Modeling De Se Belief

Bernard Molyneux

molyneux@ucdavis.edu

Paul Teller

[prteller@ucdavis.edu](mailto:prteller@ucdavis.edu)

**Abstract**

We develop an approach to the problem of de se belief usually expressed with the question, what does the shopper with the leaky sugar bag have to learn to know that s/*he* is the one making the mess. Where one might have thought that some special kind of “de se” belief explains the triggering of action, we maintain that this gets the order of explanation wrong. We sketch a very simple cognitive architecture that yields de se-like behavior on which the action-triggering functionality of the belief-state is what counts it as de se rather than some prior property of being “de se” explaining the triggering of action. This functionality shows that action-triggering change in belief-state also undergirds a correlative change in the objective involved in the triggered action. This model is far too simple to have any claim to showing how the de se works for humans, but it shows, by illustration, that nothing mysteriously “subjective”” need be involved in this aspect of self-conception. While our exposition is very different from those of Perry and Recanati, all three of us are developing the same kind of view.

**1. Setting up the Problem.**

**1.1 Introducing the Castañeda-Perry problem.**

Perry (1979), pushing his grocery cart in the supermarket, notices a trail of sugar. Clearly, someone has a leaky sugar bag in their cart. Perry, not realizing that the leaky sugar bag is his own, sets out to alert the messy shopper. For Perry to succeed in getting the leaky sugar bag patched, it appears that Perry must change his belief-state to one expressible with “I”: e.g. to a state with the content *I am the one making the mess!* Learning that Perry is making the mess won’t trigger the bag-fixing action if he doesn’t know that *he* is Perry, i.e. if he doesn‘t come to believe what would be expressed by the sentence “I am Mr. Perry.” Similarly, learning that the man in the loud shirt seen in the mirror is making the mess won’t trigger action unless he comes to believe what is expressible as “I am the man in the loud shirt seen in the mirror.”[[1]](#footnote-2)

Here is a second case. Perry works for the lottery with the job of sending out the winnings. Perry learns that one Mr. Perry has won, but doesn’t appreciate that *he* is Mr. Perry. In this case Perry acts – he sends the winnings to the Mr. Perry at the address of record. But there is something odd about the act, something different from what would happen if he knew that *he* was Mr. Perry, that *he* was the winner.

These issues were discussed earlier by Hector Castañeda (1968 and other articles), and we will refer to them as “the Castañeda-Perry (CP) problem/phenomenon.” These have come to be known more broadly as issues involving beliefde se. Stated more specifically, the CP problem involves understanding how Perry’s belief-state changes when he is ready to say, “*I* am the one making the mess” and “*I* am the winner”. What do messy shoppers who are ready to say, “*I* am the one making the mess” have in common? What is the role of this common state in activating action?

The traditional statement of the CP problem can leave one with the impression that action is triggered by acquiring a belief with a special “de se” character. We maintain that this gets the order of explanation wrong. We will proceed by asking, “What functional structure of a cognitive architecture is required to trigger action?” As a result we will see that it is the action-initiating change in cognitive architecture that makes a belief or a belief-state count as de se rather than some de se character being responsible for initiating action. No prior notion of self or self-ascription need be assumed.

In his earlier work, Perry claimed that indexicals like “I” are essential to the explanations of the messy shopper’s action, that “When we replace [the word “I”] with other designations of me, we no longer have an explanation of my behavior” (Perry 1979, p. 3). But he later changes his stance on this, noting that “[i]t is not really the first person that is needed. It is the function that the first person performs” (2011, p. 385, see also Perry 2003). In changing his view, he observes that proper names can also be used in the required way: One might knowingly refer to oneself using a third-person term – e.g. when Tarzan says “Tarzan hungry!” he expresses the de se attitude that he himself is hungry. (Perry uses the example of a child called Eloise who self-refers using her proper name.)[[2]](#footnote-3)

Just as the de se role of the first-person indexical can be played by other terms, so it is that the first-person indexical can be separated from that role and be made to be understood non-de se. Imagine, for instance, that Castañeda hurts his head and loses his memory, including all idea of who he is. In this state he reads a newspaper story about a missing philosopher, and, not knowing that he is reading about himself, says, “I do hope that they find that poor philosopher and he is able to return to his life of philosophy.” When at a later time his memory has returned he reports this incident by saying, “I hoped, not realizing that it was myself that I was thinking about, that I would return to my life of philosophy”. The third occurrence of “I” and the one occurrence of “my” are de re but not de se, while the other two occurrence of “I” and the occurrence of “myself” are de se.[[3]](#footnote-4) He did not, after all, wish the self to return to a life of philosophy but a person he considered as other.[[4]](#footnote-5)

Clearly, then, the issue cannot be simply formulated as a need for indexicals, as the early Perry sometimes seemed to suggest. The de se belief that is required for action can be expressed using a name or other referring expression rather than an indexical, and the term “I” does not always force the de se understanding.

**1.2 The “!”/”¡” notation**.

Let’s put the issue in the third person. A natural way to describe our central case is: Perry must learn that *he* is the one making the mess. Part of Perry’s point, however, is that this way of putting things does not distinguish the intended understanding from possible rivals. If Perry learned that some guy called “John Perry” was making the mess, without identifying that person with himself, he would nonetheless count as learning that he is making the mess on an alternative de re understanding. It is tempting to amend this by saying that Perry must learn that it is he *himself* who is making the mess, since “himself” pushes us more strongly towards the de se understanding. But the “himself” qualification permits non-de se understandings as well. Suppose that John is in a boat sailing down a river while checking his phone, where he is watching a satellite image of the progress of a boatsman called “Perry” also heading down a river. Ahead of Perry’s boat (according to the satellite image) is a waterfall. So John sends Perry a text message saying “You’re heading for disaster!!” Moments later, John receives a text message from Perry but he drops his phone into the water before he can read it. Now suppose that Lewis, standing on the shore, knows that John and Perry are one and the same person, and is otherwise completely apprised of the situation. So Lewis says to himself, “In thinking that Perry is heading for disaster, John (unknowingly) thinks that *he himself* is heading for disaster.” Lewis, understanding fully what is going on, does not intend to make the false claim that John believes de se that he himself is heading for trouble - if he did he would expect John to take action, but John does not. Rather, Lewis uses “himself” to emphasize the fact that John is referring to himself, but in this case unknowingly. Here, then, is a case where “himself” does not force the de se understanding.

