H₂O that means he has an H₂O concept, but rather the fact that he has historically interacted with water, which has led to the development of his concept.⁹ Similarly, those who argue that conceptual content is determined by an individual's linguistic or social community include a story about the individual's historical relationship to that community.¹⁰ Physical presence in a particular context is not sufficient for the kind of concept possession that is outlined in either of these positions. A change in context will not automatically result in a change in historically-fixed relational properties. So switching cases can still be allowed for assessing historically-fixed relational properties.

Another weakness of the 'indexical reply' is that it does not establish the fact that, where the intrinsic properties of mental states are the same, but relational properties are different, behaviour can differ. It also does not establish the failure of the cross-context test. Rather, it just shows that such a test is inapplicable in certain cases. The most it shows is that it is not possible, with respect to indexically-fixed relational properties, to identify cases where the relational properties of mental states differ, but the intrinsic properties of those mental states and the context remained fixed.

It is worth remembering that the individualist is not claiming that the *relationships* mental states have with the external world are themselves causally inert. The external world is what causes us to have mental states in the first place – it is the

⁹See, for example, Kripke (1981). Chalmers (2002) distinguishes between 'epistemic intension' (narrow content, with extension fixed by actual context) and 'subjunctive intension' (usually broad content, with extension fixed across the space of counterfactual possibilities). Chalmers' account, therefore, may have a role for versions of these two kinds of properties.

¹⁰See, for example, Burge (1979).

source of mentally represented information. If Oscar were to know that water was H₂O and his twin were to know that twater was XYZ then these differing states of the world might well have causal powers over the mental states of the two men. However, the causal powers of a relationship with the external world can be understood in terms of intrinsic properties. To know which indexically-fixed relational properties held for a particular individual would be to know nothing beyond a description of the intrinsic properties of that individual's mental states and a description of their context.

The scope of the 'indexical reply' is so limited that it barely helps the anti-individualist, as it is historically-fixed, not indexically-fixed relational properties that form the basis of almost every anti-individualist account of mental content, and the individualist has not lost their grounds for denying that the historically-fixed relational properties of mental states have causal powers. Furthermore, in section 6 I will show that knowing the indexically-fixed relational properties of an individual's mental states do not give any predictive advantage. The indexical reply, therefore, reveals a weakness of the 'behaviour isolation' approach, but should not be of great concern to the individualist.

5. Variation Four: Behaviour is Different Where Intentions are Different

The first two variations of the 'behaviour is different' reply focused on differences in the effects of behaviour. Neither of these replies worked because these effects only reveal information about different causal powers when context is controlled for. The difference in behaviour in the cases described above was due to differences

in context that were not relevant to assessing the causal powers of the mental states that caused the behaviour. However, the final reply I will consider, which I will call the ‘intentions reply’, points out that behaviour is not merely a physical action with particular effects. While it is the case that when effects vary, behaviour varies, producing different effects is not the only way in which two behaviours can be understood as being different. Adams (1993) argues that the behaviour of Oscar and his twin are different because intentional behaviour includes the intention *as part of the behaviour*. As their intentions are different, their behaviour is different. Adams (1993, p. 54): “[S]ince *bodily movement* is but one component of *behavior*, Al and twin-Al may engage in *different intentional behavior* despite the equivalence of the causal powers of their thoughts *with respect to bodily movements*.”

The idea behind the ‘intentions reply’ is that if there were a ‘switching case’ where Oscar was unknowingly transported to Twin Earth and drank XYZ, he would have drunk it unintentionally (just as if he had drunk vodka or poisoned water instead of water). So, even in the *exact same contexts* the doppelgängers’ behaviours would be different, so long as what we are concerned with is intentional behaviour. Which we are.

One problem with this reply is that it assumes that we should broadly individuate intentions. If you took Oscar and Twin-Oscar to have the same intentions (by individuating intentions narrowly) then their behaviour would once again be the same, as they would both drink intentionally, thereby confirming the individualist’s case. We are, therefore, left with something of a stalemate. The problem is, of course, that the whole exercise is meant to shed light on whether we should individuate mental states on the basis of their intrinsic or relational properties. If we must already have settled the question of how to individuate mental states before we can

run the scenario, then we needn't run the scenario in the first place. The intentions reply, rather than establishing that the relational properties of mental states have causal powers, seems instead to show that any attempt to view behaviour as a guide to causal powers is futile.

