Lund University
Department of Philosophy

Tropes and Mind

In defense of the trope solution to the problem of mental causation

Author: Erik Jansson

FTEK01 Theoretical Philosophy: Level 3
Bachelor's thesis (15 credits)
Supervisor: Prof. Johannes Persson
May 28, 2013

Abstract

The trope solution to the problem of mental causation combines a trope monism, i.e. that properties are tropes and all tropes are physical, with a type dualism, i.e. that although all tropes are of a physical type, there are subsets of tropes that are also of a mental type. It does so in order to reconcile three individually plausible yet seemingly incompatible principles that together would ensure the efficacy of mental properties in a physicalist framework: (i) That mental properties are at least sometimes relevant to physical events [*relevance*]; (ii) that every physical event has in its causal history only physical events and properties [*closure*]; and (iii) that mental properties are not physical properties [*distinctness*]. Two major objections to the trope solution are addressed: the first claims that the trope solution merely replaces one problem at the level of events with another at the level of tropes and types; the second claims that trope monism is incompatible with type dualism. The first objection is shown to be based on a flawed conception of the trope solution, but it nevertheless forces a concession that opens up for the second objection. In defense of the trope solution it is argued that what the second objection claims to be a denial of the multiple realizability argument – which is what leads to the incompatibility – might actually be a valid response to it.

Keywords: mental causation, properties, trope monism, multiple realizability.

Contents

Introduction 1

I

1. Mental causation: from dualism to physicalism 2
2. Monisms 3
3. The threat of epiphenomenalism 6
4. Three principles of mental causation 7

II

5. On properties 10
6. The trope solution 11

III

7. Noordhof's objection: the *qua*-question revived 14
8. Gibb's objection: monism, dualism and aspects 18

IV

9. In defense of the trope solution 21
10. Conclusion 25

References 26

Introduction

In the borderlands between metaphysics and philosophy of mind abides the problem of mental causation. Metaphysicians want to know the nature of causation and what the roles of events and properties are. Philosophers of mind want to know the nature of the mental and how it relates to our physical bodies. Since the mid-20th century these two lines of inquiry have begun to intertwine and it is in this area of research that the debate discussed in this essay takes place.

How is it after all that a feeling of pain – an intense, subjective and qualitative property of our experience – can cause a conscious intention to alleviate it and a subsequent physical action like rubbing the affected area or, better yet, reaching for the aspirin? Your feelings and intentions appear nothing like your muscles and their movement. So how does the mind and the body interact? In this essay, I will investigate a contemporary approach to making sense of mental causation. There are of course myriads of philosophical problems surrounding this issue, but my interest lies in a relatively narrow field of investigation.

Perhaps the most recent approach to the problem of mental causation is to move the focus from mental and physical events to what the nature of their properties are. I will investigate whether conceiving of properties as particulars rather than universals can serve to make progress towards a solution to the problem of mental causation. The model I examine is David Robb's trope monism. It purports to deliver what has been dubbed the trope solution.

The main question that I will attempt to answer in this essay is: *do the objections raised against the trope solution succeed in refuting it?* My answer will be no. My objective, then, is not to prove the trope solution right, but to show that it has stood up to serious scrutiny.

In the first section (I) of this essay I provide a background to the issue at hand; introduce identity theory and monism; and show why until recently that approach has been seen largely as a failure. In light of this, three principles are introduced that establish the requirements of a working model of the properties of mental causation. In the second section (II) I investigate two accounts of properties – universals and tropes – and go on to analyze a model based on the latter, namely Robb's trope monism. In the third section (III) I turn to the objections to trope monism and the

trope solution and suggest some flaws and misunderstandings in the critique. Finally, in the fourth section (IV), I present my defense of the trope solution where I attempt to expose a flaw in the most important objection that, when corrected, might end up supporting rather than refuting it. Let's now first turn to a short history of the mental causation debate followed by an introduction to the contemporary debate.

I

1. Mental causation: from dualism to physicalism

The pre-philosophical, common sense view of mental causation is often straightforward yet deeply problematic. According to it, there is mental stuff like mind, thoughts and beliefs; and then there is physical stuff like the floor below, the ceiling above and your body in-between. The connection between the mind and the body is what makes it possible for us to interact with the world and yet, upon reflection, this connection is steeped in mystery. In the western tradition, one's mind – or "soul" – has often been said to be in some sense independent of one's body. The tension between independence on the one hand, and interaction on the other reveals the deeper problem.

In the 17th century, René Descartes advanced a kind of *substance dualism*, combining the claim that there is a real distinction to be made between mind and body – that they are made of radically different substances – with the claim that mind and body can nevertheless causally influence each other. Descartes suggested that this interaction occurred in the pineal gland in the human brain (needless to say, modern neuroscience has put this idea to rest).[1] Today, several centuries later, a plausible theory of how Descartes' mental causation would work has yet to be formulated and substance dualism has indeed fallen out of favor in the philosophy of mind.

There are still proponents of substance dualism to be sure – and arguments for and against it abound – but in this essay I will treat that debate only to put it on the shelf. The theories and arguments discussed herein all rest on the assumption of *physicalism,* which means that there is

[1] Lokhorst (2011)

only one substance – the physical. Whatever the *mental* is according to this view, it is part of the physical world. My decision to stay within the confines of physicalism is not an arbitrary one. The ideas and arguments that sparked my interest in the subject all come from philosophers who strive to work the mental into a modern, physicalist framework and the contemporary mental causation debate is almost exclusively played out within it, so it is natural to stay within that same framework.

In the quest for a model of mental causation, physicalism is a huge leap forward compared to traditional dualism since without the great substance divide, the need for an elaborate metaphysical bridge is diminished. But physicalism has problems of its own. Although interaction within one substance is obviously less problematic than interaction between radically different ones, we must be careful not to deprive the mental of its efficacy. On physicalism, mental events are in some sense realized by physical events – and it all happens within a closed, physical system – but this doesn't seem to leave mental events with any causal work to do. At the same time, we must also be careful not to build redundancy into the model by granting both mental and physical events causal sufficiency in every case of mental causation, on pain of overdetermination.

