ACTION WITHOUT THE FIRST PERSON PERSPECTIVE

Herman Cappelen and Josh Dever

Draft October 2017

In our book *The Inessential Indexical* we argue that the various theses of essential indexicality all fail. Indexicals are not essential, we conclude. One essentiality thesis we target in the third chapter is the claim that indexical attitudes are essential for action. Our strategy is to give examples of what we call *impersonal action rationalizations*, which explain actions without citing indexical attitudes. To defeat the claim that indexical attitudes are essential for action, it suffices that there could be even one successful impersonal action rationalization.In what follows we bolster our case against an essential connection between action and *the de se* (or indexicality), first by developing a range of new action models and secondly by responding to challenges from Dilip Ninan, Stephan Torre, and José Luis Bermúdez.

Action without the first person point of view

The Inessentiality Thesis involves a modal claim concerning a necessary condition on action. In particular we argue that IIC is false:

> **Impersonal Incompleteness Claim (IIC)**. Impersonal action rationalizations (IAR) are necessarily incomplete because of a missing indexical component

We imagine a proponent of IIC claiming that the Impersonal Action Rationalization below *must* be incomplete, while the Personal Action Rationalization is or could be complete:

> Personal Action Rationalization (explanation) 1.
> • Belief : François is about to be shot.
> • Belief : I am François.
> • Belief (Inferred) : I am about to be shot.
> • Desire : I not be shot.
> • Belief : If I duck under the table, I will not be shot.

• Action : I duck under the table.

Impersonal Action Rationalization (explanation) 1.
• Belief : François is about to be shot.
• Desire : François not be shot.
• Belief : If François ducks under the table, he will not be shot.
• Action : François ducks under the table.

We say there are perfectly good Impersonal Action Rationalizations. So far this is about explanations or rationalizations. Our point is also metaphysical: an agent's action need not involve any first-personal representations. For example,she need not represent herself, qua herself, as the acting agent, nor does she need to represent the part of her body that acts on the world as her body (in a first-personal way[1]).

Action Inventory Model

In *The Inessential Indexical,* we presented one version of how Inessentiality could be implemented: the Action Inventory Model.

> **The Action Inventory Model.** Every agent has a very wide range of third-person beliefs and desires that give rise to third-person intentions, which in turn rationalize or motivate actions (via their recognition). Not all of these intentions are going to produce action, at least in normal cases (perhaps a god's intentions would). These failures occur because a given agent has an "action inventory": a range of actions that he can perform. An agent constantly seeks to match his intentions with his action inventory, and when he notices a match, action occurs. When there's no match, the intention idles, and doesn't motivate or rationalize action. So the Selection Problem[2] is solved by appealing to the physical or psychological constraints of the agent: only certain actions result because only certain actions were available in the first place.

The inclusion of this one illustration of how Inessentiality could be implemented was presentationally unfortunate: it gave the impression that this was how we thought action happened or that the Action Inventory Model was the *only* way to explain how actions could happen without appealing to *de se* attitudes. That's not what we think. We think there are literally infinitely many ways one can explain how one could act without de se representations. We outline a few below, but before giving more cases, here's a general recipe for how to get from a theory of action that includes an essential appeal to the de se attitudes to an alternative

[1] This 'in the first person way' is a way of talking employed by those who defend Essentiality Thesis: it's supposed to distinguish Nora thinking of her hand by thinking of it as Nora's hand vs thinking of it as *her* hand (the italics on 'her' is supposed to make forceful the perspectival reading). It's hard for us to sympathize with the difficulties of articulating this idea since we don't think there's anything here to articulate: so we take the hardness of stating the position as evidence of its problem.
[2] Note on what selection problem is....

non-de-se model. Just take out the de se component of your view, and as a replacement, assume that there's a god that bridges the gap, i.e. does what the work that the de se attitudes were supposed to do for you (so whatever theoretical and explanatory work the de se attitudes were doing, god does.). God might do this in different ways: she might use neurology, magic spells, send out angels, or use little gnomes to do the work.

More non-de-se Models of Action

François is about to be shot, but ducks under the table just in time and avoids the bullet. How did this action come about? The *de se* story is that the full story about how it came about *must* include certain kinds of beliefs and desires: *first-personal* beliefs and desires. François needs to believe that *he himself* is about to be shot, believe that ducking under that table near *him himself* will save him, desire that *he himself* not get shot, and so on.

We've claimed, on the contrary, that none of these special sorts of beliefs and desires are necessary. We're happy with a stripped-down account of how François' action came about. As long as François has some beliefs and desires that rationally integrate to lead to the action *François ducks under the table*, François has what he needs to act. We'll thus in many cases be happy with strict subsets of the explanations given by fans of the *de se*. Strip away all of the first-personal beliefs and desires, and there's still a perfectly good explanation of the action.

But it would be nice if things didn't just devolve into clashing judgments about which explanations are good. Thus in *The Inessential Indexical*, we also gave a picture of how action could *work* such that our trimmed-down explanations would be good ones. This was our Action Inventory Model. The Action Inventory Model has attracted considerable critical attention in discussion of the book, so we want to clarify its role. The AIM was intended as a *proof of concept*. The point was not to say that this is how human action works, but rather to say that this is one way action *could work*, and that if it did work that way, what we call impersonal action explanations would be perfectly good full explanations. But once the concept is proved, it can be implemented in many different ways. Consider some more ways that action could proceed in the absence of first-personal beliefs or desires:

1. In an occasionalist world, there's no direct causation between objects. When billiard ball B1 hits billiard ball B2, the impact doesn't cause B2 to start moving. Rather, all causation comes from the occasionalist god. When God sees B1 hit B2, God causes B1 to stop moving and B2 to start moving. The occasionalist god sustains human action as well. When God sees that Jones wants a beer, and that Jones believes there are beers in the refrigerator, God causes Jones to go to the refrigerator.

   The causal structure of the occasionalist world is different from the causal structure of our world, but it *looks* the same as in our world. Jones wants a beer, thinks there are beers in the fridge, and goes to the fridge. We think Jones *acts* in the occasionalist world

just as much as in our world. (After all, our world could turn out to be an occasionalist world.) But in the occasionalist world, first-personal beliefs aren't necessary. That's because the occasionalist god is a helpful one. Suppose François believes *François is in danger* and desires *François not be in danger*. That's enough for the occasionalist god: she then sends François under the table. There could have been a less helpful occasionalist god. The less helpful god won't send François under the table unless François *also* believes *I am François*. But it doesn't have to be that way. There's nothing wrong with the helpful occasionalist god, and in the presence of the helpful occasionalist god, François acts without first-personal beliefs.

Similarly, there could have been an *epistemically demanding* occasionalist god. The epistemically demanding god won't send François under the table unless François not only believes, but actually *knows*, that François is in danger. In the world of the epistemically demanding god, explanations of action require appeals to knowledge. That's of course one way that acting could be, but the fact that it could be that way doesn't alter the other real possibility that actions could come from beliefs that aren't knowledge.

2. Normally when we act, we need to keep track of how things are around us. If you want to get a beer from the refrigerator, you need to know where the refrigerator is in relation to you, so that you can get your hand to the right place. But that's a result of contingent limitations on our abilities. Consider the absent-minded CEO Jeeve Stobs. Stobs is too distracted plotting the future of cloud computing to keep track of such mundania as where he is and where his beers are. To avoid death by dehydration, Stobs has hired a collection of assistants. So when Stobs wants a beer, he just says, "Beer!", and one of his assistants locates a beer, gets it, and brings it to him. They are very helpful assistants. It's not just beers -- Stobs can say "Call Zark Muckerberg!", and an assistant will find a phone, place the call, and give the phone to Stobs. Stobs can say "Go to Palo Alto!", and an assistant will carry him to a car and drive him to Palo Alto.