Since “himself” won’t unambiguously mark off the de se understanding from other possible interpretations we need something that will. Accordingly, when we say that Perry learns that he is the one making the mess and intend this to be understood as de se belief, we will write: “Perry learns that he! [read “he-bang”] is making the mess.” We will use “he¡” [read he-”whimper”] when we intend the non-de se understanding. We will write ”he” with no ”!” or “¡” when we want to be neutral between the two ways of understanding such statements.[[5]](#footnote-6) [[6]](#footnote-7)

**1.3 Difficulties with Lewis’s account and our present objectives.**

In this section we will briefly characterize Lewis’s effort to understand de se belief-states, explain what we see as shortcomings in his account, and set out in what way we claim that we will do better.

In his (1979) Lewis proposes taking self-ascribed properties[[7]](#footnote-8) rather than propositions to be the objects of belief. He takes propositions to be sets of possible worlds and interprets belief in such a proposition as the self-ascription of the property of being located in one of the worlds of the proposition. Since an agent is in a possible world just in case all the world’s other denizens are in the world, such a belief can’t capture the de se character of beliefs that, like believing that one is the messy shopper, distinguish one from all other actual existents. Lewis’s strategy is to fill the gap with self-ascription of more specific, “locating” properties, such as being located at a specific place, being the messy shopper, or the like.

As Lewis recognizes (p. 543), there is a problem with the proposal as so far stated: An agent may ascribe a property to someone who they do not appreciate are themselves, thereby self-ascribing the property, but not in the ! way. Using Lewis’s example, imagine that an agent sees someone’s image reflected in a window, sees that their pants are on fire, but does not appreciate that the person they are seeing is themself. “That person’s pants are on fire!” is then a self¡-ascription. Lewis addresses this issue by appealing to a notion of acquaintance. Taking *the person I am watching* as a clear instance of acquaintance, Lewis then analyzes the pants-on-fire case as one in which the agent self-ascribes the property of watching someone who has their pants on fire. This is different from self-ascribing the property of wearing a pair of pants that are on fire, which would be the truly ! de se case. Thus Lewis draws the needed distinction.

Yes, Lewis has provided a way of drawing the distinction between ! and ¡ belief about oneself, but no, he hasn’t offered any kind of analysis of that distinction. The problem is that Lewis runs his account in terms of self-ascription, but he offers not a word on how self-ascription is to be understood. For Lewis, to believe that one’s! pants are on fire is to self-ascribe the property of wearing burning trousers, with no further details given; so we know no more about what it is to self-ascribe the property than just to believe that oneself! has the property, which is exactly what we wanted to understand.

Here is a more detailed way of setting out the gap in Lewis’s account in understanding the de se. In order to ascribe a property to a person or thing, one must have a way of representing both the property and the thing (the referent). Thinking of the pants-on-fire case, let’s characterize the means of reference as “d”, which could be a simple referring term or some more complex referential device. Nothing has been said about how “d” fits into the agent’s larger cognitive architecture. For all the analysis that has so far been given, “d” just has a referent, period. Let’s mark this neutrality in the analysis as so far given by calling this referent “the designated agent”, with no further explanation. To self-ascribe pants on fire to oneself is to ascribe pants on fire to the designated agent. So what?

Of course the *agent* knows that the designated agent is her/himself and so will take fire-quenching action. But what is involved in knowing that? The analysis, so far, is entirely dumb about this question.

The illusion that some analysis of the de se has been given is strengthened by Lewis’s comment that “Self-ascription of properties is ascription of properties to oneself under the relation of identity. Certainly identity is a relation of acquaintance par excellence.” (p. 543). But this adds nothing to the analysis. The property expressed by “having one’s pants on fire” is the same property as expressed by “being (identical to) someone who has their pants on fire.” Nothing has been added. For appeal to identity to be substantive, there have to be two logically independent referring terms or devices with identity asserted as holding between the referents of the two terms or devices. The analysis so far has, and at that only tacitly, assumed one referential device, which we are marking with the expression ‘the designated agent”. For appeal by the agent to identity to be substantive the agent would have to have some second referring device, perhaps “the self”, with identity asserted between the designated agent and the self. But this just puts off the needed analysis.[[8]](#footnote-9)

We will approach the issue afresh by asking how one might design an autonomous robot with de se behavior. In so doing we will arrive at an account that has much in common with those of Recanati (2009, 2012) and Perry (1998, 2011, 2012). Relative to their efforts what we offer will provide a simple model that shows, by illustration, the kind of thing that we take them to be characterizing quite abstractly: that the needed work is done by the referential device, “the designated agent”, being tied into the agent’s larger cognitive architecture. To be emphasized, our model is simple in the extreme – getting the details right is a job for the sciences. Our objective will be to show what kind of theoretical structure promises to unravel the de se puzzle. While the kind of view that we develop agrees very well with that of Perry and Recanati, theirs and ours have complementary virtues and deficits. They describe the role of cognitive architecture in an extremely abstract way that makes it difficult for some readers to get a real grip on how the scheme is supposed to work. Ours is transparently specific, making it easy to see how it functions, but at the cost of extreme simplification. Where we will call our scheme a model, Perry (2010, pp. 240, 241; 2012, pp. 86, 87 (note 6)) and Recanati (2009, pp. 5,6) say that they are offering a metaphor, but clearly we are doing similar sorts of things (In 2012 pp 80, 86, 87 note 6 Perry also uses the term “model”.)

Our starting point is that something must be said that provides a connection with action, that the designated agent must be in some special state that, given the right desires, will trigger action. But what action? And how? In the next section we show that the first question constitutes a dimension of the CP problem not generally appreciated.[[9]](#footnote-10)

**1.4 Refining the statement of the CP problem.**

There is a contrast not only between de se (!) and non-de se (¡) beliefs, but also for actions. Consider the following two cases:

John’s acting with the objective to make it the case that John! is rich

John’s acting with the objective to make it the case that John¡ is rich

These count as different act types since the first counts as (*prima facie*) selfish, the second as (*prima facie*) altruistic. This distinction in act types eliminates any temptation to think de se belief is needed for self¡-directed de re action. Consider the following pairs of cases:

Case 1: Perry realizes that his! is the leaky bag and fixes it.