There is, however, something counter-intuitive in saying that Oscar, when unknowingly switched with his doppelgänger, behaved differently because he *unintentionally* drank XYZ, where Twin-Oscar in the same position would have drunk it intentionally. Consider the case of Dora who drinks water containing poison, not knowing it to be poisonous. As her drinking action was caused by her intention to drink water, it would both be odd to say that she didn't intentionally drink the water, and it would be odd to say that she intentionally drank poisoned water. Dora's behaviour in this case appears to be appropriately categorised as both intentional and non-intentional. This is troubling as the Adams argument relies on the fact, not that we *classify* certain behaviours on the basis of which intentions we ascribe to the actors, but rather that the behaviour is the behaviour it is due to the intentions of the actor. His point is ontological, not taxonomic. In the Dora case there are several ways of classifying or describing her poisoned-water-drinking behaviour – both intentional and unintentional – but the behaviour itself is either intentional or not.

If, by intentional behaviour we mean behaviour that is 'not forced', then Dora intentionally drank the poisoned water. In such a case, there would be no difference between the behaviour of Oscar and Twin-Oscar when drinking XYZ, as both would do so intentionally. Alternatively, one might want to describe intentional behaviour as that which, had the individual had all the information, they would not have acted otherwise. On this account, Dora unintentionally drank the poisoned water. On such an account of intentional behaviour, you either must conclude that

the doppelgängers would behave the same (if you take the relevant information to be things like that this substance will taste like water, will quench thirst like water, isn't poison etc.), or different (if you take some of the relevant information to be indexical or about their own histories). The former supports the individualist, while the latter makes the hypothetical subjects ineligible for comparison, as to have all the relevant information would end in the two actors having different intrinsic mental states. Insofar as we determine behaviour to be intentional independently of merely listing the intentions that we take to be relevant (which, as we have seen, leads to a stalemate between the individualist and anti-individualist) it appears that the behaviour of people with intrinsically-identical mental states will be the same – if Oscar behaves intentionally, then in the same context Twin-Oscar will also behave intentionally.

Perhaps the point, however, is not about whether the difference in the doppelgängers' behaviour is *merely* about one acting intentionally and the other not, but rather about the nature of their intentions. Even if we must say that both Oscar and his Twin would intentionally drink the watery stuff in their glasses in the same situations, as their intentions are part of their intentional behaviour, and their intentions are different (which the anti-individualist argues that they are), their behaviours are different even if they are both instances of intentional-water-drinking. The point being made here is a weaker one than that which tried to show that, in the same circumstances, one of the doppelgängers would not be behaving intentionally while the other would. But the argument remains that the 'behaviour isolation' test cannot work because what counts as a particular kind of behaviour will be different depending on whether you are an individualist or an anti-individualist.

Let us accept that intentional behaviour includes both an intention and an ac-

tion. Furthermore, having any old intention merely preceding a particular physical action tells nothing about whether that action is intentional – it has to be that the intention *causes* the effect for the action to count as intentional. Intentions are themselves mental states, so this takes us back to the original question of which properties of mental states have causal powers. Or does it? Given the assumption underlying this reply that it is the intrinsic properties of mental states, not their relational properties, that are causally responsible for *physical* movements of the body, take x to be a mental state, and a to be a physical action:

(1) x causes a because of its intrinsic properties. (This is assumed, as per above).

(2) x combined with a is behaviour B. (Following from (1) and assuming, as we are, that an intention forms part of an intentional action).

(3) x has certain relevant relational properties, Rx .

(4) Therefore, Rx is a necessary part of B.