So Stobs has no need to keep track of the environment around him. He doesn't need to know where things are to act on them, because his assistants will find things when he requests them. (Does he at least need to keep track of where his assistants are? No. They make sure to stay nearby at all times. Of course, if flukishly they're out of hearing, his intentions won't manifest in action. But all of us are, under unlikely unfriendly circumstances, liable to have our intentions not manifest in action.)

We can imagine Stobs getting saddled with an unhelpful assistant, who won't bring Stobs the beer unless he's sure that Stobs himself has a belief, or a true belief, or knowledge of, or direct line of sight on, where the beer is relative to him. When Stobs has such an unhelpful assistant, he will be incapable of action without a first-personal belief, or first-personal knowledge, or direct perceptual evidence. That is indeed how it

could be. But again, the fact that it could be that way doesn't interfere with the fact that it could also be other ways, ways that don't create any need for first-personal states.

Should we worry that Stobs still needs first-personal beliefs to perform the action of *ordering his assistants*? (He needs to know where his mouth is relative to him, for example, to use it in speaking. If there's a button to summon an assistant, he needs to know where that is relative to him.) Suppose this gets to be a problem -- Stobs is so absent-minded that he can't keep track of how to notify his assistants that he's got an intention. So the assistants set up brain-monitoring equipment so that they can directly supervise Stobs's mental states. Now when they see Stobs wanting to make a call, they bring him the phone. There's no need for him to get his mouth into motion, or to push a button.

3. You might worry that our first case, with François and the occasionalist god, smuggles in first-personal beliefs. Maybe François doesn't need to have them, but what about the occasionalist god? This worry doesn't actually matter for that case. Even if the occasionalist god does need to have first-personal beliefs to act, if François doesn't, then we have a counterexample to (IIC), and one counterexample is all we need.[3]

But adding first-personal beliefs to the god is unnecessary. Gods, being gods, can act in ways that we can't. When we want a beer from the refrigerator, we need a plan, setting in motion a sequence of things sufficiently proximal to one another bridging the gap between us and the refrigerator. That's because of contingent limitations on our capacities. But gods aren't so limited. When a god wants a beer from the refrigerator, there's no need for moving of hands, or opening of refrigerator doors, or uncapping of beer bottles. How do gods act? Well, who knows? But maybe like this: gods just have direct control over the state of reality, so for a god, there's no gap between *intending that p* and *it's being the case that p*. So when a god wants a beer to be on the table, he simply intends *there is a beer on the table*, and the thing is done. Why would there be any need for first-personal beliefs for such a god?

Such gods don't have any need to keep track of their own capacities, either, because there is no limit on those capacities. But we could imagine as well somewhat more humble demigods. The demigods also act simply by intending, but their powers don't allow them to realize just any intention, but only certain kinds of intentions. Semiramis can get a beer on the table just by intending that there be a beer on the table, but can't

[3] Perhaps François doesn't really act here, but rather François and the occasionalist god act together? But we can push the theology in a more and more impersonal direction, making the occasionalist god less like another acting agent and more like a sustaining first-causal force. Perhaps (IIC) can be weakened to the claim that for every action, someone somewhere must have a first-personal belief that features in the explanation of that action? It's not clear that the occasionalist god's putative first-personal states have to be mentioned in the explanation of François' action (hard questions about what counts as a *full* explanation come up here), but even if they do, it's also unclear whether this weakened (IIC) is particularly interesting. We're in danger of having a connection to action that's weak enough that just about any belief will have that connection.

get a biscuit on the sideboard just by intending that there be a biscuit on the sideboard. So do demigods need to keep track of what they can and can't directly do, and does that require first-personal beliefs? (Again, as long as we've got the gods, it wouldn't matter for our purposes if the demigods required first-personal beliefs. But they don't.)

No. Maybe Semiramis is just an inveterate optimist. She doesn't keep track of what she can and can't do -- she just intends willy-nilly, and sometimes action results from those intentions and other times it doesn't. Or maybe Semiramis has undergone Pavlovian training to restrict what intentions she forms. When she forms biscuit-y intentions, she experiences a painful electric shock, but when she forms beer-y intentions, she receives a reward of ambrosia. Without coming to have *beliefs or mental representations about the boundaries*, her dispositions to intend can just be shaped in such a way that she no longer intends where her divine capacities give out. Or maybe one of her divinely realizable intentions was that *Semiramis intend only what she can directly do*, and she so intended early in life. Later, we'll suggest that we are ourselves can be thought of as demigods of sorts.

4. You might have similar worries about Stobs and his assistants. Doesn't this example just push the problem back a step? After all, the *assistants* will need first-personal beliefs. As with the gods, we claim that this is irrelevant even if true -- we still get Stobs as a counterexample to (IIC). But suppose corporate structure gets complicated. The demands on the assistants become so large that the assistants get their own assistants, and defer implementation of their actions to their second-order assistants. Problems, of course, just pushed back one more step. But then we add another level of assistants, and another, and another. Suppose that there is an infinite sequence of assistants. Then Stobs intends that Kim Took be fired. Assistant A1 carries out that intention, by forming a plan that (e.g.) requires the writing of an email. Assistant A2 is then responsible for the writing of the email, and forms a plan that requires turning on the computer. Assistant A3 now steps in, and so on. (Perhaps this is a gunky world, in which causal paths always decompose into smaller causal subpaths, so that the sequence of assistants maps onto the downward sequence of causal subpaths.) Every assistant has his own assistant, so we never reach a level of assistants acting unassisted. Would action result? Infinite regresses are always tricky to assess, be we don't see why not. The fan of *de se* beliefs should, we think, say that action will result if we liberally sprinkle first-personal states along the regress. So if there's a problem here, it's not in the regress itself, but in the lack of first-personal beliefs. And then we think that the structure of the example shows that the first-personal beliefs are inessential.

The assistants form a downward chain, but we could also have an upward chain. François intends to duck under the table. The occasionalist god, monitoring the situation, spots this intention and puts it into action by bringing it about that François ducks under the table. Thus François acts. But why and how does the occasionalist god act? Well, the occasionalist god intends to bring it about that François ducks under the table. And

the occasionalist metagod, monitoring the situation, spots this intention of the first god and puts it into action by bringing it about that the god bring it about that François ducks under the table. And why/how does the metagod act? Well, there's this metametagod watching things. And so on. We get an infinite sequence of monitoring gods, each one ready to plug any "gap" between intention and action. Or maybe there are only two gods (Lewis's two gods, perhaps.) Each one monitors the other, and brings it about that intentions of the other result in action. Of course, their bringing it about is itself an action realizing their intention to bring it about, and is thus brought about by the other god.

5. We are building Robbie the Robot. Robbie has a collection of basic actions he can perform -- rotating and moving various appendages. These basic actions are triggered just by activating electrical current along various wires, and we have hard-coded programming in Robbie's software and hardware for activating that current. We then program up more complex actions -- sequences of basic actions that we label "lift" or "run" or "walk", and that we set to be triggered by certain program methods.

Next we build a belief system for Robbie. Robbie has some perceptual systems: visual and auditory receptors. We use those to give Robbie an objective map of the world. Robbie also receives GPS signals, and integrates those signals with the perceptual inputs to create a partial map with a coordinate system centered on, say, the earth's center, and with objects placed at different coordinates. Among the things Robbie tracks coordinates for are Robbie himself, as well as various appendages of Robbie. (But Robbie himself isn't represented any differently in the map from other objects. It's just one more thing, tagged with a label "R".)

