The Physical Action Theory of Trying

David-Hillel Ruben

Birbeck University of London
d.ruben@bbk.ac.uk

Abstract

Metaphysically speaking, just what is trying? There appear to be two options: to place it on the side of the mind (or brain) or on the side of the world (where the latter includes the agent's body). Volitionists, who think that to try is to engage in a mental act, perhaps identical to willing and perhaps not, take the mind-side option. The second, or world-side option identifies trying to do something with one of the more basic actions by which one tries to do that thing. The trying is then said to be identical with the physical action.

After carefully stating the second, world-side view, I produce two arguments against it. The first relies on the fact that if $a=b$ and $b=c$, then $a=c$ (the transitivity of identity), sometimes put colloquially as: if something is identical to two things, then the two things must be identical to one another. In the case of trying, one might try to do something by performing a plurality of simultaneous actions, a sure sign that the relation between the trying and the plurality of actions by which one tries must be some relation other than identity.

The second argument discusses two cases, recorded in William James' *The Principles of Psychology*, of a patient who tries but who performs no action whatever. This is sometimes called 'naked trying'. A recent attempt at denying that there can be such cases of naked trying is examined and dismissed.

The article concluded with what I think is a brief restatement of the main thesis about what trying is in Gideon Yaffe's *Attempts*.

Metaphysically speaking, just what is trying? There appear to be two options: to place it on the side of the mind (or brain) or on the side of the world (where the latter includes the agent's body). Volitionists, who think that to try is to engage in a mental act, perhaps identical to willing and perhaps not, take the mind-side option.[1] Such a mind-side view might be compatible with physicalism, if the mental act is then identified with a brain state. Hobbes took this view: 'These small beginnings of Motion, within the body of Man, before they appear in walking, speaking, striking, and other visible actions, are commonly called *endeavour*' (*Leviathan*, chapter VI). The contrast I want is not between the mental and the physical per se, but between what is 'inside' the body and what is at the body's surface or outside of it.

The second, or world-side option identifies trying to do something with one of the more basic actions by which one tries to do that thing. Taylor (1973, p. 79) is an example of a philosopher who takes this option: he says, '[T]rying to do something is always. . . actually doing something with a view to accomplishing a certain result'. George Wilson and Timothy Cleveland hold roughly the same view and I shall use Wilson's statement of it[2]: 'to try to φ is to perform an action that is directed by the agent at (in an appropriate way) promoting his φ–ing' (Wilson 1989, pp. 155-156). Wilson calls his view 'the omnipresence of physical action in all trying (i.e., in all cases of trying to perform a physical action.' (159). The trying is then said to be identical with the physical action.

The purpose of this paper is to argue against the second, world-side option. I won't explore here the viability of the mind-side options, nor will I query whether these two options really exhaust the alternatives (to set the record straight, I don't think that they do; I am not a volitionist either).[3] But if I succeed in dismissing the world-side option, we will have made considerable progress in coming to a view of what trying is, because we will have learned something important about what it is not. We will have blocked the most obvious alternative taken by those who don't want trying to be placed on mind-side.

1 The Physical Action Theory of Trying. The Multiple-Realisability Claim

The initial world-side thought is that if agent A wants some end E, and so he tries to obtain E by some physical action, say, by his ψ–ing, his trying to obtain

---

[1]Hornsby (1980) identifies trying as an internal event ('..[tryings]are always internal events' (1980s, 45) and hence places tryings on the mind-side. Other volitionists include McCann (1998), Pietroski (2002), O' Shaughnessy (1980), Ginet (1990).

[2]Cleveland thinks that the trying is a physical movement: when the agent tried, he 'performed a movement, and intended of that movement that it be a raising of his arm' (Cleveland 1997, p. 142).

[3]See Ruben (2013, 2015).

E = his ψ–ing. But the world-side view is rarely described in any significant detail. I think the world-side view is best expressed as the conjunction of (1) and (2) and I shall explain why below:

1. If one can try to do something by doing some action, there may be indefinitely many physical action types by whose tokens one can try, i.e., whose instances can realise that same type of trying on the same or different occasions.

2. Every token trying is identical to some token physical action of one of those physical action types, by the token of which the person tries.

Whether there is more than one action by which one can try to do something is a contingent matter. Even if there were really only one way in which to skin a cat, there are as many ways in which to try to skin it, as there are beliefs, false or true, about how to skin it. Even if doing ϕ is really no way at all in which to skin the cat, if a person believes that it is, then he can try to skin the cat by ϕ–ing (my goodness, he tried to skin the cat by ϕ–ing. How silly. He was bound to fail). Still, (1) allows for the multiplicity of ways by which one can try, but it does not require such a multiplicity.

The view of trying expressed by (1) and (2) is tailor-made for the case of non-basic action, so that one tries to do a non-basic action by doing a relatively more basic one. But what of basic action itself, action one does but not by doing anything else? The thought must be that trying to do a basic action is identical to the basic action itself, there being no obvious physical action alternative. If one tries to ϕ, where ϕ is a basic action, then the token trying to ϕ = the token ϕ-ing. Either the case of basic action needs to be added as an explicit (3) above, or alternatively we could understand the by-relation as non-reflexive rather than irreflexive, allowing all basic actions to be done by themselves.