The relational properties, Rx , have to be *relevant* as mentioned in (3) as otherwise *any* difference in the relational properties of intentions will mean a difference in behaviour, which would mean that the same behaviour could never occur more than once, even for the same individual. Furthermore, the anti-individualist is not claiming that any and all relational properties have causal power, or that any and all relational properties are relevant to mental state individuation. They are arguing that there are *particular* relational properties that are relevant – in the case of the

doppelgängers, for example, this would be the property of being raised in an environment where the watery-stuff had a particular chemical compound. (1)–(4) are assumed by the ‘intentions reply’, and the individualist should be happy to accept them all. The thing the individualist rejects, however, is that there is any relational property referred to by Rx .

The challenge for the anti-individualist is to show why some relational properties are covered by Rx – which amounts to showing why some relational properties are relevant to behaviour individuation. With nothing else to go on we can say that certain intrinsic properties of mental states are going to be relevant to behaviour individuation: those properties that cause the physical action part of the behaviour, without which it would not be intentional. As not all relational properties of mental states are relevant to behaviour individuation, the burden of proof is on the anti-individualist to show that some are. So what would it mean for the relational property of a mental state to be relevant to individuating the behaviour of which that mental state is part? It is not relevant in a causal sense, as we agree that it is the intrinsic properties that are doing the relevant work in causing the physical movements. Until this is established, the individualist’s case still stands.

6. Conclusion: A Note on Individuating Behaviour

Let us return to the question of whether behaviour varies in cases where the intrinsic properties of mental states remain the same, but their relational properties differ. As has been shown, while the answer to this question appears to be ‘no’, one of the difficulties faced in providing that answer is that it assumes that we have some understanding of how to individuate behaviour. But the question of whether

two instances of behaviour are the same is a taxonomical question, and therefore one that doesn't necessarily have a correct answer. How we decide to individuate behaviour may vary depending on what it is that we are trying to understand or achieve with such categories, and it is possible to have multiple taxonomies that, in playing different roles, do not come into conflict with one another.

There can still be better and worse ways to develop a taxonomic system – if we classify behaviour in the right ways it can enable us to make predictions of future actions and understand previous actions, and may shed light on the mental states of behaving agents. Consider Peacocke's (1981) argument, that the *doppelgänger*s could be understood as having different behaviour for any number of reasons – their behaviour is directed at different objects, it takes place in different locations, and it could also take place at different times. This suggests that any number of relational properties of mental states could be used in behaviour individuation, and risks sounding like the trivial point that any two token instances of intentional actions will not be the same as one another. When Oscar drinks a glass of H₂O at 5 pm on a Monday, why would he not be behaving differently from when he drinks a glass of H₂O at 6 pm on the same Monday? After all, the chemical compound of the substance drunk might be constant across such cases, but the token substance itself is not the same. As any single change in any part of the universe will result in a change in the relational properties of a mental state in that universe, this point is not just trivial, but holistic. Merely saying that the relational properties of mental states are relevant to the individuation of behaviour is not an account of behaviour individuation, but rather an argument that calls for an independent theory of behaviour individuation.

The anti-individualist can argue that there are obviously some properties that

are relevant when individuating behaviour, and others that are not. The problem for the anti-individualist is how to decide which relational properties of mental states are relevant to the individuation of behaviour. What the anti-individualist cannot do in this case is invoke their theory of concept or mental content individuation to pick out the states of the world relevant to behaviour individuation. Egan (1991) points out that one cannot just say that behaviours are different because there are different intentions behind them, as this would beg the question. Rather, she argues, “the type-distinctness of behaviour is intended as a premise of an argument for the type-distinctness of mental states, rather than as a conclusion from it” (Egan (1991), p. 182).

Even if there were a non-circular method available to the anti-individualist that individuated behaviour in line with their theory of concept or mental content individuation, there would still be the challenge of showing the value of such a taxonomical system. Where we understand behaviour as varying with the relational properties of the mental states of an acting agent, we get reduced predictive power. Imagine I am observing one of the Oscars and I want to make predictions about his future behaviour. It is true that if I know only about the intrinsic properties of the Oscar’s mental states, I will not be able to make predictions about whether he will drink H₂O or XYZ, because I have no way of telling which Oscar I am observing. However, I would still be able to predict that *whoever it is* that I am looking at will drink the watery stuff found in his environment.¹¹ If I know intrinsic properties plus context then I will be able to make the more fine-grained predictions about what kind