There have been several suggestions as to what the role of mental events is, but I will focus on one kind of suggestion that has kept many metaphysicians and philosophers of mind busy since the mid-20th century, namely that mental events are in fact identical to physical events. It is often referred to as *identity theory*. I will outline an early version of it, called *anomalous monism,* which was advanced by Donald Davidson in the 1970's. Problems in the metaphysics of Davidson's monism are addressed as an introduction into a more recent debate surrounding the updated monism suggested by David Robb wherein the focus shifts from mental events to the properties of mental causation.

2. Monisms

In the article that is central to this essay, David Robb examines the role that mental properties play in mental causation.[2] The question is what, in general, makes them causally relevant and

[2] Robb (1997)

how non-physical (e.g. mental) properties can make a difference in a physical world. The origins of this mental causation debate can be traced back to an influential paper by Donald Davidson[3] wherein he argued for anomalous monism, according to which all mental events are physical events.

On anomalous monism, a mental event ***c*** would in fact be a physical event with both physical and mental properties.[4] I will outline Davidson's argument below, but first I wish to define a couple of terms. I will borrow Davidson's definition of 'event'[5], which fits well into the context of this essay. I use this definition not only for the purpose of presenting Davidson's argument, but for the further discussions to follow as well.

Davidson defines events as "unrepeatable, dated individuals" and gives some examples: "...such as the particular eruption of a volcano, the (first) birth or death of a person [...] the playing of the 1968 World Series". The difference between mental and physical events is summed up by Davidson: "an event is physical if it is describable in a purely physical vocabulary, mental if describable in mental terms."[6] Davidson gives a number of examples of mental events: perceivings, rememberings, judgements, decisions, intentional actions and the changing of belief.

In a causal relation, an event, such as a singer reaching a certain note which causes a glass to shatter, has a property that is *causally relevant*. In the case of the singer, it is the pitch of the note that is the relevant property of causation. I will go into greater detail about properties in my comparison of universals and tropes, but for now a standard, general understanding of 'property' will suffice. Let's now turn to Davidson's argument for anomalous monism:

(i) "at least some mental events interact causally with physical events."

(ii) "where there is causality, there must be a law: events related as cause and effect fall under strict deterministic laws."

[3] Davidson (2001) [orig. 1970]
[4] Robb (1997), 180
[5] Davidson (2001), 210-11
[6] Quotes from Davidson (2001), 210-11

(iii) "there are no strict deterministic laws on the basis of which mental events can be predicted and explained."[7]

(iv) For mental events to be causally related to physical events, they would have to be subsumed by strict deterministic laws.

(v) Events subsumed by physical laws (and thus a physical description) must be physical.[8]

It then follows that mental events are physical events because if they weren't, there would be no way for them to causally interact with the physical given (ii) and (iii). It is of course possible to deny (i), but, as I mentioned at the outset, denying that mental events can interact causally with physical events is exactly what we're trying to avoid. I will revisit this point in a later section about epiphenomenalism.

Davidson is primarily concerned with events and not properties. Although physical events can have mental properties on his view without violating any of the premises of the argument, as Robb points out a physical event with mental properties is not a mental cause unless the mental properties are causally relevant.[9]

When a mental event ***c*** has a physical effect ***e***, ***c*** causes ***e*** in virtue of certain properties it has. Robb illustrates this in the following way:

P M
\ /
c → ***e*** *fig. 1*

[7] Premises (i), (ii) and (iii) are quoted from Davidson (2001), 209
[8] Premises (iv) and (v) are implicit (added for clarity).
[9] Robb (1997), 179

P and **M** are the physical and mental properties of ***c*** respectively, and the diagonal lines between these properties and ***c*** represent the relation of instantiation of those properties.[10] The properties in virtue of which ***c*** causes ***e*** are its *causally relevant properties*: The Sun causes the Earth to stay in orbit not in virtue of its color, but in virtue of its mass, which is the relevant property.

So we have to ask: is it in virtue of its mental or its physical properties that the event is a cause? The answer from Davidson – who denies the kind of psycho-physical laws required for the relevance of mental properties (iii) – has to be that it is in virtue of its physical properties. This follows from the nomological character of causality (ii) and anomalism (iii): The strict laws required for causality gives us the interaction between mental events and physical events only insofar as these mental events have *physical* properties that can fall under such laws. But, as Robb and others have pointed out, that would mean that its mental properties are causally inert and would therefore entail epiphenomenalism.[11]

3. The threat of epiphenomenalism

Perhaps the greatest threat to any model of mental causation is that of epiphenomenalism. If mental and physical properties are distinct, and if the world is a closed causal system of physical events and properties, then any mental events and properties are rendered causally irrelevant. Still, we definitely do not want to deny their existence on pain of denying what Galen Strawson calls "the phenomenon whose existence is more certain than the existence of anything else", namely conscious experience.[12]

And so we're left with a model where mental events and properties silently drift alongside the physical *without any effect on the physical world whatsoever*. My experience of writing this essay, and my feeling of excruciating freedom to do otherwise would have nothing to do with whether or not it gets written. However, the mere fact that these mental properties and events and my awareness of them seem to be at least proximate causes of my writing about them casts at least some serious doubt on the thesis and we are in good company in wanting to avoid

[10] *fig.1* and its description is borrowed from Robb (1997), 180
[11] Robb (1997), 179
[12] Strawson (2006), 3

epiphenomenalism. To do this however, we must ensure the relevance of mental properties and this is what, according to Robb and others, Davidson has failed to do.

4. Three principles of mental causation

The principle of relevance

The problem Robb purports to solve then, is that of fitting this principle of relevance of mental properties into a working model of mental causation. The principle of relevance states that *mental properties are (sometimes) causally relevant to physical events.* The model must also take into account that mental properties and physical properties are distinct while both doing causal work within a closed system. In what follows, I will describe two principles that bear this out and I will then present Robb's completed model.