Now we can program even more complex actions for Robbie. We program "put book on table". This program causes Robbie to consult his map, find the book on the map with the minimal distance from R, activate sequences of motion closing the spatial gap between R and the book, activate more sequences bringing various appendages into contact with the book, and then use those appendages to move the book to the table.

We haven't given Robbie any *de se* states. He represents himself, but only as one object among many in the world. There's no sense in which he knows *which object he is*. But that's no barrier to Robbie's putting the book on the table.

Robbie doesn't act yet, plausibly. He just does things on our command. To make Robbie more useful, we add a goal system to him. Now Robbie comes with a collection of target states of the world (maybe just represented as target objective maps). Robbie monitors the environment, creates and sustains an objective map, and when that objective map gets sufficiently similar to a target map (by some similarity score we program in), he initiates action that shifts the world into the form of the target map.

We still haven't given Robbie any *de se* states, but it's hard to see why his behavioue doesn't count as full-fledged action on his part. If the fan of a *de se* requirement thinks that Robbie isn't yet acting, then we think the burden is on them to say (in a non-question-begging way) what's missing. The important thing is that there's no danger that we'll find ourselves in the end with a robot that just doesn't do anything because we failed to include some special kind of representation into its representational systems. The ability to locate objects (either absolutely, or relative to one another) and to link invocation of basic actions to object location, is enough to get action underway. There's no need for Robbie to have an additional special belief that *he himself*, in the distinctively first-personal way, is somewhere or other.

Is there a worry that we have, without realizing it, given Robbie first-personal beliefs? It's hard to see how, unless we just take "first-personal beliefs" to be *whatever sorts of beliefs suffice to bring about action*. (Of course we agree with (IIC) if "first-personal beliefs" is read in this way, because this makes (IIC) into an uninteresting definitional tautology.) Robbie doesn't really even need to track Robbie's location. It would suffice, for example, to track the locations of Robbie's various appendages. (Depending on Robbie's design, those appendages could be in locations quite different from that of Robbie himself.) And for some cases, there won't even be a need to track the locations of the appendages. If, for example, some of the appendages are fixed in location, then Robbie can just track the locations of distal objects, and trigger action when those objects end up in certain locations. (Those locations will in fact coincide with the locations of Robbie's appendages, but there's no need for Robbie to know or otherwise represent that. We, as the programmers, just help ourselves to a certain stability in the world to simplify the programming task.)

6. Robots, gods, and corporate executives can all act. So can *entire corporations*. Apple performs actions all the time. Apple adds an edge-to-edge display to the new iphone. And it does this for a reason, in the characteristic way that actions are done. Apple *wants to win a larger share of the smartphone market*, and Apple *believes that an edge-to-edge display will help win customers away from Samsung phones*. That desire and that belief combine to bring about Apple's action.

But there's no need for Apple to have any first-personal beliefs. Indeed, Apple may be incapable of having first-personal beliefs. What a corporation believes depends in part on institutional facts about the corporation. Apple knows Samsung's earnings for the year because there's a corporate-implemented representation of that fact -- a file in a filing cabinet or on a computer with the earnings, or an corporate employee tasked with tracking Samsung earnings who knows the figure. If that same employee also knows IBM's earnings, but that knowledge isn't part of his corporate responsibility, that won't count as *Apple* knowing IBM's earnings. As a result, corporate beliefs can be haphazard and gappy in a way that normal human beliefs typically are not. (Evans' Generality Principle, for example, has little plausibility for *corporate* beliefs.) And since they can be

haphazard in this way, it's possible to set up a corporation in a way that just makes it impossible for it to have beliefs about *itself as itself*.

And capable or not, Apple doesn't *need* those beliefs. Apple, for example, doesn't need to track where it is, and where it is in relation to the targets of its action. It's not clear that Apple is anywhere in particular, and even if it is somewhere (maybe in Cupertino, at corporate headquarters), where it is is irrelevant to how it acts. (Position relative to Cupertino doesn't need to be tracked to get edge-to-edge glass on the next iphone.) Similarly, Apple doesn't need to do any representation of its own capacities. Rather, its corporate architecture just builds its action structure in a way that it just doesn't try to do the kinds of things that it's incapable of doing, and does try to do the kinds of things it is capable of doing. Apple can make an iphone with edge-to-edge glass, and it's got structural features that set it up to use that capacity. Apple can't win the 100 meter dash in the Olympics, but Apple is also set up in such a way that none of its beliefs and desires could even try to trigger such an action.

Stepping back from the details: we claim that there are *possible ways acting beings could be organized* that let them act without having any first-personal mental states involved in the production of the action. There's *some sense* in which we think this is trivial. If actions are just events causally brought about by beliefs and desires, and if the causal effects of beliefs and desires *could be* pretty much anything, then there's no way any particular kind of belief is needed for action. (Forget the fancy stories told above -- suppose we just posit a being in which the belief that two and two make four and the desire to eat a sandwich reliably cause a scratching of the nose. Then there's an action without any first-personal beliefs.) But we don't want to lean too heavily on this way of putting things, because many people will deny the antecedent, and claim that such mere consequences of beliefs and desires need not be actions -- that actions need to be caused *in the right way* by beliefs and desires.

Absent the trivial argument, we need another way to convince people of the possibility claim that contradicts (IIC). Part of that way involves defusing arguments in favor of the (IIC) -- we'll come to that later in the paper. But another part of the way is vividly to display certain possibilities, to remind ourselves that the way action *typically occurs in us* isn't the only way that action could occur. The Action Inventory Model was one attempt at such a vivid display, but it's not the only one. We think there are many ways that acting beings could be organized that let them be like enough to us in the relevant ways for them to count as *really acting*, but that don't require them to use first-personal beliefs and desires. No single one of these cases is argumentatively essential. We're not saying of any of these pictures that this is how *action itself essentially works*, or that this is how we really act. We just want to say that some actions could be like this, and if *any* of those claims is right, then (IIC) fails.

That said, we are inclined to think that in fact many of our normal human actions are much like the actions described above. We are, for example, our own little demigods. Gods of very limited scope (local deities of fingers and toes), but when it comes to certain basic muscular motions,

there is for us no gap between intention and action. To twitch a finger, you don't have to make a plan for the twitching, or track where the finger is. You just *directly twitch*. And we are CEOs of our own bodies, equipped with a host of executive assistants in the form of subpersonal computational systems that work out the details of how (some of) our schemes are to be put into biochemical action. Consider two ideas that underlie both the Action Inventory Model and a number of the cases discussed above:

1. Not every action comes with a plan for its implementation, or is done by performing other actions. Some actions do. Jones writes the email *by* turning on the computer, selecting words, and pressing keys on the keyboard. Pressing a key on the keyboard, in turn, is perhaps done *by* moving the finger in a certain way. But moving the finger in that way (to pick a stopping point) is not done by doing anything else. It's just a basic action; one that we can simply perform at will. (That's not to say that we are infallible in the performance of basic actions. It's just that when we fail at performing a basic action, it's a primitive failure, not a failure constituted by a failure to do something else.) Basic actions, being basic, don't require guidance by further mental states. Once we know the intentional content of the basic action, we know all of the cognitive tools needed to perform that basic action. So if *moving a finger* is a basic action, there's no need for thoughts about *specific muscles of the finger* in performing that action.