Case 2: Perry leaves his cart looking for the messy shopper. He comes upon his own cart, not realizing that it is his! own, sees the leaky bag and fixes it

A second pair of cases also illustrate this contrast:

Case 1: Perry realizes that he! has won – and mails off the winnings to himself!

Case 2. Perry does not realize that he! is the winner, thinking only that he¡ is the winner. He sends off the winnings to Mr. John Perry¡.

Clearly the contrast between ! and ¡ actions, as well as between ! and ¡ beliefs applies to the CP cases, showing that de se belief is not needed for self¡-directed de re action. We need a more careful statement of the whole complex of the CP problem.

The CP problem has two parts or sides. The problem can be seen as that of explaining a special kind of act by appeal to a special kind of belief-state. So there is a first problem: What is the *explanandum*? What is the nature of the acts in cases 1 and how do they differ from the acts in cases 2? Second problem: What is the *explanans*? What is the special nature of belief-states that triggers the case-1 kinds of actions? These problems are distinct but, intuitively, deeply connected. The functionality of the cognitive architecture that we will sketch will provide a unified understand of both the *explanandum* and the *explanans*.

**1.5 Beliefs, belief-states, actions, and dressed propositions.**

To set up for our analysis we need to introduce a few more ideas and terminology. We will use the following methodology in analyzing belief-states: We assume that, at the present state of knowledge, no analysis of belief-states will in every respect be either complete or completely correct. Rather, an account of belief-states may be more or less detailed and more or less accurate, and in a variety of respects answering to one or another theoretical (or practical) interest. So an account – or model – may be relatively detailed, or as we will say, fine grained, or less detailed, that is, coarse-grained.[[10]](#footnote-11) With this methodology in mind we will often follow common practice and describe belief-states using description in terms of beliefs with propositional or proposition-like objects as long as we understand that so doing need not be a complete description. To keep this point firmly in mind we will occasionally insert a reminder.

We can illustrate this methodology by using it to clarify the de re/de dicto distinction. Ralph believes that the man he saw in the trench coat is a spy. Ralph’s “de re belief” – more carefully, a de re description of Ralph’s belief-state - is (just) “of the thing”. That comes to saying that the description of Ralph’s belief-state does not attribute any mode of presentation (MOP) to Ralph. When we are talking about Ralph’s “de re belief”, “man in the trench coat”“ is the speaker’s way of picking out the topic[[11]](#footnote-12) of Ralph’s belief and may or may not be Ralph’s way. On the other hand, Ralph’s “de dicto belief” – more carefully, a de dicto description of Ralph’s belief-state - does, somehow, incorporate the information that “man in the trench coat” is the MOP operating in Ralph’s belief-state. Both descriptions give a description of Ralph’s overall belief-state. But the “de re” description does so in a less detailed, or coarse-grained, way, leaving out Ralph’s MOP, while the “de dicto” description does so in a more detailed, fine-grained way, including the information about Ralph’s MOP. [[12]](#footnote-13)

This analysis of the de re/de dicto distinction is corroborated by noting that the de re/de dicto distinction does not by itself correlate with any difference in behavior. Hold desires and beliefs constant and the de re/de dicto contrast makes no difference to action.[[13]](#footnote-14) Thus, as long as we hold constant Ralph’s belief that the man he thinks of as the man in the trench coat is a spy and that the man he thinks of as Orcott is not a spy no difference in action is to be expected whether that constant belief is described by the speaker de dicto, as that Ralph believes that the man in the trench coat is a spy, or de re, as that Ralph believes of Orcott, as the speaker is referring to him, that he is a spy. This contrasts with de se/non-de se beliefs that do make the crucial difference for certain actions. Our analysis in part 2 will enable us to lay out this contrast clearly.[[14]](#footnote-15)

We will use the term “dressed proposition” for the proposition-like object in play in a de dicto description of a belief-state. A dressed proposition will include not only the topic of the belief and the property attributed to the topic as well as (perhaps) a propositional structure, but also, and crucially, the agent’s MOP. It will be relevant below that we allow that the MOP in question might not be in any way linguistic.[[15]](#footnote-16) For example, consider an agent who experiences an unknown kind of flower purely visually and comes to the attitude that the visually experienced flower has a very special scent for which s/he also has no name. This agent then is in a belief-state that we clumsy theorists might describe as a belief with a propositional content of *that* flower having *that* scent. We can then more carefully describe the agent as being in a belief-state that involves the non-verbally but sensually apprehended MOPs. If put in terms of a proposition-like object of belief this will be in terms of a dressed proposition that involves the sensory MOPs that are operative in the agent’s belief-state.[[16]](#footnote-17)

That MOPs may include very general features of the way an agent thinks about a purported referent will be important in the following analysis where we will include structural features of the way the agent’s representations are related to one another, which structural features will engender functional features of the representations. To illustrate what we have in mind, suppose that an agent takes the thing s/he thinks of as named “Hesperus” to be identical to the thing that s/he thinks of as being named “Phosphorous.” Then this cognized identity provides a structural feature that can operate as part of the agent’s MOP for both Hesperus and Phosphorous.

We need to make two observations about our use of the term of art, “proposition”. First, we have adopted the terminology of dressed propositions; but since we allow that the relevant MOPs may be non-linguistic, the term “proposition” is here perhaps stretched. However, that being acknowledged, we take it that this extension should be unproblematic. Second, in conventional terminology propositions are described as the contents of propositional attitudes. In our formulation, this will always include the topic and the property attributed to the topic as part of this content. But when MOPs are relevant, should one count them as part of such content? In our exposition, we will speak this way, but we take nothing to turn on the question of whether MOPs are part of propositions properly so called or whether they might be included in the analysis in other ways.[[17]](#footnote-18) It is precisely for this reason that we have been using the hedge “proposition-*like*” and introduced the new term of art, “dressed proposition”. We take ourselves to have said enough to make it clear how MOPs operate in the overall analysis. More broadly, we hold that such issues about the exact use of terms such as “proposition” and “content” are ones, not of substance, but of elegance of theoretical formulation that might vary with the theoretical context.