The principle of distinctness

The principle of distinctness states that *mental properties are not physical properties* (**P** and **M** in *fig. 1*). Mental properties are nothing like mass or solidity, and such physical properties are nothing like "the angst of post-industrial man under late capitalism", to borrow one of John Searle's favorite examples of a mental property.[13] This distinction not only appeals to our common sense, but is important because it resolves the issue of multiple realizability. A certain mental property can be realized by several different physical properties. An example would be the mental property 'pain' being realized in different physical systems such as a human's nervous system vs. that of an ant, say. This is the *multiple realizability argument*, and it casts doubt on any attempt to identify mental properties with physical ones. If the mental property pain in the example were identical to some physical properties of the human nervous system, then by transitivity of identity they in turn would have to be identical to some of the ant's physical properties as well as those of any other creature that was able to experience pain. I will return to this argument in later sections, but for now it's sufficient to understand the implication that there must be some distinction between mental and physical properties.

[13] Searle (2006)

The principle of closure

The principle of closure states that *every physical event has in its causal history only physical events and properties.*[14] Using *fig. 1* to understand *closure* it can be formulated this way: "for any events *c* and *e* and property **P** [of *c*], if *c* causes *e* in virtue of **P**, then if *e*, the effect, is a physical event, then *c*, the cause, and **P**, the causally relevant property, are also physical."[15]

One reason Robb gives for accepting this principle is the simple fact that scientific theories which presuppose it have often turned out to be powerful in explanation and prediction. If assuming *closure* leads to useful scientific theories, then we have reason to believe that the assumption actually tracks something about the way the world is. Robb doesn't give any examples of such theories, but it seems reasonable to suppose that what he's talking about would apply to classical physics. Theories in physics are of course in the business of explaining physical rather than non-physical events, but it is not equally clear that they preclude non-physical properties. At the least the latter is a claim that appears stronger than the former and thus more in need of proper qualification.

The main argument that Robb provides for *closure* however, is that (i) if the instantiation of a property **P** is causally sufficient for an event, then no other property is causally relevant; and that (ii) for every physical event there is an instantiation of a physical property that is causally sufficient for it. We can call these premises (i) *exclusion* and (ii) *weak closure*. From these premises *closure* follows. For *exclusion* to be true, there can be no overdetermination in the system, that is, there can never be more than one sufficient cause for any effect. This seems debatable however. For example, it is not clear how Robb would account for a causal chain *c* -> *d* -> *e* where *c* is a sufficient cause for *e* because it causes *d*, which is also a sufficient cause for *e*. In this case there are two distinct, sufficient causes for *e* which would falsify (i). Perhaps it's more forgiving to interpret Robb as saying that what he's denying is rather two wholly independent (branching rather than linked in a chain) causes *c* -> *e* <- *d* both being sufficient. In the spirit of charity, this is the interpretation I will adopt going forward.

[14] This is also endorsed by Davidson (2001), 223
[15] Robb (1997), 183

As described, the principle of closure seems to not leave any room for mental causation. Even though we can sometimes seemingly trace physical effects back to mental events and properties, there will always be a straightforward explanation that is physical through and through. Consider the following two-part example:

1. You've been hard at work at your desk all morning. A quick glance at the clock on the wall makes you realize that you're late for a meeting. You jump out of your chair and rush out the door. Suddenly there's a lot of physical activity where just seconds ago there was – let's face it – little to none. It seems clear that it is your beliefs[16] about what time it is and about today's schedule that cause your actions.

2. But we can also tell a wholly physical story about what happened: Photons bounced off the clock and landed on your retina, causing nerve impulses to activate certain regions of your brain, ultimately resulting in muscle activity and meeting-attending behavior.

You are not, however, a robot or a Chalmerian zombie[17] and if asked why you went to the meeting you would certainly give a psychological and not a physical account. Indeed, you would be ignorant of most of the underlying physical processes.

Given these three principles, there seems to be no simple way of solving the problem of mental causation. If mental properties are *distinct* from physical properties, and if the world is a *closed*, physical causal system, then it is hard to see how mental properties can be causally *relevant* at all. This is where the trope solution comes in. But before we turn to it, a more in-depth discussion of properties is needed.

[16] Here I'm using 'belief' in the pre-philosophical sense described in an earlier section.
[17] See Chalmers (1996) for a detailed account of this concept.

II

5. On Properties

So far 'property', as it has appeared in this essay, has been ambiguous. There are two ways of conceiving of 'property' that are relevant to the present debate. Consider these sentences:

(i): *"The color of this page is white."*
(ii): *"All the pages of this essay share the same white color."*

So there is the property of the individual page (i); and there is the fact that the other pages share that property in some sense (ii).

The traditional way of making sense of sentences like (i) and (ii) is by invoking universals. Indeed, in the mental causation debate, that properties are universals appears to be an underlying assumption that is rarely addressed[18]. On this view, (i) is true if the individual page is instantiating the universal white (whiteness) and (ii) is true if the other pages are instantiating the same universal as well. In both cases the property is wholly present in the individual(s) and there is no logical inconsistency in it being in two or more places at once. The pages are all are instantiating the same color-universal, so they are identical in color.

The other relevant way of interpreting (i) and (ii) is with *particulars* rather than universals. On this view, what makes (i) true is the individual page partially *consisting of the abstract particular* white. But since we're dealing with particulars and not a universal (whiteness), the pages are not identical in color. This complicates our analysis of (ii) somewhat, because we now have to make sense of how two entities can share the same color while not being identical in color.

A property construed as an abstract particular is commonly referred to as a *trope*. The particular color of this page is a trope and it belongs to a class of similar tropes that confer colors; and a subset of that class of exactly similar (again, not identical) tropes that confer "white" color.

[18] Gibb (2004), 467

Some tropists prefer *concepts* or *predicates* instead of *classes*, but I shall go along with Orilia, who in commenting on the present debate vouches it safe to neglect these variants.[19] All of the works referenced in this essay also assume classes.

In this essay, I will use the word "type" when talking about similarity-classes of tropes – for example color-type tropes. I will also use the word "type" for properties construed as universals in the sense that this page is of a white type. This might appear confusing, but there's a straightforward rationale for it: a type of trope – being a similarity-class of tropes – is analogous to a universal. In the special case of classes of *exactly* similar tropes these classes can be substituted for universals. When talking about this essay's pages being of a white type we are talking about either:

universals: pages that all instantiate the universal whiteness or

tropes: pages all with tropes belonging to the same similarity-class of exactly similar white tropes.