¹¹This is a point made by Loar (1985/1996) who describes an example where we read the diaries of two people with intrinsically-identical, but relationally-different mental states. We do not need to know whose diary we are reading, or what the referents of their words are for their diary account to fully explain their actions.

of watery stuff the Oscar I am observing will drink. However, as ‘switching cases’ illustrate, even if I know the intrinsic properties of the Oscar I am observing, and I know the relevant features of his context, this will still not tell me which Oscar I am observing. It could be that the man I am observing is Twin-Oscar who has been switched with his doppelgänger and is therefore on Earth. The only way I could know this would be if I knew the historical-relational properties of his mental states. However, knowing the historical-relational properties of the Oscar I am observing would not give me any more information upon which to make my prediction – it would not enable me to make a better prediction than I could have merely knowing the intrinsic properties of the man I am observing and knowing that he is on Earth. In fact, in the switching cases, knowing the historical-relational properties of the Oscar’s mental states, but not knowing his context would lead to *worse* predictions, as I would predict that Twin-Oscar was going to drink XYZ, not knowing that he was now on Earth. Contrast this with the case where I know the intrinsic properties of his mental states, without knowing his context. Here my predictions might be less fine-grained, but they do not turn out to be false.

In cases of ‘incomplete’ concepts, predictions made on the basis of relational properties are even more likely to be wrong. Again borrowing Burge’s (1979) example, we can imagine a case where Alf goes to an arthritis specialist because he mistakenly believes that he has arthritis in his thigh. Knowing that Alf is wrong about arthritis may help us predict what would happen to him after he goes to the doctor, but if it led to any prediction of his behaviour that differed from the prediction made on the basis of intrinsic properties, it would be getting it wrong. People who have pain in their thighs typically don’t go to the doctor and ask for arthritis medication. Similarly, if we wanted to explain why Alf went to the doctor (in a way

that would lead to future predictions) we would not focus on what he didn't know, but rather on what he believed. Alf thought his thigh pain was cause for concern – that is why he went to the doctor.

Furthermore, changes in context alone might result in our wanting to classify behaviour differently – completely independently of the relational properties of the mental states of the agent acting. If I am the last person to vote in an election it is possible that mine is the casting vote that causes the Prime Minister to lose her job. One might view my voting in this context as being significantly different from my voting in the case where my single vote would have made no difference whatsoever in the outcome of the election – my action in the latter case is merely voting, while in the former it is deciding an election. In this case the historically-fixed relational properties of my mental states, just like the intrinsic properties of my mental states, will not be able to explain the differences in behaviour. The only thing that will be significant in explaining or making predictions about the difference in these cases will be information about context. So variation in behaviour consistent with variation in context is not enough to show us anything about the explanatory power of the relational properties of mental states.

In a final case to illustrate the redundancy of relational properties of mental states to explanations of behaviour, consider Celeste, who has a normal life on contemporary Earth. Unbeknownst to her she has a doppelgänger, Twin-Celeste, who has indistinguishable phenomenal states, and whose brain is identical to Celeste's at every moment. However, while Twin-Celeste has all of Celeste's beliefs, they are all false – Twin-Celeste is in a Matrix-style scenario, where her mental experiences are the product of being hooked up to a network of computers with her body physically located in a pod of which she is unaware. When Celeste kicks a football,

Twin-Celeste only believes that she is kicking a football. This is a case where the intrinsic mental states of the doppelgängers are indistinguishable, while their behaviour clearly differs – in the case of Twin-Celeste there is no physical movement at all accompanying her being in a particular mental state, so no behaviour.

In response to such a case the individualist can point out that such an example does not show that the mental states of Celeste and Twin-Celeste differ in causal powers, even if they differ in effects. If their situations were switched, Twin-Celeste, now in the real world, would behave in exactly the same way that Celeste did when she was in the real world. We cannot know much about the behaviour, behavioural dispositions or (by extension) the mental states of an individual unable to move any part of their body. However, such a case tells us nothing about the causal powers of the mental states of that individual – it is not the fact that Twin-Celeste has a particular series of mental states that explains the fact that she is not moving her body. Rather, what explains the lack of behaviour is something external, but *non-mental*.