   What's characteristic about many of our cases above, then, is that they involve beings whose range of basic actions involves actions with no first-personal component. In the limiting case, gods are beings for whom *every* action is basic. It's common to think of first-personal cases when thinking of ordinary human basic actions (*move my finger, close my eyes*), but there's no clear reason why actions that aren't first-personal (and everyone agrees that there are such actions, even if they think that performing those actions requires other first-personal states) can't also be basic.

   2. Relatedly, it's a mistake to think that because successful performance of an action involves being rightly related to some object or concept, that we have to *mentally represent* that object or concept. This mistake is what we called the *Overrepresentation Fallacy* in *Inessential Indexical*. For example, for a baseball player to catch a fly ball, he must run at a speed and in a direction such that the first derivative of the tangent of the apparent angle of elevation of the ball is constant. But of course it doesn't follow from that that baseball players need to have *beliefs* about first derivatives of tangents. Not even tacit beliefs of this sort are needed, and not even mental representations in subpersonal systems. It suffices if something in the full physical implementation of the agent, connecting mental states to bodily movement, encodes an appropriate sensitivity to the derivative. (This could, for example, be achieved by having the visual system set up so that it creates visual seemings in a way properly correlated with the relevant derivative.)

   For creatures whose causal powers have any kind of locality constraint, where they are in relation to the targets of their action will matter to how the action is to be performed.

> But it would again be a fallacy of overrepresentation to think that such creatures thus had to *mentally represent* where they are in relation to their targets. Jeeve Stobs doesn't need to keep track of where things are, because he has assistants that do the tracking for him. That fact is key to Stobs' ability to act without first-personal beliefs.

When we set aside the modal claim of (IIC) and consider the question of how we in the actual world really act, our (admittedly armchair) suspicion is twofold. First, our basic actions are pretty expansive (and able to expand with experience and expertise), and that as a result we often act directly beyond the scope of our body. (When the car skids, we perform a basic action of righting the car's course. Even *turning the steering wheel*, much less *tensing our arms*, is not a subaction of that action.) Second, there's a lot less representation going on than one might initially suspect. We routinely rely on contingent regularities of our environment and of our capacities to save attentional energy by not representing things that aren't likely to change or matter. Given those two pieces, we suspect that a lot of our ordinary action is indeed impersonal in our sense.

Replies to Torre, Ninan, and Bermúdez

We turn now to 3 efforts to defend IIC. The defenses we consider are from Dilip Ninan, Stephan Torre, and José Luis Bermúdez.

Reply to Stephan Torre

Stefan Torre in "In Defense of *De Se* Content" (Torre 2017) argues for a principle he calls (CDDS):

> (CDDS) Necessarily, for any subjects, S and T, if S and T agree with respect to the content of their beliefs, then they have the same *de se* beliefs.

We won't be concerned here specifically with objecting to (CDDS). Torre wants to work with a notion of "*de se* belief" that is broad enough to allow *de se* skeptics like us to agree that there are such beliefs (roughly, beliefs that subjects are disposed to report using first-person pronouns). But in arguing for (CDDS), Torre relies on a theory of action that conflicts with things we say in *The Inessential Indexical* , and which we think lead to an objectionable theory of the *de se*, so we do want to show why we disagree with that theory of action.

Torre's argument for (CDDS) is the following:

> My argument for CDDS can be summarized as follows: (1) Suppose we have two subjects with different *de se* beliefs. (2) Then they will act differently or be disposed to act differently. (3) Appeal to difference in content is essential to explain the difference in action or disposition to act differently. (4) Therefore there is a difference in content

> between the two subjects. So difference in *de se* belief entails a difference in content. (Torre 2017:3)

It's step (3) in this argument that particularly concerns us. Note that step (3) doesn't itself make mention of *de se* belief (in Torre's "neutral" sense), so we don't need to engage with that idea to discuss what worries us here.

Step (3) concerns us because, of course, it's the thought that underlies one standard way of responding to the Perry cases. Consider Torre's version of a Perry case:

> Let us consider a case in which David's pants catch fire and Susan, who is standing nearby, sees it happen. Suppose David forms a belief that he expresses by saying "My pants are on fire" and Susan, upon observing David and hearing his utterance, forms a belief that she expresses by saying "Your pants are on fire". ... David will stop, drop, and roll, and Susan will run to get the fire extinguisher. What explains the difference in action? (Torre 2017:3)

The strategy in such cases, from our point of view, is to stipulate cases in which (i) two agents perform different actions, but (ii) *when we restrict ourselves to non-de-se contents* (here not in Torre's neutral sense of "*de se* belief", but in the theoretically laden sense of distinctively *de se* contents), the two agents have all the same beliefs and desires. If we add to such cases Torre's step (3), we're then forced to stick in some *de se* contents to explain the difference in action.

Before getting into the central issues concerning step (3), let's consider a somewhat subtle dialectical issue. Suppose we just conceded this point and endorsed Torre's (3). Then we'd be forced to say that there is a difference in *de se* content between David and Susan, and so (presumably) at least one of David and Susan is such that the explanation of their action appeals to *de se* contents. But notice that this is still far from an argument for (IIC). (IIC) is a necessitated universal claim: it claims that *every possible action* has some *de se* contents in its explanation. All we have here is *one particular action* that has a *de se* content in its explanation; it's fully compatible with that that other actions (maybe even one of David's and Susan's actions) don't have *de se* contents in their explanation. And thus it's compatible with this case that (IIC) is false.

Maybe we could try to generalize this reasoning to provide a full argument for (IIC). Presumably the strategy would be to say that for every action, it's possible that there be a Perry case involving that action, in which another agent has all the same non-*de-se* contents but doesn't perform that action. There's a worry that there's quite an extrapolation here from some specific cases, but we also think that even if the extrapolation were fine, it still wouldn't be enough -- all we would get from the extrapolation is that *one of the two* actions in the paired cases involved *de se* contents, which isn't enough for (IIC). That said, even if we can hold on to the rejection of (IIC), there's still more concession to the *de se* even in the single case than we'd like to make.

We'd prefer a view on which there was *never* a need to appeal to special *de se* contents in explaining action.

As a result, we want to reject step (3). That is, we reject the principle:

> (DADC) If two agents perform different actions, there is some difference in (non-*de-se*) attitude content between them.

But if difference in attitude content doesn't explain the difference in action between David and Susan, what does? As Torre notes, in *II* we suggested that the difference could be explained by a difference in what they were *able to do*. If David can perform *David stops, drops, and rolls*, but Susan cannot, then of course it's no surprise when David *does* perform that action and Susan doesn't.

Torre then notes that this proposal isn't adequate to all cases. He notes that we can remove the difference in ability between David and Susan without disturbing the crucial elements of the case:

> The appeal to difference in available action in order to explain difference in action performed is unsuccessful. This can be seen by considering a scenario in which the same actions are available to both subjects. Suppose that, *unbeknownst* to her, Susan has magical powers and is able to cast a spell that will result in the action that David stops, drops and rolls. Or perhaps, *unbeknownst* to her, her neurons are connected (perhaps wirelessly) to David's motor cortex so that she is able to perform the action that David stops, drops and rolls. Having the same impersonal beliefs and desires as David, Susan's belief-desire-obligation-intention set produces the same input action as David's belief-desire-obligation set: that David stops, drops and rolls. Furthermore, this action matches one of Susan's available actions: that David stops, drops and rolls. But Susan does not perform this action. The action switchboard appears to have malfunctioned. (Torre 2017:5, our emphasis)

We agree with Torre that such cases are possible. People can fail to act on abilities that they have, even when they have every reason to act on them. So we agree that some further explanation is called for here. (Although we'll suggest below that we should be open to the thought that there's only a rather thin sense of explanation available in some cases.)