This makes possible a comparison that reveals a strength in the trope-account to which we will return in the next section.

6. The trope solution

By introducing tropes into his model of mental causation, Robb is able to construct a new kind of monism – namely *trope monism* – that purports to be compatible with the principles discussed above. By monism, Robb means that all tropes are physical – there is no dualism of the mental and the physical at the level of events or at the level of properties *qua* tropes. Both the events involved in causation and their properties are constituted by tropes.

So, how does the trope solution show that *relevance, distinctness* and *closure* are not inconsistent? It does so by distinguishing between properties *qua* tropes on the one hand and properties *qua* types of tropes on the other. In this essay, when talking about a trope of type **T**, I will refer to it as a **T**-trope. In Robb's ontology of mental causation there are only tropes[20], and

[19] Orilia (2008), 60

[20] Robb (1997), 186, is sympathetic to the sort of trope nominalism endorsed by Williams (1953) and Campbell (1990). The distinction between concrete particulars and abstract particulars (events and tropes) on this view is quantitative rather than qualitative since concrete particulars are constituted by complexes of tropes (see Williams (1953) for the formal argument).

tropes are physical. They are what constitute the world. In other words, the world consists of **P**-tropes (physical) that belong to similarity-classes that are all subsets of the class of **P**-type tropes with **M**-tropes (mental) being one such subset (see *fig. 2* below). So it follows from trope monism that every **M**-trope is also a **P**-trope.

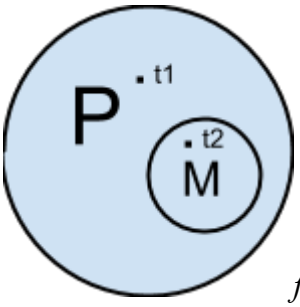

fig. 2: ***t1*** *is a* ***P****-trope;* ***t2*** *is an* ***M****-trope as well as a* ***P****-trope*

There are, for example, color-tropes, shape-tropes and size-tropes (physical type- or **P**-tropes from *fig.1*) etc. As a crude example, when a red color-trope causes you to bring your car to a halt (speed-trope), that is an instance of mental causation where some of the relevant mental properties (i.e. mental type- or **M**-tropes from *fig.1*) of 'you' are your beliefs about traffic law, brakes and the police. These beliefs are all comprised of **M**-tropes.

Since **M**-tropes are also (a subset of) **P**-tropes, they are just as relevant to mental causation as any other **P**-tropes are. This satisfies the *relevance* principle: mental properties are relevant because they are tropes and all tropes are physical. It also satisfies the principle of *closure* because tropes of a mental type are still unproblematically part of the physical chain of events. Then there is the principle of *distinctness*. As we've seen, the principle of distinctness states that *mental properties are not physical properties.* If we read 'properties' as universals, then mental and physical properties would be wholly distinct and we would face epiphenomenalism once again. But if we read 'properties' as tropes and types of tropes, the problem goes away: *distinctness* would merely mean that **M**-types are distinct from non-**M**-types. This would not require any concessions in the handling of *relevance, closure and exclusion*.

It would also steer clear of the problem of multiple realizability that was introduced above. As we saw, theories that identify mental properties with physical properties have the problem of

different physical systems sometimes realizing the same mental property. Different physical systems don't usually share the same physical properties however, which they would have to do if they realized the same mental property, because of the transitivity of identity. But whereas the multiple realizability problem is devastating to an identity theory which posits that properties are universals, it is not as effective against trope monism. This is because the identity between the mental property and the physical property holds only at the level of tropes (they are the same trope). The differences in the physical systems that realize the same mental property can be accounted for by the fact that the tropes belong to different physical similarity classes – they are of different types.

Having distinguished between property *qua* trope and property *qua* type, Robb restates the principles in light of the trope solution:[21]

Relevance: mental tropes are (sometimes) causally relevant to physical events.
Distinctness: mental types are not physical types
Closure: every physical event has in its causal history only physical events and physical tropes.

The principle of *exclusion* that supports closure also gets a trope reading: if trope T's being instantiated is causally sufficient for an event, then no trope distinct from T is causally relevant to that event.

So, we've seen that the trope solution is compatible with the three principles, but it remains to be seen whether it really holds up to scrutiny. In the next section we turn to the two main objections to the trope solution. The first states that it doesn't fare any better than Davidson's model because it only serves to push the problem from one level of analysis to another. The second objection builds upon the first, and maintains that the trope solution is fundamentally flawed unless it is modified in a specific way. Taken together, the objections appear to refute Robb's model since the necessary modification fails to avoid the first objection.

[21] Robb (1997), 188

III

7. Noordhof's objection: the *qua*-question revived

In a reply to David Robb[22], Paul Noordhof compares Robb's solution to that of Davidson. He observes that the main difference between their respective approaches lies in the consideration – or lack thereof – of properties. As Robb pointed out, the crucial flaw in Davidson's model is revealed when one asks the question *is it in virtue of its mental or its physical properties that an event is a cause?*. I will refer to this type of question as *qua-questions.* As we saw in the preamble to my analysis of Robb's model, Davidson's argument leads him to deny mental properties their causal relevance.