Consider a type like 'trying to stop the flood'. On different occasions of flood stopping, doing different things might stop the flood. Agent A on one occasion might try to stop the flood by his starting of the pump (hereafter, his starting), agent B might on a different occasion try to stop it by his lowering of the flood barrier (hereafter, his lowering), and on a third occasion agent C might try to stop the flood by his closing of the sluice on a dam (hereafter, his closing).

No type of trying, like trying to stop the flood, can be identical to any one of the physical action types, by whose tokens one can try. The relationship here is well known from the philosophy of mind more generally: the multiple-realisation relation.[4] What one can say on the physical action theory is that a type, like trying

---

[4]The core of the theory uses the multiple-realisation relation, and in this context it is much less problematic than its use in the brain-mind identity theory. Some critics of the multiple-realisation argument, as it occurs in the brain-mind identity theory literature, have suggested that it arises only

to stop a flood, can be realised by many different physical action tokens on different occasions, each of which may be an instance of a different physical action type. Yesterday, trying to stop the flood was realised by a token starting of the pump, realised today by a token lowering of the flood barrier, to be realised tomorrow by a token opening of the sluice. Indeed, as I indicated above, given oddball enough beliefs about what causes what, a token of almost any type of physical action can on some occasion realise a type like trying to ϕ. In the case I have sketched, the type is realised by three token actions, each a token of a different physical action type. Hence, (1).

2 The Physical Action Theory of Trying: The Token-Token Identity Claim

What is the relation between a token trying and the token actions that realise the trying? Consider now only a single occasion of flood-stopping by a single agent. In a simple case, the agent might try to ϕ by doing a single token action, ψ-ing (say, starting the pump). So the physical action identification theory would have it that his token trying to ϕ = his token ψ-ing. Hence, (2). (For the sake of simplicity, I omit time-indexing.)

---

as an artificial creation of describing the mental type at one level of abstraction or generality and describing the brain state at a different, more specific level of generality. The suggestion is that were we to insure that the levels of generality in description match, we might find that we had a type-type identity theory of mind and brain after all. Another reply in favour of retaining type-type identity is to restrict the type identity to individual species, thereby making the identification more plausible.

These issues do not appear to affect the use of multiple-realisability in the case of trying. The level of generality required both for the content of the trying and the action by which one tries is set by the intentions or plans of the agent and hence both are taken from the personal level. The agent will have a plan to try to do some action, and that plan will include the salient steps he needs to take in order to try and obtain the goal or end. In the more general mind-brain use of multiple-realisability, there is a question of the right level of generality needed for the scientific description of the brain state, and the consequent need to match up the right personal and impersonal levels. But no such problem arises in the case of trying-both the action and the trying are located in the personal level of description. The comparison between uses of multiple-realisability in mind-brain theories and trying theories is instructive. There appear to be two different uses of multiple-realisability, a horizontal use and a vertical use. Its use in mind-brain identity theory is a vertical use: something from one level ('the personal') is said to be multiply realised by many things at a different level, the impersonal level of neuroscience. Since each level has its own discourse, a question can arise as to how to match types from the two discourses. The use of multiple-realisability in the trying-physical action theory is a horizontal use: different levels are not involved in the sense that both the action descriptions and trying descriptions are part of the discourse of folk psychology. As far as species specificity is concerned, to whatever extent some specific species of animal can try and act, it remains true that for at least some cases of the animal's trying, there can be more than one way in which the same animal can try to do that thing. The application of multiple-realisability in a horizontal use is much less problematic than in a vertical use. These issues are covered in (Gozzano & Hill 2012, pp. 129-141).

But many cases are not like that. There are many cases in which A tries to do something by doing numerous things. Unlike a simple case like trying to open the door by turning the handle, many action plans will typically have many nodes on them, with the end that one is trying to achieve as the last node on the chain: construct the hull outline, fix the planks, put pitch in the joins, paint the result, fix a mast, run up a sail, and so on, and finally-voila!- a ship. Each node is a separate action on its own. The action plans most loved by philosophers of action are ones where each action is accomplished by its prior by-action: he reduced the world's population by one by killing the Queen, he killed the Queen by shooting her; he shot her by firing the gun; he fired the gun by pulling the trigger; he pulled the trigger by bending his finger. The action plans under discussion here are NOT like that. No one fixes the mast by painting the result, etc. The example of an action plan under discussion is one in which the agent must take a series of (usually) successive and independent actions in order to achieve some end. Each action on the plan is not related to its predecessor by the by-relation (except insofar as the agent may aim for the final result by doing all of the other actions equally).

In our example, Agent A might try to stop the flood (a) by his starting of the pump, and (b) by his lowering of the flood barrier, and (c) by his closing of the sluice on a dam. In this case, let us suppose that the same agent must do three separate actions, each of which is necessary but insufficient to stop the flood (and he knows this). Imagine that none of the actions on its own can even partially stop the flood. Suppose a threshold effect so that all three actions must be undertaken, in order for any stopping at all to occur. In this case, there is one token trying and three token actions taken in order to try.