We then want to make a two-fold response to Torre's "hidden ability" cases. First, it's also possible to construct such cases when there is sameness of *de se* beliefs as well. There are cases in which two agents have all the same attitudes *both objective and de se*, and all the same abilities, but nevertheless perform different actions. Presumably Torre's thought is that while (DADC) is false, (DADC*) is true:

(DADC*) If two agents perform different actions, there is some difference in attitude content (broadly construed to include *de se* as well as objective contents) between them.

If (DADC*) isn't true, then it's hard to see why we would accept Torre's step (3). But (DADC*) is incompatible with cases of the sort we'll go on to suggest. That much alone is enough to show that there's no consideration in favor of *de se* contents coming out of the Perry cases. If there's a puzzle here, it's a *general puzzle*. It's the observation that sometimes, despite everything anyone wanted to cite by way of explanation of action, two agents can have all of those things in common yet act differently. That would mean that there's a gap in our theory of action. But that gap would also be a gap in the *de se* friendly theory of action, so there's no reason to think that the mysterious *something* that fills the gap has anything to do with the *de se*. But second, we think that we can make at least some steps toward solving the general puzzle. We'll thus end the discussion of Torre by gesturing at what we think *does* differentiate David and Susan.

Torre's problem for us, again, is that an agent can have an ability and have every reason to act on that ability, yet still fail in fact to act on it. Why wouldn't Susan perform *David stops, drops, and rolls* when she can, and when she has reason to? A natural thought is that she doesn't know/believe that she can. That natural thought suggests a general principle:

(Control) For agent A to perform action F, A must believe that she *can* F.

Susan then doesn't perform *David stops, drops, and rolls* because she doesn't believe that she can perform that action.

But once (Control) is explicitly formulated, it's not hard to see that it can't be right. There are straightforward cases in which agents perform actions that they don't think they are capable of. (Such is the stuff that heroes are made of.) Suppose that Susan is in the middle of running an ultramarathon. She is capable of completing the race, but she doesn't think that she is capable of it. She feels exhausted, and her assessment is that she can't make it to the end. (Note thus that she can not only *fail to believe that she can perform the action*, but also positively *believe that she cannot perform the action*.) Nevertheless, she keeps running,And does indeed finish. There is, we think, nothing unusual about such a case.

Two concerns that can be raised about the case:

1. Surely at some level Susan believed that she could complete the race? We don't see why we should accept that. We want to stipulate a case in which Susan is thoroughly pessimistic. She doesn't believe at any level that she can complete the race. (Try running an ultramarathon if you're skeptical that participants reach such levels of pessimism.) Maybe you think that such cases aren't even possible, but the burden is then on you to justify that impossibility claim. (And of course an appeal to (DADC*) in arguing for the impossibility of the case would be question-begging in context. We could have made the same move in response to Torre's David-and-Susan variant.) Minor

variants of the case will show that Susan also need not believe that she *might be able* to complete the race, or that *she is capable of trying* to complete the race. None of these beliefs are needed for her to act.

2. The thought that Susan surely must at some level believe she could complete the race springs, we think, from the further concern that if she didn't believe that, there's nothing that would explain her continuing to run. But we think that's wrong. There's a perfectly good explanation for her continued running. She wants to finish the race, and running is the way to finish the race. Why wouldn't this explain her running?

In addition to such an ordinary case, we think there's also a recipe for constructing *extraordinary* cases in which agents act without the belief that they are able to perform the action. The recipe picks up on tactics that Williamson uses in objecting to versions of epistemic analyticity. We imagine agents with strange theoretical commitments that undermine certain ordinary beliefs. Thus consider:

> **Quinean Susan**: Quinean Susan has read too much Quine and become a modal skeptic. She think that modal expressions in English don't express any real concepts, and she thus doesn't believe any claims involving modals. As a result, she doesn't believe that she *is capable of finishing the race*, because that is a modal claim.
>
> **McTaggartian Susan**: McTaggartian Susan believes that there are no events. (Maybe because events are temporally extended, and she has become committed to the unreality of time.) Unfortunately, she's also a Davidsonian on the semantics of action verbs, so she thinks "running" claims involve existential quantification over running events. Since she thinks there are no such events, she thinks no one ever runs. As a result, she doesn't believe that she is capable of finishing the race, because that claim involves quantification over events.

But both Quinean Susan and McTaggartian Susan are runners. They were both experienced runners before undergoing their philosophical conversion, and they carried on their *activity* post-conversion (although McTaggartian Susan would no longer describe that activity in the same way). Both are thus examples of people who do things they don't think they can do.

(Control), then, is false. And from counterexamples to (Control), it's easy to build counterexamples to (DADC*) -- cases in which two agents have all the same abilities, and all of the same attitudes, objective and *de se*, but in which the two agents nevertheless act differently. All we need to do is add to Torre's David and Susan case the further stipulation that David, like Susan, doesn't believe that he can perform the action *David stops, drops, and rolls.* With that addition, David and Susan are perfect psychological matches both objectively and in *de se* terms. But nevertheless, David and Susan act differently. David stops, drops, and rolls, while Susan calls for help.

If you don't accept the possibility of our case (but do accept the possibility of Torre's original case), we suggest that that's because you're still in the grips of (Control). Once we remind ourselves that people sometimes do what they don't believe they can do, we should see that David could be one of those people. And if he is, then our case is possible, and it counterexamples (DADC*), refuting step (3) and undermining the positive case for *de se* contents.

One final point. We think we've said what needs to be said to respond to Torre's challenge to our view at this point. But we agree that a mystery still remains. In our modified David-and-Susan case, David acts on an ability he has but doesn't think he has, while Susan doesn't act on an ability she has but doesn't think she has. That's a difference between David and Susan, and one that we haven't offered an explanation for. All we've done is argued that Torre *also* doesn't have an explanation for it, and thus that these sorts of differences don't give us reason to posit special difference-explaining *de se* contents.

But we also suspect that lingering mysteries are one of the things that push people to get interested in *de se* contents, and that if the mystery lingers, people will just convince themselves that there's some *other de se* difference between David and Susan that we've missed. (We're not sure why people find acceptable to *stipulate* a lack of objective difference between David and Susan, but still allow themselves to *discover* a *de se* difference. We're a bit suspicious that it's because the *de se* contents are just a magical black box, and that since we don't know what they really are, people don't take seriously stipulations about them.) So we want to offer some brief gestures toward an explanation of why people sometimes do and sometimes don't use abilities they have, when they have reason to use them.

David and Susan both enter a race. Both are capable of both running the race and skipping through the race. But neither thinks (in the manner we've discussed above) that they have either of these abilities. Nevertheless, David runs the race, and Susan skips along the race. Why, then, does David run and Susan skip?

Roughly, we want to say: because those are their habits. David is a habitual runner, so in situations that can be dealt with by running, his habit is to run. He's developed a disposition to deploy that ability, independent of his beliefs about the ability. Susan is a habitual skipper, so in situations that can be dealt with by skipping, her habit is to skip. She's developed a disposition to deploy *that* ability, independent of her beliefs about the ability.

We enter this world with a collection of abilities, but with no information about what abilities we have. Early in life, we stumble on to some of those abilities during the general flailing-about of infancy. The reward mechanism of that stumbling-on produces in us a disposition to try more things of that sort (a disposition that, again, need not be reflected in a *belief).* Two people can both wiggle their ears. One does and the other doesn't, at the crucial time -- that's because one is a wiggler in general, and thus disposed to try wiggling solutions, while the other isn't. If Susan has magical abilities to control David's action, she presumably doesn't use them because she's

not in the practice of exercising magical abilities, and thus it never occurs to her to act in that way. (Note that Susan's failure to act becomes more puzzling if we build into the case that she's a practicing magician.)