Does Robb's trope monism provide the tools needed to succeed where Davidson's monism failed? Noordhof's claim is that it does not, because according to him the trope solution cannot evade the *qua*-question, appropriately conceived:
Q: Is it in virtue of a trope's being a mental trope that mental tropes cause things?. Noordhof purports to show that Robb fails to break any new ground, and that he merely moves the problem from the level of events and properties to the level of tropes and types. If there are both physical and mental types of tropes, and these types are in some way relevant to causation, then we can ask in virtue of which type the trope is a cause. This forces the trope theorist to exclude one of the types from causal relevance by choosing the other. Since this is a variant of the original *qua*-question, the trope solution – as it is interpreted by Noordhof – is in violation of what he calls *the bulge in the carpet constraint:*

"No candidate solution to a philosophical problem should raise another problem which appears just as intractable and which requires the resolution of an issue similar to that which made the original problem so intractable."[23]

This constraint seems commonsensical – philosophy is after all in the business of solving problems rather than replacing them with new ones. However, philosophy is also in the business

[22] Noordhof (1998)
[23] Noordhof (1998), 223

of finding the most fruitful approaches to problems. It could after all be the case that a new problem raised in lieu of an old one has a better chance of being solved. Noordhof claims that we should avoid raising another problem that "appears just as intractable", but appearance can be deceiving. Insofar as the bulge in the carpet constraint takes into account this subtlety, I shall accept it, and in a reply to Noordhof, Robb agrees with it too.[24] So we must ask: does it apply in this case, and if so, what is the new "bulge"?

First, let us examine the question (*Q*) more closely to see if it rests on interpretations of Robb that are fair and correct. *Q* states that "mental tropes cause things" and, upon inspection, we're given no reason in his article to doubt that what Noordhof assumes is that the mental tropes in question are the *causal relata*. In fact, he states explicitly that "tropes are causes".[25] If this is what Robb says, then there might indeed be a new bulge. This is because whereas the cause in Davidson's case was the event, in Robb's case it would be the trope and in both cases we can ask if it is in virtue of its mental properties/type that it is a cause. The fact that all tropes are physical wouldn't help Robb because if a mental trope is a cause in virtue of being a physical type (to not violate closure) then its mental properties would have to be causally inert, just like in Davidson's account and trope monism too would lead to epiphenomenalism.

But Noordhof seems to be missing something about the trope model. One of its strengths is that it is *three-tiered* as opposed to Davidson's model which is *two-tiered*. By 'tiers' I mean levels of similar entities in different models of causation. A tier has no ontological significance but is merely my construct for comparing different models. In Davidson's case, there are these two tiers:

D1. the concrete particular (the event) that is the cause and
D2. its causally relevant properties.

This forces Davidson to read *properties* in the same (the only) way in *relevance, closure* and *distinctness* which leaves the three principles inconsistent – which was Robb's indictment in the first place. In Robb's model however, there are three tiers:

[24] Robb (2001), 91
[25] Noordhof (1998), 222

R1. the concrete particular (the event) that is the cause[26],

R2. its tropes (properties as particulars) and

R3. the tropes' types (properties as similarity-classes of particulars (tropes)).

As described earlier, on trope monism *property* is to be read as 'trope' in *relevance* and *closure* but as 'type' in *distinctness*. Noordhof seems to get this wrong. As I've shown, he assumes that the tropes – and not concrete particulars – are the causal *relata*, i.e. that they do the work that events do in Davidson's model. But this fails to take into consideration Robb's first tier (R1) and actually misses one of the main points of the trope solution. It is R1, and not R2, that corresponds to D1. And it is R3 that corresponds to D2, since – as we saw in the section on properties – types can be seen as both similarity classes of tropes and as universals. R2, then, is a tier of what Robb calls "intermediate entities"[27] and there is no corresponding tier in Davidson's model.

Thinking in tiers makes it easier to grasp the nuances of Robb's model since it is not as clearly stated in his article. Robb seems to realize this need for clarification only a few years later in a reply to Noordhof where he ends up confirming my, rather than Noordhof's, interpretation.[28]

With this clarification in mind, we can turn back to Noordhof's question and see if it applies to the trope solution, properly understood. Since we established that it's the concrete particular that is the causal relata, we must now restate Noordhof's question and not ask in virtue of what a mental trope is a cause, but rather in virtue of what the mental trope is causally relevant: *Q': Is it in virtue of being mental or in virtue of being physical that the trope is causally relevant to the effect?*.[29] What makes the original question posed to Davidson so devastating is that it is asked about events that have both physical and mental properties and Davidson's monism cannot account for the relevance of the mental properties.

In the case of tropes, however, the question seems to lose its force. The trope *is the property* and it is one and the same while being of two types: It is a member of both the similarity-class of

[26] A concrete particular, as opposed to an abstract particular (or trope), is an object or event (etc.) that *has* properties. In the case of universals, a concrete particular instantiates universals (this is what I take to be the most plausible reading of Davidson's view).

[27] Robb (1997), 187

[28] Robb (2001), 91

[29] I've borrowed this formulation of the question from Maurin (forthcoming), 1

physical tropes and the subset of mental tropes, but it is not in virtue of its type that it is causally relevant – it is in virtue of being the causally relevant trope, full stop.

Perhaps there is another level of analysis to which Noordhof's question is pertaining – perhaps tropes have aspects. This would raise a new version of *Q'*, namely *Q'': Is it in virtue of its mental aspect that a mental property (a mental trope) is causally relevant?*. We might wonder what a "mental aspect" is supposed to be. In Robb's reply he takes this to mean secondary properties, that is, properties of properties, which my understanding of Noordhof's objection leads me to think is fair. He then provides an analogy to show why trope monism does not need aspects: *"[A] ball's shape is not roundness in virtue of this or that property it [the shape] has, it is just roundness full stop."*.[30] The property roundness is a particular property (its shape-trope) which belongs to a similarity-class of round shape-tropes and it is in virtue of this shape-trope that it has the shape it has and not in virtue of any secondary properties (aspects) belonging to the shape-trope. Robb simply denies that tropes have aspects according to his model, so Noordhof's objection fails since none of the supposed "new bulges" can be found (by means of *Q, Q'* or *Q''*). At least on its own it fails, but we now move on to an objection that takes into account the concessions made by Robb in his exchange with Noordhof and attempts to deliver a refutation of the whole trope approach.

Before getting into the next objection it is helpful to recap some of what we have now learned about trope monism. It is a monism only at the level of events and tropes because they are all of the same type (physical). There is however a kind of dualism lurking at the level of types (between **P**-tropes and **M**-tropes), in fact it is this type-dualism, as I have called it, that resolves the apparent inconsistency of the three principles *relevance, closure* and *distinctness.* The fact that all **M**-tropes are also **P**-tropes (see *fig.2*) is what keeps the dualism barred from the causally relevant tiers of events (R1) and tropes (R2). Furthermore, we learned that tropes do not have causally relevant aspects but are themselves the causally relevant properties. With this improved understanding of trope monism we may now turn to Gibb's objection.