Here, in preliminary sketch, is why we introduce our extended notion of dressed propositions. Perry (1979) rejected the possibility of understanding de se belief-states by analyzing belief-states in terms of beliefs with propositional or proposition-like objects. His argument was eliminative, arguing for each kind of propositional object then available that it would not supply the needed analysis. Our approach will be to apply our notion of dressed propositions to introduce an option not considered in Perry’s (1979).[[18]](#footnote-19)

As we noted in section 1.3 a thorough account must say not only why de se beliefs are necessary for action, but for which actions de se beliefs are necessary. So we must say something about how actions will be treated in the analysis. We take act types to be individuated, at least in large part, by objectives, i.e., by the goals of the acts. Perjury requires the intent to tell a falsehood; murder requires malicious intent.

Objectives, just like beliefs, can be analyzed in relatively fine- or coarse-grained ways, in terms of proposition-like objects, omitting or taking into account various nuances of the agent’s MOPs. Because objectives are specifiable to various degrees of grain, and because objectives help individuate act types, acts themselves are specifiable to various degrees of grain. Whether or not two objective types count as the same kind of type will turn on what is at issue in the context. This and other conclusions concerning objectives will straightforwardly carry over to the acts that they individuate. For these reasons, the de se phenomena that we marked using the symbols “!” and “¡” can appear for objectives - and hence for acts - just as they do for beliefs. As we illustrated in section 1.4: *John’s acting with the objective to make it the case that John! is rich* is a (prima facie) selfish act, while *John’s acting with the objective to make it the case that John¡ is rich* is a distinct, (prima facie) altruistic act. Similarly, Perry can have the objective of getting his! sugar bag patched, in which case he will lean down and fix it, or he can aim to get his¡ sugar bag fixed, in which case he will search around the store trying to find himself¡.

Our distinction between ! and ¡ objectives and acts will enable (partial) answers to the questions we raised earlier. For it is in having the same *de se* content that objectives and beliefs come into agreement. When we ask for which actions de se beliefs are necessary the answer is actions with appropriately related de se objectives. When we ask which beliefs are necessary for such de se action, the answer is the beliefs with the appropriately related de se status. We will see that the special character of such beliefs and actions is grounded in the special functionality of the cognitive architecture that we will now sketch.

**2. The Model.**

**2.1 What the model will and will not do.**

We will be constructing a simple model of the CP phenomenon. No claim is made that the model thoroughly captures the way !-states lead to action in humans. We take ourselves to be modeling the *relevant* functionality involved in CP phenomena, not all of the functionality of human belief. Our hope is to provide, by example, some understanding of the ultimate functionality involved in de se belief and action, though in a highly simplified form, and to show, were there any doubt, that there need be nothing mysterious or “extra-natural” about the Castañeda-Perry phenomena. For those who find the model too simple, future more complex models can be expected to address features not covered. So, for example, a model could be built that handles de se phenomena the way ours does, but, where ours handles just singular propositions, the extended model would also handle beliefs directed at propositions involving quantification.

The following notions will be presupposed: We assume an understanding of the idea of agents, assumed to have the capacity to deliberate and to act as a result of their deliberation. In particular, we will model action-producing deliberation, in part, using the idea of practical inference. We will assume normal, that is non-!-intentionality. So our aim is not to reduce intentionality per se, but to reduce de se intentionality to regular intentionality. What we have to say will be neutral about whether non-!-intentionality can or cannot be understood naturalistically, but will, we believe, show that the step from non-!-intentionality to !-intentionality can be done in an entirely naturalistic way.  Nothing about our account either assumes or sheds light on phenomenology.

**2.2 Preliminaries**: **File systems and de dicto beliefs.**

We follow others in modeling agents as realizing the functionality of concepts via the use of internal files.[[19]](#footnote-20) On our simple model, the agent sets up a file for each object encountered, where the files are labeled “John Perry”, “Man in the Trench coat”, “Orcutt”, “Man wearing a loud shirt”, etc.[[20]](#footnote-21)

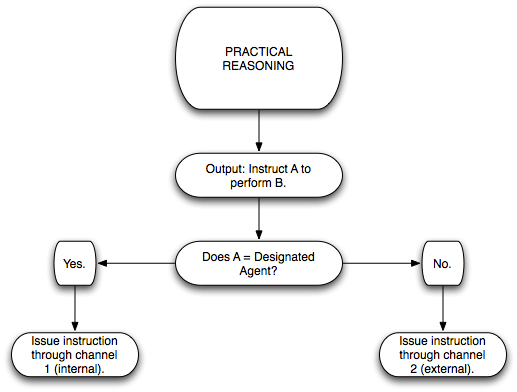
Each file has several sub-files, corresponding to the agent’s attitudes. For example, the file on Cicero has a belief sub-file, the information in which encodes the agent’s beliefs about Cicero. Properties represented on the belief sub-file - e.g. *was a great orator -* are thereby modeled as believed to be true of the file’s object - i.e. Cicero. In order that each belief have as content the corresponding dressed proposition, the content of the belief is determined by the fine-grained contents of the belief sub-file - e.g. if the agent has in the belief sub-file, known under the name “Cicero”, it is Cicero *so called* that the belief is about and being a great orator *so described* that is attributed to Cicero; it might not have any fine-grained content concerning Tully, so called; or concerning the skill for which Cicero was widely known, so described.

As well as a belief sub-file, each file also contains a sub-file pertaining to the agent‘s objectives that relate to the referent of the file. So, e.g., if the file on Cicero has *learn about* in the objectives sub-file, then the agent aims to learn about Cicero. If the agent has “known under the name ‘Cicero’ ” in the belief sub-file, it is Cicero *so-called* that the agent seeks to know more about. He may not have any fine-grained objective to learn about Tully, so called. So there is interaction between the belief and objective sub-files insofar as objectives are treated under representations. Quite generally, then, propositional objects for such objectives work analogously to those for beliefs. In particular, an objective is understood as a state of affairs represented by application of the predicative expression in the objective sub-file applied to the referent of the file as represented in the belief sub-file.[[21]](#footnote-22)

When context makes it amply clear we will often leave out the “sub” in talk of belief (sub)files and objective (sub)files. When the agent comes to believe that two files correspond to the same referent this is modeled by the files becoming linked understood in such a way that all the information on one counts as information on the other. That makes the two together function as one “super file”, which is exactly the functional effect of identity beliefs.[[22]](#footnote-23) It will be crucial to our account that, for example, an agent’s belief that Hesperus is identical with Phosphorus is modeled by the linking of the Hesperus and Phosphorus files.