Note the similarity here to our earlier case of the CEO and his assistants. There could be two executives E1 and E2 with two assistants A1 and A2. E1 and E2 both order their assistants "win the race!". But A1 can make E1 run, but not skip, so A1 starts E1 running down the course. A2, on the other hand, can make E2 skip, but not run, so A2 starts E2 skipping down the course. The result is that E1 and E2 act differently, in a way that reflects their different abilities to act, without ever knowing about what they can and can't do. They act differently because they have different assistants -- that is, different non-attitudinal mechanisms for implementing their goals.

Reply to Dilip Ninan

We turn now to Dilip Ninan's objection in his (....) paper. Ninan says:

> A notable feature of Cappelen and Dever's proposal is the use of agent-specific action types such as 'the action that DN curls up'. But do Cappelen and Dever also allow into their ontology agent-neutral actions, such as the (much more familiar) action of curling up? Unlike the former, the latter is an action that agents other than me can perform. Cappelen and Dever certainly write in various places as if they do accept action types of this sort (e.g. Cappelen & Dever 2013, 47). And it is a good thing too, since the idea that no two agents can perform the same action is absurd. (Ninan 2017: 106)

Ninan presents us with the following familiar scenario: A and B both think a bear is attacking A, but they act differently. A curls up in a ball, and B runs away. The first draft of the problem says:

> *First draft of problem*: that's mysterious. Why do A and B perform different actions if they have the same beliefs and desires?

Ninan then concedes that we have a response to this:

> *Response to first draft of problem*: the action in question is "A curls up in a ball". Both A and B have reason to perform this, but only A can, so only A does.

Ninan then has a second draft of the problem:

> *Second draft of problem*: but what about another action: curls up in a ball. A and B can both perform this action, but only one does. Why?

First stab at an answer: “curls up in a ball” is an action *type*, not an action. And we never said that if two people had all the same psychological states, they would perform the same action types. So there’s no problem here.

Second stab at an answer: “A curls up in a ball” is an action type, too. It must be, if we’re contemplating both A and B doing it. Two people can´t both perform the very same token action.This shows that there’s something wrong with Ninan’s Explanation principle:

> *Explanation*: Suppose the fact that x performed action α is explained by the fact that x has beliefs B^x_{p1} , ..., B^x_{pn} and desires D^x_{q1} , ..., D^x_{qn} . Then, if y has beliefs By B^y_{p1} , ...,B^y_{p2} and desires D^y_{q1} , ..., D^y_{qn} , then, other things being equal, y will also perform α. (Ninan 2016: 102)

Put a bit loosely:

> If two agents have all the same (relevant) beliefs and desires, then, other things being equal, they will behave in the same way.

Note that this must mean they will perform the same action *types*, since they have no chance of performing the same action tokens. Once we see this, we see that we need to pay attention to the way actions are typed. Consider the type “Action performed by A”. It´s definitely not true that if both of A and B have the same beliefs and desires, then A performs an action of the type “action performed by A” if and only if B performs an action of the type “action performed by A”. That way of typing actions isn't a way that we expect to be captured by patterns of psychological attitudes. And there are lots of types like this, not all as trivial as the “action performed by A” type. Consider the type “dangerous action”. If A and B have all the same beliefs and desires, does it follow that A performs a dangerous action if and only if B performs a dangerous action? No. A might be courageous and B cowardly, so that A acts on the beliefs and desires, and B doesn´t. Or A’s situation might be dangerous and B´s not. Again, belief desire psychology is the wrong place to look for explanations of that kind of action type similarity.

So here’s the general picture. There are many many potential token actions, and there are many many potential agents. As a result, there are:

- Many different relations between agents (*having the same psychological state, being in the same room as, being braver than, being identical to*).
- Many different ways of typing actions, where an action type can be thought of as a set of action tokens.

Now pick some relation R between agents and some typing T of actions. We can then ask the following question:

(Typing Question) Does relation R guarantee T-sameness in action? That is, is it true that if agent A stands in relation R to agent B, and A performs an action of type T, then B also performs an action of type T?

The answer to the Typing Question will vary depending on both R and T. At one extreme, if R is the identity relation, then (trivially) the answer to the Typing Question is "yes" for any type T. But for any choice of R other than the identity relation, the answer to the Typing Question will sometimes be "no". (For example, when T is *action performed by A*.)

When we're in the business of understanding action, then, one thing we might be doing is presenting various choices for R relations, and making correlation claims about what action types T get preserved by a given choice for R. Ninan endorses one very specific such correlation claim:

- When R is the relation of *having the same belief and desire contents*, then we get correlation with respect to any type T of the form *is an action whose content is given by ϕ(a) for some object a*.

So Ninan endorses a principle according to which, if A and B have all the same beliefs and desires, and A performs some action N, then B also performs an action whose content is the same as the content of N, but with mention of A in the content replaced by mention of B. If A performs *A runs away*, then B performs *B runs away*.

We, on the other hand, claim that correlation happens at a slightly different place. We endorse:

- When R is the relation of *having the same beliefs and desires and available actions*, then we get correlation with respect to any type T of the form *is an action whose content is given by* ϕ.

Since no one can get any *perfect* correlation principle, getting correlation for all types T (for any R relation short of identity), the question is just whether we are getting *enough* correlation. Enough for what? Presumably, enough for our ordinary concepts of belief and desire to play a reasonable role in predicting, explaining, rationalizing, and causing ordinary actions. We think we are. We're not claiming a correlation pattern that does *all* the work of explaining why people do what they do, but we've argued that no one can have such a pattern, so that's not the success condition. But we are claiming a correlation pattern that's robust enough to make sense of ordinary folk psychology, and as long as we have that, we don't take it as a serious objection that we don't get exactly Ninan's preferred correlation.[4]

[4] It's worth noting that we can get pretty much the pattern that Ninan wants. We agree that if we group agents not by sameness of belief desire content, but by sameness of belief desire content minus specification of agent, then we get prediction of action typed at a level of propositional content minus agent. This in fact seems like a natural pattern to

Reply to José Luis Bermúdez

José Bermúdez in "Yes, indexicals really are essential" and in Chapter 1 of his book Understanding "I": Language and Thought (Bermúdez 2017) responds in a three step argument:

> *Premise 1:* No action rationalization can correctly reconstruct an agent's practical reasoning if it is possible for some agent to hold every propositional attitude in the set and not perform the action.
>
> *Premise 2:* Even if an agent holds every propositional attitude in an impersonal action rationalization, she will not perform the consequent action if she believes that she is not the person referred to in the action rationalization.
>
> *Premise 3:* For any impersonal action realization it is possible for an agent to hold every propositional attitude in the set and nonetheless believe that she herself is not the person referred to in those attitudes.

From these three premises, Bermúdez derives:

> (IICa): Impersonal action rationalizations are necessarily incomplete.

We reject (IICa), and Bermúdez's argument for it. We give two replies to the argument. In our first reply, we observe that premise 2 begs the question against an important element of our view. In our second reply, we observe that Premise 1 unnecessarily excludes all *ceteris paribus* action rationalizations. We show that a consequence of this exclusion would be the essentiality not just of indexical beliefs but of all beliefs, and on the basis of this observation conclude that we should allow *ceteris paribus* rationalizations.