[30] Robb (2001), 93

8. Gibb's objection: monism, dualism and aspects

S. C. Gibb suggests another problem with the trope solution.[31] She argues that it is not possible to combine a trope monism with a type dualism, because trope monism entails *type monism*. When the trope solution avoids violating the principle of *distinctness* – that mental properties are not physical properties – it does so by appealing to a dualism of types (**M**-tropes and **P**-tropes). If this turns out not to be possible, then Robb's model is indeed in trouble. To see how Gibb reaches her conclusion, let's first get reacquainted with the argument from multiple realizability (MRA).

A mental property, like 'pain' (**M** in *fig.3*), can be realized by multiple, different physical systems with different physical properties: 'pain' in humans (**P1** in *fig.3*) vs. 'pain' in ants (**P2** in *fig.3*) to use a familiar example. The lines between **M** and **P1** and **P2** respectively in *fig.3* represent the realization of a mental property by physical properties.

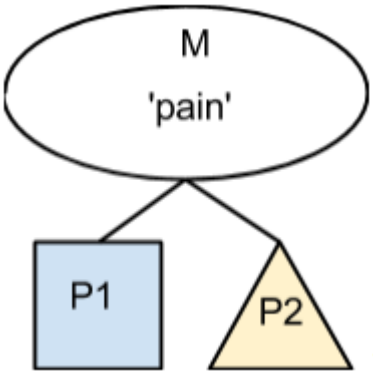

fig.3

If monism is true, and mental properties are identical to physical ones, then by transitivity of identity, the different physical properties that realize 'pain' must also be identical. Furthermore, this must apply to *all* mental properties and across all physical systems the physical properties of which realize them. The sheer implausibility of this state of affairs should lead us to reject the identity between mental and physical properties.

If **M**, **P1** and **P2** are properties *qua* universals, then the problem is obvious: **M = P1**, and **M = P2** entails that **P1 = P2**. But whether you choose the color or the shape of the figures in *fig.3* to

[31] Gibb (2004)

represent the physical property that realizes the mental one, it is apparent that **P1** ≠ **P2** and therefore, given MRA, that mental universals cannot be identical to physical universals.

On trope monism however, the problem is not as apparent. If **M** = **P1** and **M** = **P2** then what this means is merely that **P1** is a physical pain-trope (of a mental type) and that the same goes for **P2**. The fact that **P1** ≠ **P2** is a trivial matter since tropes are particulars and only identical to themselves. **P1** ≠ **P2** is true by definition.

Where trope monism runs into trouble is when it comes to accounting for how it could make sense to talk about 'pain' being the same thing in both cases at all. At the level of particulars (tropes) we're given no reason to think that what it's like to be in 'pain' for one person has anything at all to do with what it's like for another person or other animal. This, incidentally, is not an issue for a universalist model, since my pain and your pain would be the same kind of thing because they are instantiations of the same universal mental property. And so we need a relation between the tropes **P1** and **P2** such that this becomes possible for trope monism too. Similarity is just such a relation. If **P1** and **P2** belong to a similarity-class of mental tropes, then of course my pain is similar to yours. In my discussion of properties I described the analogy between universals on the one hand and types of tropes on the other. In the above case we see how types can give trope monism the benefits of unity that is built into universalist models, without the hassle of identity.

But this is an unsatisfactory solution, says Gibb, and this is where her main argument for trope monism entailing type monism gets its foothold. For trope monism to reap the benefits of unity that universalist models enjoy, the relation between tropes such as **P1** and **P2** needs to be stronger than mere similarity, according to Gibb. It needs to be that of *exact similarity*. To deny this would be to deny MRA.[32] That is, since only classes of exactly similar tropes are analogous to universals, for Robb to appeal to *inexact* resemblance (of the mental tropes) in MRA-cases would not be a valid response to MRA. I will return to this point in a later section and attempt to reveal a flaw in Gibb's reasoning. Now, trope monism does allow for the possibility of exact similarity between tropes. Again, the pages of this essay all have color-tropes that are exactly similar. The problem however, is that if it's right as Gibb says that exact similarity is needed in

[32] Gibb (2004), 471

response to MRA, then trope monism fails just like any other theory that identifies mental properties with physical properties.

To see this, let ***M*** in *fig.3* be the class of exactly similar, mental, pain tropes to which the mental tropes **M1** (which is identical to **P1**) and **M2** (which is identical to **P2**) belong. Since all tropes are physical, **P1** and **P2** also belong to some similarity class(es) of physical tropes. Since **P1** is identical to **M1,** and **M1** is exactly similar to **M2** (both belonging to ***M***), by transitivity of (exact) similarity, **P1** must also be exactly similar to **P2** – they must both belong to the same class of exactly similar, physical pain tropes – call this class ***P***. From this it follows that ***M* = *P*** because since all tropes in ***M*** are exactly similar, they must also belong to the same similarity-classes, and the same is true of all tropes in ***P.*** So both ***P*** and ***M*** must consist of all, and only those, tropes that belong to *both* classes. What this shows is that trope monism is incompatible with type dualism. So, if Gibb's argument is sound, trope monism does fall prey to MRA because it cannot achieve *distinctness* by appealing to type dualism. Thus, if Gibb is right, the trope solution fails.

Gibb goes on to suggest that there is one – and only one – way for Robb to avoid this fate. He could allow for tropes to have aspects. We can see this by importing aspects into the case above: If **M1/P1** has aspects **X1** and **Y1**, and **M2/P2** has aspects **X2** and **Y2**, and aspects **X1** and **X2** (let's call it the pain's mental aspects) exactly resemble each other while **Y1** and **Y2** (the physical aspects) do not, then *distinctness* could again be achieved, only at the level of aspects rather than at the level of types.