On whatever theory of act individuation one might chose to adopt, these are three actions and not three descriptions of one action. The agent simply does each particular action seriatim, and the order in which he does them may not matter. He tries only once to stop the flood, but he must do three quite different things (let us suppose) in order to fully try even that one time. So he tries to stop the flood by doing three distinct actions. What is the relationship between the token trying to stop the flood and the three token actions the agent takes? On the physical action theory of trying, with what physical token action would the token trying be identical?

Choice of one rather than the others of the three token actions with which to identify the token trying would be unmotivated. The equal apparent claim of all three is an indication that the relationship between the token trying and each of the several distinct token actions, whatever it might be, is not identity. The trying≠the lowering, the trying≠the closing, and the trying≠the starting. What other relationship might we substitute in place of identity, as a relation between the token trying and the multiple token physical actions?

An alternative suggests itself. Tryings go on for a period of time-'he tried for several days'-and as such they have parts; they are the temporally restricted slices of the whole trying. If an agent tries from t_1-t_{10}, and if the trying has parts, then the parts of the trying would be the temporal segments of it: the trying from t_1 to t_2, from t_2 to t_3, and so on.

A natural suggestion then, for the identification theory, would be to identify each of the temporal slices of the trying with a different action on the action plan. So the temporal trying slice from t_1-t_2 might be identical to (say) the lowering of the flood barrier. Each of the different token actions on the action plan could be identical to a different trying temporal slice.

Most of us can walk and chew gum at the same time. The temporal slice strategy won't work if the agent engages in two or more of the independent actions at the same time rather than seriatim. With one hand, the agent lowers the flood barrier; with the other, he closes the sluice. The same argument applies as before. The two actions, the lowering and the closing, are not identical, so the temporal slice can't be identical to both, and neither has a better claim than the other to the identification. What option is left to the identification theorist? Perhaps this: the appropriate temporal slice of the trying would need to be identical to the mereological sum (or fusion) of the two actions.

The mereological sum strategy has its own problems. Two of the axioms central to extensional mereology are (a) the axiom of unrestricted composition (AURC) and (b) the axiom of the uniqueness of composition (AUNC). The first axiom asserts that whenever there are some things, there is a sum of those things. The second says that two sums, α and β, have exactly the same parts iff $\alpha = \beta$. (the AUNC has many different formulations. I have written it as a biconditional to make its extensional implications more salient).

The (AURC) raises a question that the mereological proposal would need to address. That proposal claims that some time slices of trying are sums of actions. It certainly does not say that every sum of actions is a time slice of a trying. However, while the person lowers the flood barrier and closes the sluice, he may also deliberately wink at someone (after all, both of his hands are fully engaged). But although there is, according to the (AURC), a sum of the winking and the lowering, that sum is not identical to any time slice of the trying. So the question, which I shan't pursue further here, is why the one sum of two simultaneous actions but not another is identical to a trying time slice.

The (AUNC) entails that, like sets, sums or fusions are modally fragile. Two non-identical sums cannot have the same parts. Sum α cannot gain or lose a part without ceasing to be. Sum α merely gets replaced by a different but very similar sum β, which differs from α by having at least one different part.

Ordinary physical objects are not mereological sums because they do not obey

the axiom of uniqueness of composition. On most accounts of objecthood, ordinary objects are not modally fragile. They can gain or lose at least some parts. Ordinary objects are modally robust.

Are ordinary events and actions, like tryings (if, on the assumption of the argument we are addressing, they are actions), modally fragile or modally robust? They are not as robust as physically objects but on the other hand they do not seem to be as fragile as mereological sums.[5] Helen Steward (2013, pp. 802-809) thinks that processes are modally robust, but that events are modally fragile. I am not interested in debating the alleged event/process distinction, but I think that what she says about processes is true of events, at least as ordinarily understood.

The Great War could have had one less battle than it did have, and still it would have been the numerically same Great War. Imagine that two disputants are disputing whether or not the Great War included some battle, the Battle of the X. If the Great War were a mereological sum, the two disputants would not even be disputing about the same world war; they would have two numerically different although very similar world wars in mind.

If a time slice of a trying were identical to the mereological sum of the individual actions the agent undertakes during the temporal extension of that time sclice, and the whole trying in turn were a mereological sum of the time slices, then any addiction or subtraction of an individual action in the sum that could have occurred would destroy the numerical identity of the trying. Sir Edmund Hilary tried to climb Mt. Everest (and succeeded on 29 May 1953). Suppose two disputants are arguing whether Hilary forged a route through the treacherous Khumbu Icefall (the forging).

If the mereological sum view were correct, they can't be disputing about the same attempt by Sir Hilary. On the mereological sum view, one is probably speaking about the actual attempt; the other, about a similar possible but non-actual attempt that he thinks was the actual attempt. That doesn't seem correct. They are arguing about one and the same attempt and disagreeing about whether it included the forging. Of course, there are some changes that attempts could not survive without losing their numerical identity, but then there are some changes physical objects cannot survive either without losing their numerical identity.