We also have a third point of disagreement with Bermúdez. He takes (IICa) to be insufficient to establish genuine essential indexicality -- he wants to show that impersonal action rationalizations are incomplete *because* they do not cite indexical attitudes. He thus argues that action rationalizations that are neutral on indexical matters, requiring neither that the acting agent has nor that the acting agent does not have such attitudes, are incomplete. In a third reply, we also reject Bermúdez's further claim that neutrality on indexical matters creates a problem for completeness of action rationalizations, observing that there are straightforward cases in which neutrality is unproblematic.

First Reply: Bermúdez's argument assumes incorrectly that all impersonal action rationalization is subject-referring.

---

expect, better than the patterning Ninan looks for. And we think it will let us do anything more in the way of psychological explanation that Ninan might still have wanted after our first proposition level typing principle.

In the book we introduce what we call *impersonal action rationalizations*. Those come in two kinds and these are our illustrations:

> *Impersonal Action Rationalization 1.*
> • Belief: François is about to be shot.
> • Desire: François not be shot.
> • Belief: If François ducks under the table, he will not be shot.
> • Action: François ducks under the table.
>
> *Impersonal Action Rationalization 2.*
> • Belief: Nora is in danger.
> • Desire: Nora not be hurt.
> • Belief: If the door is closed, Nora will be safe.
> • Action: Herman closes the door.

Throughout the book, we point out an important feature of *Impersonal Action Rationalization 2*: the propositional attitudes referred to in the rationalization *don't mention or refer to the agent, i.e. to Herman*. We say: "Note that in this case the impersonal action rationalization doesn't attribute "Herman"-beliefs or "Herman"-desires to Herman. Instead, the rationalization is entirely third-person." (37) Our picture was this: When Herman sees his daughter in danger, he just acts. Not via some representation of himself - he doesn't, for example, need to go via the belief that Nora is Herman's daughter. He just acts when Nora is in danger.

With that in mind, consider Bermúdez's Premise 2. In saying "she will not perform the consequent action if she believes that she is not the person referred to in the action rationalization," this premise simply presupposes that all action rationalisations are of the subject-referring kind and that none are of our second kind. Bermúdez gives no argument for that assumption.[5]

In sum: If you take into account the non-subject referring cases that we rely on throughout the book, Bermúdez's argument is unsound because premise 2 is false (as is assumption 3, as noted in footnote 4). We are of course open to arguments to the effect that we shouldn't have included such cases, but the focus of Bermúdez's paper isn't to establish that, it simply assumes it. Moreover, several of the models in the first part of this paper share that feature with Impersonal Action Rationalization 2: the gods, for example, when they act simply by intending that the world be so-and-so, need no mention of themselves in the explanations of their actions.

---

[5] Bermúdez's Assumption 3 also presupposes that there are no non-subject referring rationalisations:

> *Assumption 3*: An impersonal propositional attitude is one that refers to the agent non-indexically – e.g. through a proper name or non-indexical definite description.

Second Reply: Bermúdez's argument assumes incorrectly that all practical reasoning is not *ceteris paribus*.

Our first reply shows that Bermúdez's argument has no force against non-subject-referring action rationalizations. In our second reply, we show that it also fails even when its scope is restricted to subject-referring action rationalizations. Bermúdez's Premise 1 places a very strong necessity requirement on such action rationalizations. For a set S of propositional attitudes to reconstruct an agent's practical reasoning, it must be *impossible* for any agent to hold all of those attitudes and not perform the action.

But as stated, this principle is clearly false. Suppose Alex wants a drink of water and believes there is water in the glass, and she then picks up the glass. *Prima facie*, we'd like to say that her desire and belief explain her action. But it's *possible* for Alex to be tied to the chair, or to be paralyzed, or to be given drugs that interfere with her practical reasoning. In any of these cases, she would *not* pick up the glass. These possibilities, however, do not prevent her desire and belief from rationalizing her actual (unimpaired) action.

Action rationalizations, like most explanations, are typically *ceteris paribus*. They take place against a background of normal conditions, and they don't bring about the explanandum when conditions are sufficiently abnormal. Bermúdez's Premise 1 disallows *ceteris paribus* explanations, because it requires that there be no possibility of having the explanans without getting the explanandum as well, no matter how abnormal the situation. Premise 1 should thus be rejected. Without Premise 1, Bermúdez's argument collapses, even if (contrary to our remarks above) Premise 2 is accepted.

Suppose we rewrite Premise 1 to allow *ceteris paribus* explanations:

> Premise 1*: No action rationalization can correctly reconstruct an agent's practical reasoning if there are normal conditions in which some agent holds every propositional attitude in the set and doesn't perform the action.

For Bermúdez's reasoning to go through, Premise 3 would need to be revised to guarantee that the addition of identity-confusion beliefs fell under normal conditions:

> Premise 3*: For any impersonal action rationalization there are normal conditions in which an agent holds every propositional attitude in the set and nonetheless believes that she herself is not the person referred to in those attitudes.

But we see no reason to accept Premise 3*. Our grip on 'normal conditions' is through our understanding of the sorts of agents we are trying to explain, so to convince us that adding identity-confusion beliefs always preserves normality, you have to convince us that our explanatory practice isn't sometimes *designed* for non-identity-confused agents.

To see what's going on more clearly, remember that identity-confusion beliefs are not the only beliefs that will be covered by *ceteris paribus* clauses. Alex desires water, and she believes that there is water in the glass. But if she believes that everyone on earth will be tortured if she picks up the glass, she will not pick up the glass. We would be inclined to think that a belief that global torture results from lifting a glass is an *abnormal* belief, and thus gets swept up into the *ceteris paribus* clause. Consider an analogue of Bermúdez's argument. Call an action rationalization atortural if it says nothing about whether torture results from lifting glasses. Then:

> Premise 1*: No action rationalization can correctly reconstruct an agent's practical reasoning if there are normal conditions in which some agent holds every propositional attitude in the set and does not perform the action.
>
> Modified Premise 2: Even if an agent holds every propositional attitude in an atortural action rationalization, she will not perform the action if she believes that global torture will result from performing it.
>
> Modified Premise 3*: For any atortural action realization there are normal conditions in which an agent holds every propositional attitude in the set and nonetheless believes that global torture will result from performing the action.

We then conclude that atortural explanations are always incomplete. Perhaps we've just discovered a new kind of essentiality. In addition to indexical essentiality, there's also torturous essentiality. No explanation of action without beliefs about torture. We would rather modus tollens on the implied conditional, and conclude that neither kind of attitude is essential, because both cases are abnormal. But for those not convinced, we note that the new essentiality argument is very powerful. Take any belief B of Alex's. If Alex believed that if B and she lifts the glass, then everyone on earth will be tortured, then she will not perform the action. So for any B-free action rationalization, we can find a B-containing expansion of it that blocks realization of the action. If that expansion is normal, then the B-free rationalization will be incomplete.

Third Reply: Neutrality on indexical matters is not a threat to the completeness of action rationalizations.

We have now rejected Bermúdez's IICa twice over. We thus don't think that there is any incompleteness in impersonal action rationalizations to be repaired. As mentioned above, Bermúdez proceeds to reject one specific strategy for providing (unnecessary, in our view) repair for impersonal action rationalizations. We briefly discuss this portion of Bermúdez's response to show that no plausible separate argument for essential indexicality can be extracted from it.