But in his defense against Noordhof's critique, and in compliance with *the bulge in the carpet constraint*, Robb himself has already closed that door. Gibb too believes that allowing aspects is not the way forward, not only because it re-invites the *qua*-question as demonstrated by Noordhof, but also because it's ontologically dubious. But it is only if Robb allows aspects that the trope solution can escape Gibb's refutation, so given the two objections represented above, Robb ends up in philosophical checkmate.

IV

9. In defense of the trope solution

First let us delineate Gibb's argument:[33]

(i) In terms of universals, MRA concerns a single mental universal being realizable by different physical universals.

(ii) Only classes of exactly similar tropes can be substituted for universals.

(iii) In terms of tropes then, given (i) and (ii), MRA concerns members of a class of exactly similar tropes of a mental type being realizable by physical tropes that do not themselves exactly resemble one another.

(iv) But given (iii) the trope solution cannot succeed because of the transitivity of exact resemblance – if the tropes in question belong to a class of exactly similar mental tropes, then they must also belong to the same class of exactly similar physical tropes and these classes would be identical.

(v) To claim instead that classes of *inexactly* similar mental tropes can take the place of the tropes in (iii) is to deny MRA[34]. This follows from (i), (ii) and (iii).

(C) From (iv) and (v): The trope solution fails.

According to the first (i) claim, if properties are universals, and mental properties are identical to physical properties, then the physical properties of all systems that realize the same mental universal must be identical. This is of course deeply problematic for any universalist identity

[33] Henceforth I will assume – as both Robb and Gibb does – that tropes do not have aspects. As such, trope monism is at least potentially vulnerable to Gibb's argument and to the multiple realizability argument (MRA) as we saw in the previous section. For a thorough description of MRA, see page 18 of this essay.
[34] This point will be elaborated in a discussion to follow.

theory and since it's grounded in a cornerstone of logic, namely the transitivity of identity, its force is undeniable.

The second claim, (ii), is about something that has been addressed earlier in this essay – that types of tropes are analogous to universals, but only when they are classes of exactly similar tropes. The corollary of this is what's important to Gibb's argument, namely that it is *not* possible to substitute *inexact* resemblance classes for universals. To see that this is true – which I agree that it is – imagine the color-universal *sky blue*. For the sake of argument we'll postulate that this is one very specific color (property) – the one which you can hopefully see if you look out your window. The equivalent to this universal in trope terms is the class of color tropes that are all exactly similar in that they are all sky-blue-tropes. Now let's modify our class of tropes to also include turquoise tropes. Turquoise is very similar to sky blue, but it isn't exactly similar. In universalist terms these turquoise tropes cannot, as it were, be included in the sky blue universal – turquoise must be conceived of as a universal of its own. However close the inexact similarity of tropes in a class, any difference between them will break the analogy with universals since they can then only be accounted for in terms of different universals.

The third (iii) and the fifth (v) claim bears this out: Trope monism must approach MRA as an argument against identifying mental types and physical types where these types are construed as classes of *exactly* similar tropes, on pain of denying MRA. As we saw in (iv), this renders trope monism just as vulnerable to MRA as a universalist identity theory because transitivity works the same way for exact resemblance as it does for identity.

But what exactly does Gibb mean by *denying MRA*? And why would it be a problem if the trope solution denies MRA as it is stated in (iii)? It seems to me that Gibb gives insufficient reason for us to accept that this is a problem for trope monism.[35] The only reason she gives is that sets of inexactly resembling tropes are to be substituted within a theory of universals for *different* universals, which we saw in the sky blue / turquoise example above, and that this somehow makes it ill-fitting for MRA. This, however, presupposes the specific reading of MRA in (iii),

[35] See Gibb (2004), 471, 475, for her mentions of this issue.

and I will now attempt to show that the tropist is in fact not required to yield to that version of the argument.

To see this, first we must understand the ontological significance of MRA. Traditionally it's been leveled against identity theories where properties have been assumed to be universals. 'Pain' is the staple example of a mental universal, and it's one that I myself have used in this essay. For a universalist identity theory, the fact that different creatures can be in pain *qua* the same mental universal is a problem because their respective physical properties that realize pain are not identical, whereas their mental 'pain' property is, and this violates transitivity of identity. But let's say that we try to resolve this by postulating slightly different 'pains' so that the mental property 'human pain' is not identical to, say, 'dolphin pain'. Perhaps the slight difference between these mental 'pain' properties could account for the slight difference in neurology (the relevant physical properties). This, however, is not a move that the universalist is free to make because just like in the sky blue / turquoise example, any slight difference between two properties renders these properties as *different* universals and that *would* deny MRA. But while this move is not open to the universalist, I believe that it *is* open to the tropist. On trope monism, 'human pain' and 'dolphin pain' are both tropes that belong to the same class of (very) similar mental tropes. So the tropist can say that tropes from different (i.e. not identical) classes of similar physical tropes can realize tropes that belong to *the same* similarity class of mental tropes. This would then pass the test of MRA.

Is this denying MRA as claimed in (v)? What I take to be the denial of MRA that Gibb warns us of in the case of inexact similarity is that if we don't talk about either one and the same universal, or whatever the equivalent of it would be within another theory (see (ii) above for example), then we cannot account for the fact that we're talking about the same thing when we're talking about for example 'human pain' and 'dolphin pain'. While this is trivially true in the sense that we're not talking about the same *qua* identical thing, it is not clear that *very close similarity* is not a good enough substitute for identity in MRA when we have tropes in the picture. I would argue that it is at least as intuitive to speak of pain in terms of close similarity as it is to speak of it in terms of identity – and trope monism gives us the freedom to do so.