As we indicated already, this simple model will not attempt to deal with many complications, e.g., quantified propositions, generics, the fact that, when the agent believes that John loves Mary, “loves Mary” should occur on the John file if and only if “is loved by John” occurs on the Mary file, and many others. Again we stress that the account is far from a general account of beliefs or belief-states. Our aim is to create a simple system, capable of exhibiting the functionality of de se belief, from which we can learn appropriate lessons. We can get the relevant insights by having the system handle simple singular propositions. We don’t need it to be as powerful as humans, or to handle every doxastic phenomenon.

**2.3 The relation** **Agent A2 is the Designated Agent for agentA1.**

In our simple model for each agent, A1, there is some agent, A2, such that A2 is the agent designated for agent A1 (A1’s *Designated Agent*) and each agent has a special file headed “Designated Agent”. *Designated Agent* is a functional notion, characterized as follows. Agents are modeled as deliberating using practical reasoning[[23]](#footnote-24), with conclusions of the form: Instruct agent A to perform action B.[[24]](#footnote-25) When an agent reaches such a conclusion it issues the corresponding instruction. The agent looks to whether or not it has described agent A as its Designated Agent, that is, whether the A-file is or is not linked to the Designated Agent file. If yes, the instruction is issued through channel 1; if not, then through channel 2. Channel 2 instructions are issued externally on a “public broadcasting system”. Channel 1 instructions are directed internally.[[25]](#footnote-26) The agent responds to internally directed instructions by implementing the instruction. In other words, internally routed directives are obeyed, externally routed directives are passed on.[[26]](#footnote-27)

**2.4 New kind of MOP.**

Recall that in section 1.5 we included MOPs’ structural, and more generally functional, relations among representations, now more specifically, file headings. So, for example, Hesperus and Phosphorus are functionally treated as distinct if the agent keeps distinct unlinked files on them and are treated as identical when their files are linked. The Designated Agent file has a distinctive functional relation to the action-triggering structure of the larger cognitive architecture. Consequently, this functional structure can operate as what is now a distinctive MOP for the Designated Agent file, its file heading, and all the files to which these are linked.

To illustrate with our running example, when on the Designated Agent belief sub-file there is recorded “Has won the lottery” or “Is making a mess”, it follows that the agent has beliefs, respectively, to the effect that the Designated Agent has won the lottery, or that the Designated Agent is making the mess, where these beliefs utilize the special mode of presentation. The agent, in effect, believes that the specially represented agent has won the lottery or has made the mess. So, corresponding to the special kind of MOP that comes from having a special kind of file, there is a special kind of belief with such dressed propositions as their propositional object. Or, more carefully, there is a kind of fine-grained model of belief-states that treats beliefs as having such dressed propositions as propositional objects.

This last more careful statement is essential for understanding the relation of “de re”, “de dicto” and “de se” beliefs. This is because, when more carefully stated, as we explained in section 1.5, the de re/de dicto distinction is not a difference in kind of belief but a difference in kind of description of one and the same belief-state. By way of contrast, de se belief-states really are a different, distinctive *kind* of belief-state. The beliefs states, described de dicto, that the man in the loud shirt is the messy shopper and that Mr. Perry is the messy shopper are different belief-states because they involve different MOPs in using different files with different file headings. But while different beliefs, with respect to our questions about triggering action, these two beliefs are the same kind of belief. By way of contrast, the belief-state in which “is the messy shopper” is written on the agent’s Designated Agent file involves a functionally distinctive kind of MOP from either of the prior cases, and so represents a different kind of belief-state from either of those cases. Once this is all clear, we can revert to the less careful but more familiar way of speaking and say that we have three different beliefs, all described “de dicto”[[27]](#footnote-28), that the man in the loud shirt is the messy shopper, that Mr. Perry is the messy shopper, and that the Designated Agent is the messy shopper, with the third a genuinely different kind of belief from the first two. All this with the understanding that the man-in-the-loud-shirt-file, and the Mr.-Perry-file have not been linked to the Designated-agent-file. In this circumstance, only the third constitutes the kind of distinctive belief we call “de se”. If the first two files become linked to the Designated-Agent-file, then the first two beliefs also become de se.[[28]](#footnote-29)

**2.5. Identity of an agent and the agent designated for that agent.**

The agent designated for an agent is, of course, that very agent. Viz, if A2 is the Designated Agent for A1, A2 = A1. All identities are necessary, but their statements need not be, as in the case of a statement identifying the evening with the morning star. In the case of the de se the informative nature of the identity statement is provided by the functionality of the file for the Designated Agent. We could have called it the “Me” file. But by postponing making the identity explicit, we have explicated the functionality in a neutral way, without assuming any notion of the de se up front. We have not begged relevant questions in the way the label “Me” might have suggested.

Our model provides a natural interpretation for the “!” notation. Appending a “!” to a term signals that the file picked out by the term, or its anaphoric antecedent, has been linked to the Designated Agent file. This, of course, means that the referent is represented as being identical with the Designated Agent. It is essential, however, that it is not just this identity that is represented, but that the identity is represented by the linking of these two files. The special functionality of the Designated Agent file is then also crucial for the identity, so represented, to play the needed role in the analysis [[29]](#footnote-30)

Though the “cat is out of the bag”, we will continue to use the artifice of the Designated Agent file. In spelling out the functionality of the Designated Agent file we are spelling out certain important parts of the concept of the self. But only parts. We want to be very clear that we have no pretensions to anything like a complete analysis. By building up a functional account of a clearly circumscribed notion we facilitate future discussion of the question, what parts of the conception of the self are covered by the limited idea of the Designated Agent and what parts are left out. By sticking with the term “Designated Agent” rather than using “self”, ‘”I”, “me”, etc., we keep this intent clearly in mind. In section 3 we will further develop this functionality with attention to the ways in which an agent determines which agent is designated for him/her. We will also mention the essential afferent – epistemic – considerations that must be included with the efferent – action-triggering – considerations.