Call an impersonal action rationalization *indexically neutral* if it not only does not cite any indexical attitudes of the acting agent, but actually requires that the acting agent be neutral on

all indexical matters (neither believing nor disbelieving them). Bermúdez suggests that indexically neutral action rationalizations cannot explain action:

> If it has never occurred to the agent that she might be phi, why on earth would she act upon a set of propositional attitudes conceptualized in terms of phi? If it has never occurred to me that I might have won the raffle, why would I go to collect the prize? (Bermúdez 2017:xx)

This consideration, of course, has potential force only against subject-referring action rationalizations. But even when directed against subject-referring rationalizations, we are unconvinced. We can ask the same questions about no-torture beliefs. Suppose Alex neither believes nor disbelieves that if she lifts the glass, there will be global torture. Why would that make it mysterious that she then act upon a set of propositional attitudes conceptualized in terms of glass lifting? There is no grip to the question: *if it has never occurred to Alex that if she lifts the glass, there might or might not be global torture, why would she lift the glass?* There is no reason for Alex to entertain global torturing contingencies in getting her thirst and glass beliefs coordinated into action -- she's just not built to need to consider (without special prompting) those sorts of alternatives. And things could easily be the same way with us. We don't need to consider whether we are the owners of the body parts we see about us to use beliefs and desires about those body parts to get us into action -- we're just not built to consider (without special prompting) those sorts of alternatives.

So we don't see any reason to think that there is any special role for indexical attitudes here. There are infinitely many propositions, drawing on every possible conceptual resource, that are potential barriers to realization of action. It can't be right that *all* of those propositions are essential to action rationalization (either to block their negations, or to block indifference with respect to them). We've been given no reason to think that the right thing to say about these infinitely many other cases doesn't extend equally well to indexical cases.

A Diagnosis Of Why Our Opponents Won't be Convinced

The reactions to our book have struck us as interesting in at least two ways: First, we've been encouraged by the fact that the defenses of Essential Indexicality have been (i) varied, and (ii) have not appealed to arguments in the original literature. We had a slight worry, having finished the book, that there was some kind of obvious point out there that we had overlooked or failed to grasp that the essential indexicality proponents would all converge on. However, a) no such convergence has occurred, b) no one has claimed that we fail to respond to the articulated arguments in the literature that existed before our book, and c) all the responses have been different. That, to us, indicates that there weren't any good arguments for essential indexicality, i.e. that our conclusion was right.

The second thing that has struck us is that no one in the essential indexicality camp seems to have changed their mind. They tend to think of us as sceptics who throw up weird sceptical scenarios that are hard to refute, but don't convince.There's a spectacularly strong attachment to the significance of the first person perspective and a very strong reluctance to take de se skepticism seriously. We will end with some brief reflections on why the de se has such a firm grip on many thinkers - why essential indexicality is an idea that strikes many so obviously right that hardly any arguments can shake their conviction. We suspect that at least three factors play a role:

(i) *Obsession with Misidentification Cases*: We find that many of the arguments just keep returning to the misidentification cases. We'll present an argument against a reply that has nothing specifically to do with misidentification, but at the end, the respondent tend to return to the question: "Well, but what if I make it into a misidentification case, what do you say then?" Here, as a reminder, is how that discussion will go:

> *We*: Well, our case *isn't* a misidentification case.
> *Opponent*: Okay, but it *could* have been one.
> *We*: Right, but that's another case, our case doesn't involve any kind of identity confusion. Look, for the sake of argument we'll grant you that identity confusions will affect action explanation in various interesting ways. But we just need *one single* case where there's no de se content involved and *we choose to focus on cases where the agents have no identity confusion*.

Maybe a source of the resistance to putting aside cases that involve identity confusion is the thought that if Susan is not identity confused - i.e. doesn't believe *I'm not Susan* - then she must believe another de se content: *I am Susan*. So, maybe the thought is, if she doesn't believe *I'm not Susan*, then she must believe *I am Susan*. One of these *must* play a role in the action explanation/rationalization. But this, as we saw in the reply to Bermúdez, is a mistake - Susan don't need to believe either *I'm Susan* or *I'm not Susan* in order to act. In the cases we imagine, neither of these play a role. In response to this we tend to get just a brute insistence that it *must*, but that's where the arguments run out (or they turn into one of the arguments above - and then those arguments can't return to misidentification data alone.)

(ii) *The Multiplicity of Argumentative Avenues*: We started *The Inessential Indexical* by noting that there are many different ways to argue for an essential indexicality thesis. Some arguments have to do with agency, some have to do with semantics and opacity, some have to do with epistemology and others again with perception. And that's not an exhaustive list. For example, in response to our book, L.A. Paul has argued that we've looked at the wrong motivation: empathy is the important phenomenon to focus on, she says (Paul 2017, Cappelen and Dever 2017). In *The Inessential Indexical*, we take several of these arguments on one by one and show that they fail. However, the problem is that if the focus is on a particular argument, then it will always seem that one of the others can pick up the slack, and so there's a tendency to gradually sneak over to another arguments when one starts crumbling. For example, what we

say about the irrelevance of identity confusion for explaining action doesn't explain opacity and the failure of substitutivity that will arise. That's a separate argument and it's tempting for de se proponents to slide over to it. What this shows is that to become a de se skeptic takes a lot of work: you have to convince yourself that all these pillars crumble and since each of the pillars have the weight of both tradition and authority behind it, it's easy to understand that the whole framework is hard to liberate oneself from. In sum: De se scepticism is hard work, but since it is true, and its truth is significant, it's worth it.

(iii) Finally, our opponents tend to retreat to the claim that we're attacking a straw-woman or man. The strong modal claim, they say, is too strong. What proponents of essential indexicality had or should have had in mind is that contingent features of actual human beings make it the case that we have to rely on de se contents. This isn't an essential feature of agency - it's an empirical claim about human beings and the way they happen to act. In *The Inessential Indexical*l we give two responses to this objection. First, we prove that the stronger claim has been repeatedly defended and relied on in the literature (so we are not misrepresenting the view - the thesis we attack is the thesis that has been repeatedly proposed). Second, an argument for the weaker claim would have to based on empirical research into human agency, not armchair reflection:

> Whether indexicality (or "the *de se*") is involved is a very, very detailed question about the implementation of complicated mechanisms in the human head. Armchair reflections about us moving our fingers won't get us such conclusions; nor will philosophical reflections about generalizations or opacity. In saying this we are not taking a stand on how action mechanisms are in fact implemented in human heads. It would be absurd for us to think that such a complicated empirical issue can be settled that way. We're just saying that our opponents don't know how the implementation goes, and that there are feasible non-indexical implementations.

We stand by these two components of the reply. One point we didn't elaborate on in the book was the last conjunct of that passage. Here we've provided some further elaboration. We continue to think that armchair speculation into these complicated empirical matters should be treated with caution and suspicion. But if people want to play the armchair speculation game, we're willing to play it too, and we've attempted here, by setting out some possible non-indexical implementations in more detail, to show why we think it's compelling not just that it's *possible* for there to be creatures like that, but that we ourselves are creatures that often act without any indexical thoughts.

References

Bermúdez, Luis Jose. 2017. Understanding 'I': Language And Thought. Oxford: OUP.
Cappelen, Herman & Dever, Josh (2017). Empathy and transformative experience without the first person point of view. *Inquiry* 60 (3):315-336.
Paul, L. A. (2017). First personal modes of presentation and the structure of empathy. *Inquiry* 60 (3):189-207.

Cappelen, Herman & Dever, Josh (2013). *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*. Oxford University Press.
Ninan, Dilip (2016). What is the Problem of De Se Attitudes? In Stephan Torre & Manuel Garcia-Carpintero (eds.), *About Oneself: De Se Thought and Communication*. Oxford University Press.