One could argue that, in this case, were we to know the indexically-fixed relational properties of Celeste and Twin-Celeste's respective mental states, we would know whether the woman we were contemplating was in the real world or the Matrix. This would not only explain the differences in their behaviour, but it would allow us to make predictions about their future behaviour.

However, it is unclear that we get more or new information from knowing the indexically-fixed relational properties of an individual's mental states. To know what indexically-fixed relational properties a mental state possesses we already have to know the second part of the relation – we already have to know the thing about the world that such properties were supposed to reveal over and above their intrinsic

sic constituents. *We just wouldn't know what mental state they possessed without already knowing what environment they were in.* So we end up with something circular: Celeste has a particular mental state because she is, say, about to drink x . Therefore, Celeste having this mental state is an indicator that she is about to drink x .

It is true that our predictions would be better if we knew more about context, but not in any way that related to an individual's mental states. As we already need to know about context to know the indexical-relational properties of an individual's mental states, knowing which indexical-relational properties her mental states possessed would *hardly* be a better predictive aid, given that the extra detail we get from an account of a situation that includes relational properties of mental states is knowledge *that we already had*.

In this paper I have argued that, even in the face of objections, observing whether behaviour changes when intrinsic properties of mental states remain the same and relational properties are different is a good way of establishing whether relational properties have causal powers. As has been established, in such situations the behaviour will not (indeed cannot) be different, and so we must conclude that the relational properties of mental states do not have causal powers. Indeed, not only are causal powers local, but when it comes to explanations of behaviour we gain nothing by talking in terms of the relational properties of mental states that we would not get from talking in terms of intrinsic properties plus context. Furthermore, when individuating behaviour in accordance with relational properties we see that it is very difficult to do so in an informative, non-circular way, and such a behavioural taxonomy is more likely to lead to erroneous behavioural predictions. The 'behaviour is different' reply to individualism does not stand.

References

- Adams F (1993) Fodor's modal argument. *Philosophical Psychology* 6(1):41–56
- Burge T (1979) Individualism and the mental. *Midwest Studies in Philosophy* 4(1):73–121.
- Burge T (1988) Individualism and self-knowledge. *Journal of Philosophy* 85:649–663
- Burge T (1989) Individuation and causation in psychology. *Pacific Philosophical Quarterly* 70(4):303–322
- Chalmers DJ (2002) The components of content. In: Chalmers DJ (ed) *Philosophy of mind: Classical and contemporary readings*, Oxford University Press, Oxford, pp 608–633
- Davidson D (1987) Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association* 60(3):441–458.
- Egan F (1991) Must psychology be individualistic? *Phil Review* 100(2):179–203.
- Evans G (1982) *The Varieties of Reference*. Clarendon Press, Oxford
- Farkas K (2008) *The Subject's Point of View*. Oxford University Press, Oxford
- Fodor JA (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press, Cambridge, Mass
- Georgalis N (2015) *Mind, Language and Subjectivity: Minimal Content and the Theory of Thought*. Routledge, Cambridge, Mass.

- Jacob P (2002) Can mental content explain behavior? In: A M Galaburda YC S M Kosslyn (ed) *Languages of the brain*, Harvard University Press, Cambridge, MA, pp 91-101
- Kripke S (1981) *Naming and Necessity*. Blackwell, Maldon, Mass.
- Loar B (1985/1996) Social content and psychological content. In: Pessin A, Goldberg S (eds) *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's "The Meaning of 'Meaning'"*, M. E. Sharpe, New York, pp 180-191.
- Mendola J (2008) *Anti-Externalism*. Oxford University Press, Oxford
- Molyneux B (2007) Primeness, internalism and explanatory generality. *Philosophical Studies* 135(2):255-277
- Peacocke C (1981) Demonstrative thought and psychological explanation. *Synthese* 49(2):187-217.
- Putnam H (1975) The meaning of meaning. *Minnesota Studies in the Philosophy of Science* 7:131-193.
- Segal GMA (2000) *A Slim Book About Narrow Content*. MIT Press, Cambridge, Mass.
- Williamson T (2000) *Knowledge and its Limits*. Oxford University Press, Oxford