**3. Applying and expanding the model.**

**3.1 How change of belief leads to action.**

When Perry learns that there is someone with a leaky sugar bag he sets up a new file concerning this “Messy Shopper”. As Perry acquires data about the Messy Shopper he puts the information on the Messy Shopper belief file, writing things like:

Has a leaky sugar bag on the bottom of his cart.

Is making a mess.

Is wearing a loud shirt.

Is named “Mr. Perry”.

Etc.

Perry also forms an objective: Get the Messy Shopper to patch her!/his![[30]](#footnote-31) leaky sugar bag. That is, Perry puts “Patch her!/his! leaky sugar bag” on the Messy Shopper objective (sub) file.

None of this, of course, will yet trigger the needed act. Perry, at this point, is not appreciating that he! is the Messy Shopper.[[31]](#footnote-32)

Recall that a believed identity for the referent of two terms is modeled by linking the files with those terms as file headings. Thus Perry’s coming to appreciate that he! is the Messy Shopper is modeled by his linking the Messy Shopper file to his Designated Agent file. The link models Perry’s belief changing from:

(a) The Messy Shopper¡ is making a mess.

to:

(b) The Messy Shopper! is making a mess.

The *very same* (token identical) act of linking *also* models Perry’s change of objective from:

(c) Get the Messy Shopper¡ to patch her!/his! sugar bag.[[32]](#footnote-33)

to:

(d) Get the Messy Shopper! to patch her!/his! sugar bag.

The one link thus models *both* the change of belief and the change of objective from ¡ to !, altering the character of both the belief and the action-individuating objective. It follows, then, that in replacing the belief (a) with the belief (b) an agent with objective (c) necessarily goes to having objective (d). And because !-marked instructions are directed internally, the instruction (d) is directed in this way, that is, Perry’s sugar bag patching behavior is triggered.[[33]](#footnote-34)

In section 1.4 we parsed the CP problem into the problems of understanding both the *explanandum* (what special kind of action) and the *explanans* (what special kind of belief) and promised to show how these are intimately linked. Mission accomplished: There is both a belief and an objective involved in the case, and both are shifted from ¡ to ! by the one linking of files.

In Perry (1979), much subsequent literature, and preanalytically one naturally takes change in belief to be the cause, and so the explanation, of action. On our account this is, at best, an oversimplification. External circumstances cause a change in belief. In the cases of identification here under consideration we model this change of belief with the linking of two files. But formation of this link equally constitutes the change of objective, and in the end it is objectives, when transformed into instructions, that most immediately lead to action. If we identify the cause of action as a change it should be the whole package comprised by setting up the link together with the change of belief and change of objective that arise with this link. Given the “wiring” described by our model, the three, change in link, in belief, and in objective, are architecturally tied together. For an alternative description of the etiology we can distinguish between the new belief and the new objective that are both products of this change. If we want then to think of the causal pattern in terms of these end states of the process, we naturally focus on the new belief because this is the way we think of the immediate response to external input. But it is the new objective that provides the proximal cause of action. We should not, however, given the way in which they arise, think of the new belief as a cause of the new objective.

Might this gloss on the relation between new belief and new objective in the triggering of action be an artifact of the model? We don’t think so. For the moment putting the model aside, think of one’s own case. If I learn that I! am the messy shopper, what I have learned also transforms my objective from getting someone else to patch their sugar bag to the objective that I should do the patching myself. It is then in light of this new objective that I stop and do something about the mess that I am making.

**3.2 An illustration.**

To consolidate these considerations we illustrate how identification as the Designated Agent results in action. We model the agent as engaging in the following practical reasoning, where we have annotated the steps in the argument with consequent changes in the file system.

1) Someone is making a mess with a leaky sugar bag.  
 (New information: The agent sets up a Messy Shopper file.)

2) The person making the mess should patch her!/his! sugar bag.  
 (The agent writes “Patch her!/his! leaky sugar bag” on the Messy Shopper objective (sub)file.)[[34]](#footnote-35)

3) The man in the loud shirt seen in the mirror is making the mess.

(New information: A file is set up for “the man in the loud shirt seen in the mirror”, and this file is linked to the Messy Shopper file.)

4) Therefore the man in the loud shirt should patch his! sugar bag.

(The link effectively puts “Patch his! sugar bag” on the objective (sub)file for the man in the loud shirt.)

5) The man in the loud shirt is the Designated Agent.

(New information (discussed below): File for “the man in the loud shirt” is linked to the Designated Agent file.)

6) Therefore the Designated Agent should patch his! sugar bag.

(Effectively, “Patch his! sugar bag” is now on the Designated Agent objective (sub)file.

7) Conclusion of practical reasoning: Issue the instruction, PATCH YOUR! SUGAR BAG to the Designated Agent.[[35]](#footnote-36)

with the result that the instruction is directed internally. This is the simplest of examples. But it should be enough to indicate to the reader how linking to the Designated Agent file can enter in all kinds of ways in the process of deliberation. Such linking constitutes, in our model, the effect of de se considerations, effects that can be direct or indirect, even highly so.

**3.3 Expanding the functionality of the Designated Agent file.**

How does an agent get the kind of information used in step 5 in the above illustration?[[36]](#footnote-37) This is a practical problem, resolved in different ways in different contexts. The agent may simply query the Designated Agent file, for example, looking to see whether “is named “Perry” is on the file. Or our agent may correlate action with observation. For example, to see whether the man with the loud shirt seen in the mirror is the Designated Agent, the agent could issue the internal instruction: WAVE ARM. If the man with the loud shirt seen in the mirror is then seen waving his arm, the file for The Man in the Loud Shirt Seen in the Mirror is linked to the Designated Agent file. This provides an example of the ways in which the Designated Agent file can incorporate the functionality that reflects the identity of the agent designated for an agent with that very agent.

While we have provided only one very simple example of how an agent might establish such identities, the example makes clear how a functional system can incorporate a multitude of such stratagems each of which enriches the functionality of the Designated Agent file. In section 2 we said that the Designated Agent file was characterized functionally. But there the only thing said about this functionality was that instructions directed to the Designated Agent were issued internally – call this the “basic action-function” of the Designated Agent file. The function of the Designated Agent file is now further characterized by the way links between the Designated Agent file and other files represent believed identities and how such links are set up. This allows the Designated Agent file’s basic action-function to tie more, less, and sometimes very indirectly with any consideration that might come up in the kind of practical reasoning illustrated above. This tie now embeds the Designated Agent file’s basic action-function within cognitive functioning very broadly, as is characteristic generally of functional characterizations of mental states.