I acknowledge that intuitions about the modal robustness of token events and actions differ. It is generally accepted that they are not continuants. The person determined to use a mereological model for grasping the way in which the sum of two actions compose a time slice of the trying might simply insist that in the examples above, there are two different world wars or two different attempts at mountain ascent under discussion, one actual and the one merely possible. So let me offer a further, second argument against the mereological solution to the

---

[5]Simons (2000), who disagrees.

difficulty I have identified.

The thought was that the time slice of the trying = the sum of the lowering and the closing. Suppose that the lowering *must* happen before the closing in order to be effective. In this case, the view would be that the trying from t_1-t_2 = the sum of the lowering at t_1 and the closing at t_2. And suppose that at t_2 when he closes, he lowers again (ineffectively of course as the lowering that counts is the one that occurred at t_1). So now we have at least these sums to contemplate: the lowering at t_1 and the closing at t_2; the lowering at t_2 the closing at t_2; the closing at t_2 and the lowering at t_1. Mereologically speaking, the first and third of these sums are the same sum of course, since they have the same parts.

Does mereology have the resources to explain why the first and third sums are different at least for our purposes (because of the order of the parts) or why the second sum couldn't a part of the trying? Following an argument of (Fine 1999, p. 66), it might be suggested that what the first sum would need to distinguish it from the third is the inclusion of a trope of ordered successiveness, absent from the second sum altogether and reversed in the case of the first and third.

But, as Fine argues, tropes can't be parts of mereological wholes in the same way as events (or substances) are. The whole mereological project is simply misguided. The part-whole relation of mereology isn't the right model for grasping what trying is (or more generally what 'complex' events are, or what physical objects are, for that mater). These actions, the stopping, closing, and lowering, are parts, in a non-mereological sense, of a plan, and plans incorporate ideas of simultaneous or successive actions; mereological wholes do not.

Lest it be thought that a possible rejoinder to this is to let trying time slices be identical to ordered pairs, in order to capture the idea of temporal structure, one needs to remember that ordered pairs are a special type of set. If they are anything else, trying or trying time slices are certainly not identical to abstract objects like sets.

To be sure, there are alternative accounts lurking in the same neighbourhood: constitution and non-extensional mereological composition are good candidates. The relation between composition and constitution is a matter of some controversy (Evnine 2011). There is a vast literature here and I will by necessity be brief.

Types of matter (wood, plastic, glass) constitute or make up objects. It may be that Goliath = the lump of clay, but Goliath is constituted, not by the lump of clay, but by clay (a mass noun). The table is constituted by wood. There is no identity associated with constitution so understood and hence constitution provides no model for grasping the alleged identity of a trying time slice and the mereological sum of two or more actions.

However, there clearly is another sense of composition other than the mereological one described above. Physical objects obviously have parts in some sense,

even though such parthood cannot be understood as extensional mereological part-hood. Our ordinary non-technical discourse licenses this non-extensional way of thinking about objects and their parts. But the thought that tryings have parts, and these parts are the actions by which one tries, is hardly a non-technical, ordinary idea. It would be to use 'part' as a specialised term of art, introduced for theoretical purposes. It is hard to see how the plausible and quite ordinary idea that physical objects have parts in a non-extensional sense could provide similar plausibility for the thought that tryings have parts in a non-extensional but ordinary sense, which parts being the actions by which they are done.

I won't pursue the counterarguments to this critique of the constitution or the parts strategy offered here (I acknowledge that there are possible replies). However, there is another problem that affects equally the case when there is only one and when there is more than one action associated with a single temporal slice of the trying. Suppose that the only action the agent undertakes between t_1-t_2 in order to try and stop the flood is to lower the flood barrier. In this simple case, the physical action theorist might say that the time slice t_1-t_2 = the lowering of the flood barrier. So what's wrong with that?

3 The Physical Action Theory and Problems of Action Failure

The agent may try to lower the barrier but fail. But if he fails, so the barrier is never lowered, there was no lowering of the barrier. It won't be true that he lowered the barrier. With what could the relevant time slice of trying to stop to flood be identified, if not with the lowering of the barrier (since it never occurred)? The only plausible candidate appears to be with his trying to lower the barrier, for that surely did occur even in the case of failure. But the problem here is obvious. Far from identifying trying with any physical action, we have only another trying, (say) *trying* to lower the barrier, with which to identify the time slice of trying to stop the flood. That isn't an illuminating identification theory of trying with a physical action at all. It might be true but it doesn't permit us to go beyond trying to a physical action in the way in which the physical action theorist had hoped.

Why not say that the time slice is identical to the lowering of the barrier when the agent successfully lowers the barrier, but not otherwise? This is a perfectly possible theory but it creates a partition among tokens of trying in an unexpected way that otherwise has no motivation. It commits us to two kinds of trying when from all appearances they seemed to be a homogenous class, and it still leaves us with the identification question for the tryings in the case of failure.

But the physical action theorist of trying has a reply. True, trying to stop the flood, or rather the appropriate temporal slice of the trying, might be only identical

to trying to lower the barrier. So far, no break out of the trying circle. But trying to lower the barrier, in its turn, might be identical to the physical action by which one tries to do that. Perhaps trying to lower the barrier = turning the handle of the gizmo. So the temporal slice of his trying to stop the flood from t_1 to t_2 = his trying to lower the barrier = his (actually) turning of the handle of the gizmo. But what then happens if the agent fails to turn the handle although he tried to do that? Well, he moved his hand, and perhaps that is a basic physical action, done by means of no other physical action of his. If so, we could anchor the tryings all the way down if needs be, until we hit a rock-bottom basic physical action.