Compare this to a trope account of nuances of red color. Scarlet and crimson are very similar, but not exactly so. Both nuances belong to a class of inexact resemblance that we call 'red'. There is no doubt that when you see a scarlet color, and I see a crimson color, we can both say that we're talking about red, rather than blue, colors. Let's now apply MRA to this state of affairs. Red, as described above, is multiply realizable. A scarlet dress and a crimson rose both realize tropes in the same class of red tropes. The physical properties that realize this red type are not exactly similar however (scarlet tropes and crimson tropes). So even though the scarlet trope is identical to a trope in the class of red tropes (they are the same trope), and the crimson trope is identical to a trope in that same class (they are the same trope), there is no transitivity of exact similarity, because the class of red tropes is a class of *inexactly* similar tropes. We can both nevertheless still say that the dress and the rose are red. While this is perhaps not a perfect analogy to a human and a dolphin both being in pain despite not having exactly similar physical properties that realize the pain, it should serve to make my point.[36]

Even the universalist could drop the identity requirement (one universal) and instead talk about slightly different pains (different universals), but that would be problematic for another reason since, in the limit, there would be different mental 'pain' universals for every single physical system that can realize pain. In other words, it would actually lead to a trope-like account (add classes of similar universals and you have the complete conversion).

Employing tropes and similarity-classes in response to MRA is not denying the argument if 'denying' means avoiding or sweeping it under the rug. On the contrary, proponents of trope monism are free to accept MRA and even to accept that it is as devastating to trope monism as to any other identity theory *if* types are read as exact similarity classes. Appealing to inexact similarity is not a denial of the argument, it is a lesson learned from acknowledging it. And I believe that it is a valid response to it – perhaps the only one.

So, whereas it makes sense to use the traditional MRA as stated in (i) against universalist identity theories, we have not been given reason enough to accept that trope monism must go the same way. If the significance of MRA is to show that a theory must be able to account for the fact that

[36] Robb (2012) argues along the same lines in his recent reply to this kind of objection.

different physical properties can realize the same mental property – an intuition that I share – then we should take into account the different tools with which different theories can do this. As I've tried to show, trope monism is uniquely qualified to do so by utilizing inexact similarity classes of mental and physical tropes. If we allow for inexact similarity in MRA – as I've argued that we can and should – then we have no reason to accept that trope monism denies MRA, nor do have any reason to accept that trope monism succumbs to MRA or that trope monism is incompatible with type dualism.

In sum, Gibb fails to argue for the truth of her conclusion that the trope solution fails.

10. Conclusion

David Robb's trope solution to the problem of mental causation purports to reconcile three individually plausible but seemingly incompatible principles of mental causation: That mental properties are at least sometimes relevant to physical events [*relevance*]; that every physical event has in its causal history only physical events and properties [*closure*]; and that mental properties are not physical properties [*distinctness*].

The solution is to combine a trope monism with a type dualism. By reading mental properties as tropes (which are fundamentally physical), *relevance* and *closure* can be achieved while *distinctness* is ensured by reading properties as types of tropes, e.g. in response to the multiple realizability argument.

In this essay, two objections to the trope solution were addressed. The first attempts to show that trope monism fails to make any progress over other identity theories and that it instead replaces one problem with another, similar problem. It turns out that this objection fails against the trope solution properly understood. One concession made by Robb in responding to the first objection – that tropes do not have aspects – opens up for a second objection: If tropes do not have aspects, then trope monism is incompatible with type dualism and the trope solution fails.

The main thrust of this second objection is that the trope solution must either deny or concede the multiple realizability argument. I attempted to show that the 'denial' option is in fact not a

denial, but rather a valid response to the multiple realizability argument properly conceived. I have not shown, nor have I attempted to show, that the trope solution succeeds on all counts in solving the problem of mental causation. I have however tried to demonstrate that the objections addressed herein fall short of refuting it.[37]

References

Campbell, Keith. *Abstract Particulars* (Oxford: Blackwell, 1990).

Chalmers, David. *The Conscious Mind: In Search of a Fundamental Theory* (Oxford University Press, 1996).

Davidson, Donald. "Mental Events", in *Essays on Actions and Events* (Oxford University Press, 2001). [orig. 1970]

Gibb, Sophie. "The Problem of Mental Causation and the Nature of Properties", *Australasian Journal of Philosophy*, Vol. 82, No. 3 (Sep. 2004), pp. 464-476.

Lokhorst, Gert-Jan, "Descartes and the Pineal Gland", *The Stanford Encyclopedia of Philosophy* (Summer 2011 Edition), Edward N. Zalta (ed.). Available from: http://plato.stanford.edu/entries/pineal-gland/. [Accessed: 21 May 2013]

Maurin, Anna-Sofia. "Tropes and Issues in the Philosophy of Mind [5.2]", *The Stanford Encyclopedia of Philosophy*, [forthcoming]. Available from: http://www.flov.gu.se/digitalAssets/1445/1445652_tropes.pdf. [Accessed: 18 May 2013]

Noordhof, Paul. "Do Tropes Resolve the Problem of Mental Causation?", *The Philosophical Quarterly*, Vol. 48, No. 191 (Apr., 1998), pp. 221-226.

Orilia, Francesco. "Basic Ontology, Multiple Realizability and Mental Causation", in *Tropes, Universals and the Philosophy of Mind: Essays at the Boundary of Ontology and Philosophical Psychology*, Gozzano, Simone and Orilia, Francesco, eds. Vol. 24. Ontos Verlag, 2008

Robb, David. "The Properties of Mental Causation", *The Philosophical Quarterly*, Vol. 47, No. 187 (Apr., 1997), pp. 178-194.

Robb, David. "Reply to Noordhof on Mental Causation", *The Philosophical Quarterly*, Vol. 51, No. 202 (Jan., 2001), pp. 90-94.

[37] I would like to thank Johannes Persson for his help and invaluable input, and Anna-Sofia Maurin for introducing me to the debate and the surrounding literature.

Robb, David. "The Identity Theory as a Solution to the Exclusion Problem", in S.C. Gibb, E.J. Lowe, and R.D. Ingthorsson, eds., *Mental Causation and Ontology.* (Oxford University Press, 2013), pp. 215-232.

Searle, John. Pufendorf Lecture. (2006). *Lecture 1 - Consciousness*. [Online Video]. 19 June. Available from: http://www.pufendorf.se/2006/pufendorf_2006_day1d.wmv. [Accessed: 16 May 2013].

Strawson, Galen. *Consciousness and its Place in Nature* (Exeter: Imprint Academic, 2006).

Williams, D. C. "On the Elements of Being: I", *The Review of Metaphysics* (1953), pp. 3-18