At the end of section 2 one might have plausibly raised the following objection: We have said no more about the functionality of the Designated Agent file than that instructions directed to the Designated Agent are directed internally. How is that more substantive than just saying that de se beliefs are the ones that trigger action, just our complaint about Lewis’s analysis?  
  
 Our analysis provides substantive functionality in a number of respects and provides a foundation for much further articulation. While we leave it to roboticists and cognitive psychologists to consider how the internal direction might be implemented, we have described how, at least on a simple model, the node from which instructions are directed internally is connected with other elements of action-directing cognitive structure. In particular we have outlined the structure of the contrasting kinds of objectives and acts, de se and non-de se, and the roles they play in the larger functional structure in the interplay of de se and non-de se beliefs; and how all of these lead to the point at which instructions are issued, internally or externally. In our simple model the functionality inherits a great deal of the functionality of deliberation characterized as practical reasoning; and we can expect that analogous functionality will apply again in more subtle characterizations of the processing of information as it effects determination of actions. Indeed, as such characterizations become more realistic, the richer structure will further enrich the functional characterization of the role of the Designated Agent.

We can expect further enrichment also when one adds to such models characterization of how information flows into the system, especially through perception. For example, in what would still be an extremely primitive model, we can think of characterizing the agent as having perceptual apparatus the output of which is processed by an interpretive system to produce predicative expressions specifying the relation of external objects to the Designated Agent. Such a perception and interpretive component could write predicative expressions such as “there is a large round red object five feet away from\_” and “The temperature is rising on the surface of \_” on the Designated Agent file. This kind of epistemic function of the Designated Agent file enormously expands the file’s functionality.[[37]](#footnote-38) This is not the place to begin work on such considerations. The point here is that accounts with or with variations on the kind of structure we have outlined have the potential for developing an extraordinarily rich functional characterization for the Designated Agent role.

**4. Prospects for broader application.**

The style of analysis that we have proposed promises to apply very broadly: Reflexive perception of people: Mary saw herself!/¡. Action verbs generally: John tickled, praised, chased himself!/¡… Verbs that can take propositional objects that occur in other constructions: remembered/liked/anticipated… herself!/¡, remembered, liked, anticipated [PRO!] going home…. Generally, the kinds of considerations we have examined potentially bear on all verbs marked for subject as experiencer or as agent.

Given the identity of an agent’s Designated Agent with that very agent, together with the functional characterization, our style of analysis also promises to have implications for a functional individuation of agents. We have tacitly been assuming that the “decision” and “instruction following” modules are housed in the same body so that the full agent is naturally seen as located in this one body. But our characterization of the source of agency is functional and so could be realized in scattered units.[[38]](#footnote-39) Our style of analysis also makes possible giving content to forms of extreme pathology such as having multiple Designated Agent files.

Lastly, if we're right that the Designated Agent file must be characterized functionally then this leads to the observation that de se content can only be encoded in a functionally-understood body-related format. It would thus stand as the kind of content involved in embodied cognition in the sense outlined by Goldman (2012). Combined with Lewis‘s hypothesis that all content is reducible to the de se, this leads to the radical hypothesis that all content is ultimately reducible to embodied content.

Appendix

Though we differ from the analysis of Perry and Recanati in a great many details, on the whole we take our development to complement rather than to conflict with theirs. Here we sketch what we take to be the relation between their analyses and ours.

Perry presents the view in (1998, 2011, 2012). Here we will follow (2011), which we find clearest.

Perry introduces his “ideas”, “notions” and “files”:

I’ll call just call ideas of the sort we associate with general terms, verbs, prepositions, and the like “ideas”. I’ll call ideas associated with particular things --- people, places, things, universities,… “notions”. A notion, plus all of the ideas associated with it in belief, makes up a *file*. (2011, p. 386)

The account introduces a special notion, the “self-notion”, “with which all of the information picked up in self-informative ways [feeling hungry, proprioception…] is associated, and the beliefs associated with this notion motivate self-sensitive actions.” (2011, p 387) Into the self-file the agent also collects “all sorts of additional information, about our telephone numbers, salaries….” (2011, p. 387) Such externally collected information is obtained in the same kinds of ways that one obtains information about others. When the agent takes the agent in question to be her/himself the information goes into the self-file. If the information is about the agent her/himself but thought to be about someone else, the information goes into files for others – these are the cases like that of the messy shopper and Kaplan failing to recognize himself in his reflection. The problematic cases provide mere “knowledge of who one happens to be” (2011, pp. 373, 378, 383 ff.).

The self is just the agent to whom the self-file belongs, and what is special about the self-file is the way it is connected with the larger cognitive architecture:

*Self* is a role persons play. One’s self is like today or one’s home or one’s father; a perfectly ordinary object, thought of in virtue of its relation to the thinking agent. The relation associated with “self” is just identity. That is,

If *x* is *y’s* self, then *x = y*

There are self-based methods of inquiry, and self-sensitive ways of accomplishing things … [that are] architecturally and contingently guaranteed to be self-informative [and relevantly connected with action] (2011, p. 280 - 381).

We emphasize that this statement of identity is made by the theorist, not made or thought by the agent.

In all four papers Perry mentions the connection with reflexive action along with reflexive ways of knowing. This is put more specifically in (1998 , 11) where Perry writes that there are “reflexive ways of acting. These are ways of bringing it about that someone has a property, that each person can use to bring it about that he or she has a property, but cannot use to bring it about that others have it.” These correspond to our stylized “internally directed instructions”.