4   The Physical Action Theory and Total Action Failure

This reply only works if the possibility of failure is not recursive, all the way down. But failure all the way down can occur in the following kind of well-known case. An agent could try as hard as he might, without any action whatever happening, not even a basic action, as in the case of Dr. Landry's or Professor Strümpell patients.[6] In William James' description of these cases (let's call them both 'pathological cases'), the patient lacks any 'kinaesthetic impressions' or any proprioception in some limb, so is unable to sense any movement of that limb or its location. The patient is blindfolded, so that he cannot see where his limb is. Finally, another person forcibly holds down the patient's hand, so that he, being anaesthetised, is unable to feel this pressure on his hand either.

After being asked to move the limb, he is surprised to discover when the blindfold is removed that he has failed to do so. He believed that he had moved it. In these cases, some muscular contractions occur even in the absence of bodily movement. In other cases of total paralysis, not even muscle contractions occur; there is only some brain activity on the part of the patient, which results in neither bodily movement or muscle contractions.

It is plausible to say in these cases that the patient had tried but utterly failed to do anything whatever. The thought is that the patient must have at least tried, because he would not have been surprised if, for example, he had only had the intention to move the limb but nothing more. No matter how many or which beliefs, desires, and intentions, the agent had, even proximal intentions, no one would be surprised about his action failure only in light of them. Even when an agent has a proximal intention, akrasia or inertia can set in and the agent can fail to act. Such a scenario would not lead to the agent's surprise about his action failure. If the agent is surprised, some sort of endeavouring must have gone on. Since no action whatever results from his trying, this type of trying is sometimes called

---

[6]William James describes the cases of the patients of Dr. Landry and Professor Strümpell. See James (1950, pp. 490-492).

'naked trying'.

In order to explain his surprise, do we really need to assume that he really did try? Would it not be enough to assume that he falsely believed that he had tried? If he believes that he tried, but unbeknownst to him his belief is false, then he would be surprised in the same way as he would have been had his belief about his trying been true. So the idea would be that he might falsely believe that he has moved his limb, not because he did try to move it, but because he also falsely believes that he has tried to move it.

What is the relation between trying and believing that one has tried? I would defend two theses: incorrigibility and self-presentation (although I only need the former here). Self-presentation claims: 'If P tries to φ, P believes that he has tried to φ.' A weaker self-presentation thesis says only that if P tries to φ, then he believes that he has tried to do something. But I think the stronger thesis is true, and that apparent counterexamples can be dealt with adequately.

What is at issue in the suggested reply is not self-presentation but incorrigibility, since the reply we are considering supposes that a person's beliefs about his trying are corrigible. Incorrigibility claims: If P believes that he has tried to φ, he has tried to φ. (There is also a weaker incorrigibility thesis similar to the weaker one about self-presentation.) I won't engage in a defence of the incorrigibility of trying here, although I believe it to be true. But I would add that if one denies it, it is then incumbent to produce a credible error theory of why an otherwise rational person might falsely believe that he has tried, and I can't see that the prospects for that are very good.

5   The Denial of Naked Trying

Some philosophers agree that there is such a phenomenon as naked trying. Hornsby defended the idea: 'even if all actions are tryings, not all tryings are actions' (Hornsby 1980, p. 42). More recently, Gideon Yaffe has also taken this view: '… trying need not be, itself, an action' (Yaffe 2010, p. 91). Yaffe doesn't mention the Landry case explicitly, but he has something like that in mind: in the case of a paralysed patient, all that may happen is that acetylcholine might be released by the postsynaptic nerve that normally activates the muscles but which may in the case of the paralysed patient be blocked (Yaffe 2010, p. 91). Other philosophers have denied the possibility of naked trying.[7] Wittgenstein might have denied the possibility (if we equate willing and trying) but it is not easy to decipher what exactly he means by his aphorisms on this topic (Wittgenstein 1984, p. 86).

Thor Grünbaum, George Wilson, and Timothy Cleveland have offered several

---

[7] See for example (Grünbaum 2008; Taylor 1973, pp. 82-85; Cleveland 1997, ch. 6; Wilson 1989, pp. 151-167).

arguments against the possibility of total action failure, and these are the only sustained philosophical discussions of the Landry phenomenon post-William James of which I am aware (in contrast to the Landry case merely being rather frequently mentioned or cited en passant). Grünbaum produces several arguments and I shall only reply to one of them.

What is crucial in this debate about naked trying is to separate trying from a sense of effort or innervation. William James denied, in opposition to Ernest Mach, that there was such a thing as a sense of innervation (James 1950, pp. 511-515), but he never denied that there was such a thing as trying, especially in the pathological cases. Grünbaum confuses the two (Grünbaum 2008, pp. 75-76). James identified trying with the will, not with a conscious feeling of the efferent commands. James' denial of the latter (say, in a pathological case) does not entail his denial of the former.

As for cases of failed action, James is clear that in paralysis, '[T]he volition occurs, but the hand remains as still as the table... He tries harder, i.e., he mentally frames the sensation of muscular effort... but the palsied arm lies passive as before' (James 1950, p. 561). In brief, James has all the ingredients in his theory necessary to explain why trying or endeavouring occurs in both successful action and total action failure, in spite of his dispute with Mach over a sense of innervation.

So too, it is very clear in the scientific literature on effort, that a distinction is being made between effort and trying, and that even when effort is absent or attenuated, say in the pathological cases they investigate, no scientific writer of whom I am aware has denied that the agents tried.[8] Even in discussions of effort,

---

[8]Here are several recent major discussions of effort or exertion in which it is simply assumed that trying always also occurs:

1. 'Exertion is accompanied by a sensation of strain and labour, a feeling that intensifies the harder a person tries', 'even if capacity is completely extinguished so that efferent activity is not possible (e.g., in paralysis), then no effort is felt even when intentionally trying to move', 'When people are trying to solve problems together, effort experiences can be the basis for unintended plagiarism' (Preston & Wegner 2009, quotes from 570, 571, 578 respectively).

2. 'Similarly in a second patient... futile attempts to abduct or extend the fingers were described as follows: 'My fingers felt normal but I could not move them. I knew what I was trying to do... but I could not feel any effort in it at all', 'To determine whether sensations of effort could arise when the limbs were not only paralysed but also insentient four patients with spinal transections were interviewed. These patients were aware of sensations of effort when first attempting to move their paralysed limbs... when specifically asked to try and move their limbs, patients said that the feeling of effort or heaviness remained', '... during the stage of complete paralysis of upper motor neuron type patients were aware of their failed attempts to move. But they did not feel the sense of effort...' 'Two patients, who suddenly became hemiplegic, without sensory symptoms, noted that attempts to move when movement first returned were accompanied by distinct sensations of effort or heaviness' (Gandevia 1982, quotes from 154, 155, 157).

3. 'Subjects with experimental paralysis of one limb experience strong sensations of effort when

psychologists simply assume that trying has also taken place. The important point is that the question of trying is not the same as the question of effort, the latter of which is closely associated with the idea of heaviness. Effort and a sense of innervation have a phenomenology; there is something it feels like to expend effort. But there is no obvious phenomenology to trying. For the many philosophers who think that one tries whenever one acts, this must be so, since if there were a phenomenology to trying, and since there is no sense of trying in cases of easy success, it would be implausible to hold such a view.

Of course, even when no action occurs in a case of naked trying, not even the flexing of muscles, some sort of brain activity must go on while, at the same time as, the agent tries to move his hand. We are not like the Straw Man in the Wizard of Oz, who can act in the absence of a brain. Perhaps the persistent physical action theorist could argue that the agent's token trying = his token brain activity.

This is the line of argument taken by George Wilson and Timothy Cleveland. Using the idea of de re intention, Wilson says that in a pathological case, '...to put it bluntly, electro-chemical activity...[occurs] in the brain...Why can we not say that it is this activity of the paralyzed man that was intended, by him, to make his arm rise?' (Wilson 1989, p. 161). 'Activity' can refer either to an action or to a non-actional process, as in 'there was a great deal of sun spot activity recently.' By activity, both Wilson and Cleveland mean action. Wilson and Cleveland say repeatedly that the brain activity (or movement) is something the agent *performs*.

In the case of the paralysed man, the idea would be: 'The agent performed a movement and intended of that movement that it cause his arm to rise'. In such a case, the brain activity is the action the agent performs, but his intentions about it are claimed to be de re, so the agent may have no real knowledge about the nature of the brain movement or even that it is brain movement. This allows Wilson (and Cleveland) to dodge the objection that the agent 'must have some intrinsic awareness of' the brain activity. The agent 'can intend of activity that it do so-and-so, even if [he has] no awareness of what that activity consists in' (Wilson 1989, p. 162). Wilson says that the agent might only be aware that he is doing something or other intended to lead to a certain result, but he may have no other knowledge about what that something is. So, for Wilson and Cleveland, the agent's trying = his brain activity. The brain activity is the agent's action. If so, there would be no naked trying after all.

I think that there are several things wrong with the Wilson-Cleveland proposal. First-and here I guess we are only swapping intuitions-they are committed to the

---

they attempt to move that limb...The same hypothesis would account for permanence of sensations of effort in all cases of distal paralysis, where corticofugal pathways are not altered' (Jeannerod 1995, p. 1429).

.

view that in the paralysed arm case, the agent has an intention *about* something-but he is in the dark about exactly what the intention is about-that constitutes his trying to raise his arm. I see no grounds for attributing any such intention to the agent. Perhaps we simply have to accept that in such a case, the agent fails to have any de re intentions in Wilson's sense and has only a general intention. The agent might only have a general intention in such a case; he intends that he perform some action that is his raising of his arm (but of course he fails to perform any such action). What grounds are there for attributing a de re intention to the agent in these pathological cases?

Second, Wilson raises what I take to be the major objection to his view: '. . . there may be some residual worry about whether electro-chemical activity in the brain... can be activity that the agent performs' (Wilson 1989, p. 164). He says that he cannot see that this is a substantive issue, because he sees the issue as one about the ordinary use of 'performs', which he suggests can simply be extended to cover brain activity if it doesn't cover it already. But the issue is much deeper: it is a question of what can count as bona fide action. Wilson begs the question when he argues for the proper use of 'performs' in connection with muscular contractions, which he says 'are part of the exercise of my normal control over my body' (Wilson 1989, p. 164), thereby conflating the idea of things an agent can control with the idea of necessary conditions for his control. The expression 'are part of' could be taken in either sense.

It is easy to state the contrary case to Wilson and Cleveland: brain activity is not an action; it is nothing the agent does (or performs, as they say). Not everything internal to an agent that is a necessary condition for his action is itself an action. But here is a possible reply, in defence of Wilson and Cleveland, to that too easy dismissal. Joel Feinberg spoke of the accordion effect of action: 'we can if we wish puff out an action to include an effect. . . Instead of saying that Smith did A. . . and thereby caused X in Y, we might say something of the form, "Smith X-ed Y". . .' (Feinberg 1970, pp. 119-151).[9] By puffing out the action, many of the things that Smith does, like X-ing, will be unintended, or anyway unintended under the description 'X-ing'. The unintended event effect, X, can be made intrinsic to an unintentional action, his X-ing. For example, in a well-known example, the agent turned on the light and the prowler was unintentionally alerted. It is true that the agent alerted the prowler although his action of alerting him was unintentional (or unintentional under that description).

But instead of only puffing out action forward, why not also deflate action inwards? Where X is a causally necessary condition for Smith's A-ing that is still internal to Smith, can we not also say that Smith X-ed, perhaps unintentionally? Smith's neurones fired. So since causal chains extend backwards as well as forwards,

---

[9]See also (Bratman 2006).

we might say that Smith fired his neurones, and he fired them by (say) moving his arm. The 'by' here is of course teleological and not temporal. His firing of his neurones occurs before his moving of his arm, although he only does the former by doing the latter. There can be unintended causes as well as unintended effects of action, both of which might be made intrinsic to unintended actions by the agent.

Deflating the accordion effect isn't going to help the identification theorist who wishes to identify trying with some brain activity. On the backwards-and-forwards accordion effect model, every unintended action (or action under its unintended description, but I shan't keep repeating this qualification) has to be built up as an effect or a cause of an intended action. But in the case of the totally paralysed arm, there is no other action X such that it can be true that the agent performs a range of brain activity and the brain activity is an unintended cause of his X-ing intentionally. If performing some brain movement is an unintended action, in the case of naked trying what would be the intended action in virtue of whose causes the agent can unintentionally perform some brain activity?

So in normal cases, it may be true that the agent fires some neurones, whether he knows it or not, just as he may or may not know that he has alerted the prowler. In the paralysed arm case, however, there are no grounds for taking the brain activity as intrinsic to any action the agent does. Although there is certainly brain activity going on in this case as in all others, he doesn't *perform* a brain movement at all, let alone one about which he has de re intentions. The identification that Wilson and Cleveland propose assumes that we have grounds for thinking that there is such an action as firing one's neurones (or performing brain activity), and *then* we can identify it with the trying. But there simply are no independent grounds for believing that the agent fires his neurones or 'performs' any brain activity in this sort of case. The brain activity could be an unintended action only if it is the cause of something else the agent does intentionally.

6   Gideon Yaffe's Attempts

Although the main focus of his book is on attempts in a legal context, Gideon Yaffe's *Attempts* is probably the fullest account of the nature of trying in the recent literature. I have already cited Yaffe's acceptance of the possibility of naked trying, whether or not these would have any place in a legal context. On this, we agree. In light of this acceptance of the possibility of naked trying, what exactly does Yaffe think trying is? I will present a rational reconstruction of his view, which I hope brings out more clearly the various strands of that view. These are not always the terms in which he presents his view. His view of trying is close to, or at least consistent with, the one I have defended elsewhere (Ruben 2015).

On his view, an agent tries only if he is guided or moved by his relevant inten-

tion (Yaffe also speaks of this being a 'striving') (Yaffe 2010, pp. 73-74). Although to try is in large part to be in a certain mental state (Yaffe 2010, p.90), the agent must be moved by the mental state or anyway by the commitments they constitute. In the light of the first sentence of this paragraph, the mental state referred to in the second sentence must refer to an intention. There is no suggestion that trying is a sui generis mental state in addition to an intention.

In successful action, the intention motivates and guides the agent to perform some token action. As Yaffe uses 'motivates', it could either mean 'gives the agent a reason to do an act of type ϕ' or, 'causes the agent to do an act of type ϕ'. 'guides' on the other hand seems unambiguously to be a causal notion. There seem really to be two parts to this requirement, however we label them: a causal part and a reason-giving part. In the case of a successful action, on a certain theory of reasons at any rate, the same token action could both be caused by the intention and be the action that fulfils the intention, the action at which the intention aims.

In the case of a naked trying, in which there are no relevant action tokens, if the agent tries to ϕ, then the intention will be a reason for the agent to do some action of the *type*, a ϕ–ing, some token of which the agent believes would fulfil his intention were it to occur. But there must be some causal requirement as well. '. . . without being moved by one's commitment, there is no striving towards success, and so no trying' (Yaffe 2010, p.73).

So how would this work in the case of a naked trying, since there are no actions for the intention to cause? '. . . the intention that constitutes the commitment non-deviantly causes an event as part of motivating an action that the agent believes will. . . serve as a means to fulfilling his intention' (Yaffe 2010, p. 92). The intention causes something 'on the way towards performance' of what he is trying to do (Yaffe 2010, p. 74). In private correspondence to me, Yaffe has said (substituting my own symbols and numbering): 'My view is that that's an instance of guidance just in case there is an act [of type] ϕ such that (3) S believes that doing some action of that type would be a means to the satisfaction of the intention, and (4) the intention causes [some token event] E while motivating ϕ-ing. The crucial point here is that ϕ and E need not be identical. They can be, but they need not be.' I have altered his response slightly to clarify what I think is best construed as a type and what as a token: I think that 'ϕ' is best regarded as an act type; 'E', a token event. Of course, on this construal, E and ϕ could not be identical, since one is a type and the other a token. In the case of naked trying, Yaffe holds that E is the release of the acetylcholine by the postsynaptic nerve. For the agent to have tried, his intention must have had some physical effect or other, and moreover one on the relevant causal pathway which would have otherwise lead to the action one intended to do.

To summarise: Yaffe's analysis of trying seems to be this. One tries to ϕ iff

one's intention to φ causes something on the causal pathway that leads or would have lead to a token action of type φ, such that an action of that type does fulfil or would have fulfilled that intention.

7   Conclusion

The result of this discussion then is to support the view that there are naked tryings, tryings such that there are no physical actions with which they can be identified. Let me stress the obviously close connection between the physical actions by which one tries and what it is that one tries to do. That there is a close connection is surely obvious. The question is: what is the nature of that close connection, if not identification? My view is the most natural one: they are the actions *by which* the agent tries. Just as physical actions can have by-acts, there are by-acts when one tries. Just as agents can φ by ψ–ing, so too an agent can try by ψ-ing. But this offers no comfort to a physical action theorist of trying. The conclusion is that there is no physical action the agent does such that his trying in the pathological case, or indeed in any case, is identical to that action.

References

Bratman, M. E. (2006), 'What is the accordion effect?', *Journal of Ethics* **10**(1-2), 5–19.

Cleveland, T. (1997), *Trying without willing: An essay in the philosophy of mind*, Ashgate Brookfield, VT.

Evnine, S. (2011), 'Constitution and composition: Three approaches to their relation', *ProtoSociology* **27**.

Feinberg, J. (1970), *Doing & Deserving; Essays in the Theory of Responsibility*, Princeton, N.J.,Princeton University Press.

Fine, K. (1999), 'Things and their parts', *Midwest Studies in Philosophy* **23**, 61–74.

Gandevia, S. (1982), 'The perception of motor commands or effort during muscular paralysis.', *Brain: a journal of neurology* **105**(Pt 1), 151–159.

Ginet, C. (1990), *On Action*, Cambridge University Press.

Gozzano, S. & Hill, C. (2012), *New Perspectives on Type Identity: The Mental and the Physical*, Cambridge University Press.

Grünbaum, T. (2008), 'Trying and the arguments from total failure', *Philosophia* **36**(1), 67–86.

Hobbes, T. (2012), *Leviathan*, Clarendon Press.

Hornsby, J. (1980), *Actions*, Routledge and Kegan Paul.

James, W. (1950), *The principles of psychology*, Dover Publications, New York.

Jeannerod, M. (1995), 'Mental imagery in the motor context', *Neuropsychologia* **33**(11), 1419–1432.

McCann, H. J. (1998), *The Works of Agency: On Human Action, Will, and Freedom*, Cornell University Press.

O' Shaughnessy, B. (1980), *The will*, Vol. 2, Cambridge.

Pietroski, P. (2002), *Causing Actions*, Oxford University Press, Oxford.

Preston, J. & Wegner, D. M. (2009), Elbow grease: When action feels like work, *in* E. Morsella, J. A. Bargh & P. M. Gollwitzer, eds, 'Oxford Handbook of Human Action', Oxford University Press, pp. 569–586.

Ruben, D.-H. (2013), 'Trying in some way', *Australasian Journal of Philosophy* **91**(4), 719–733.

Ruben, D.-H. (2015), 'A conditional theory of trying', *Philosophical Studies* pp. 1–17. online at:
  http://link.springer.com/article/10.1007/s11098-015-0490-5.

Simons, P. (2000), *Parts: A Study in Ontology*, Oxford University Press, Oxford.

Steward, H. (2013), 'Processes, continuants, and individuals', *Mind* **112**(487), 781–812.

Taylor, R. (1973), *Action and Purpose*, New York,Humanities Press.

Wilson, G. M. (1989), *The Intentionality of Human Action*, Stanford University Press.

Wittgenstein, L. (1984), *Notebooks, 1914-1916*, Vol. 47, University of Chicago Press.

Yaffe, G. (2010), *Attempts: In the Philosophy of Action and the Criminal Law*, Oxford University Press, Oxford.