We were misled by the exposition in (1998):

The view I advocate is simply that identity is a basic relation, and that our idea of self (“being me”) is the idea of the agent-relative role, is identical. This is the role we each play in our own lives. That is, identity, like being in front or behind or above, is a basic relation relative to which we have epistemic and pragmatic methods. There are certain methods for picking up information about the person identical with us, and certain methods for having an effect on that person. The notion that is the repository of information gained via those methods, and the motivator of actions associated with that relation, is our self-notion. The person this notion is of, is the person we take ourselves to be. (1998, p. 95)

Perry addresses a potential problem with the foregoing, that it might appear to appeal to the notion of *myself*, which is what is to be explained. He addresses this concern by appealing to his notion of an unarticulated consistent. Earlier (1998, pp. 86 ff.) Perry had argued that, e.g., in the case of perception, one does not have to think something like *I perceive an apple*. The first term in such a relation can be an “unarticulated consistent”, so that the thought takes something like the form: *apple there*. His strategy appears to be that rather than thinking, *I am identical to [the referent of my self-notion]*, the first term of the identity relation can be left as an unarticulated constituent:

It is agent-relative knowledge that keeps the account from circularity. It is not the idea, person identical with me that I need, but only the role-idea, person identical. My idea of me is not a part of this idea. (1998, p. 95)

But things can’t work this way because when the self-notion is introduced in this way, we ARE introducing the self-notion, so the self is (conceptually) no longer an unarticulated constituent. Or again: since it is identity that is in question, if it is on one side of the relation it’s on both. This whole way of introducing the self-notion is, anyway, we feel, misleading because the role relation, identity, is itself very different from role relations such as perception.

We address the apparent circularity by using the neutral terms “designate agent” and “file for the designated agent” rather than “self” and “self-file”. The neutral terminology enables us to state clearly that it is the cognitive architecture that makes this file more than an idle floater: it is the cognitive architecture that gives this file its functional substance. Without any such architecture, the identity between the agent and referent of the self-file would be, as we said in the case of Lewis, barren. It is then the theorist, not the agent that makes the identification. Perry clearly agrees with this order of explanation when he writes that “self-notions are about their possessors *because* of their causal and informational role” [our emphasis] (2012, 99).

In the end, the difference in this respect is one of exposition, not substance. But the use of the neutral terminology makes it easy to see that, while the referent of the designated agent/self-file is the very agent whose file it is, it is the architecture, not this identity that is doing the work.

Turning to Recanati, he works from the background of what he calls, “epistemically rewarding relations” with the relation of perceiving being the leading example. These correspond to Perry’s relations of roles that things play in an agent’s life through, for example, perception. When the agent perceives a man dancing the relation, *the agent perceiving the man*, underwrites a file whose referent is the man perceived. In Recanati’s hands, the relation is no part of the content of the agent’s thought, *that is a man dancing*, but the relation is operative in the truth conditions for the thought – the thought is true just in case there is a man that is the object of the perceptual event in question and that man is dancing. (This corresponds nicely to Perry’s distinction between conditions of truth and truth conditions mentioned above in footnote 16.) The relation of perception is epistemically rewarding in the sense that it provides a channel through which the agent gathers relevant information.

In Recanati’s presentation identity is just one such epistemically rewarding relation:

According to John Perry, the concept of self is a mental file that is based upon a special relation which every individual bears to himself or herself, namely identity. In virtue of *being* a certain individual, I am in a position to gain information concerning that individual in all sorts of ways in which I can gain information about no one else, e.g. through proprioception and kinesthesis. The mental file “self” serves as repository for information gained in this way. (2009, p. 255)[[39]](#footnote-40)

The wording here initially misled us, as it did in the case of Perry’s (1998) to which Recanati is apparently referring, to think that it is supposed to be the relation of identity in virtue of which the special, personal epistemically rewarding relations hold. Whatever Recanati might have intended here, our view is that the referent of the self-file IS (that is, is identical to) the agent whose file it is; but this fact, noted by the theorist, not the agent, is the presuposition that must hold for the “self-based methods of inquiry” and “personal epistemically rewarding relations” to hold in the complex structure of the cognitive architecture. That is, given the identity, it is the cognitive architecture in virtue of which information obtained by perception, proprioception, and the like is delivered to the designated agent/self-file and in virtue of which the “reflective modes of action” can take place.

One more difference between us and the Perry/Recanati development is the focus. Although Perry developed the issue in his (1979) with reference to self-directed action, in all his later positive work on the issue he focuses on “self-based methods of inquiry” with “reflexive ways of acting” mentioned only in passing. Recanati likewise focuses on “epistemically rewarding relations”. We focus on the connections with action, mention the connections with perception and self-awareness only in passing. Clearly both must be addressed.

References

Hector-Neri Castañeda (1968). “On the logic of attributions of self-knowledge to others” *Journal of Philosophy*, 65: 439-456.

Daniel C. Dennett (1983). “Beyond Belief.” In Andrew Woodfield (ed.), *Thought and Object*. Oxford University Press.

Heim, Irene (1988). *The semantics of definite and indefinite noun phrases*. New York: Garland Pub.

Holton, Richard (to appear). “Primitive Self-Ascription: Lewis on the De Se” in Barry Loewer and Jonathon Shaffer, eds., *The Blackwell Companion to David Lewis*. Blackwell.

Liao, S. (2012). “What Are Centered Worlds?”. *The Philosophical Quarterly* 62: 294–316.

Ludlow, Peter (2000). “Interpreted Logical Forms, Belief Attribution, and the Dynamic Lexicon.” In K.M. Jaszczolt, ed., *The Pragmatics of Propositional Attitude Reports*,  
Oxford: Elsevier Science, 2000, pp.31-42.

Mach, Ernst (1914). *The Analysis of Sensation*, p. 4, n1.

Ninan, Dilip (2010). “De Se Attitudes: Ascription and Communication” *Philosophy Compass*.

Perry, John (1979). The Problem of the Essential Indexical. *Nous* 13(1):3-21.

Perry, John (1998) “Myself and I.” In Marcelo Stamm, editor, *Philosophie in Synthetisher Absicht* (A festschrift for Dieter Heinrich), Stuttgart: Klett-Cotta, pp. 83-103.

Perry, John (2010) “Selves and Self-Concepts” in *Time and Identity,* eds. J.K. Campbell, M. O’Rourke, and H.S. Sliverstein. Cambridge: MIT Press. pp. 229-247.

Perry, John (2011). “On Knowing One’s Self”, in *The Oxford Handbook of the Self*, ed. S. Gallagher, Oxford University Press, 2011. Pp. 372-393

Perry, John (2012) “Thinking About the Self” in *Consciousness and the Self: New Essays*, L. JeeLoo and J. Perry, eds. Cambridge: Cambridge University Press

Recanati, François (2009). “De re and De se”. *Dialectica* 63 :3 pp. 249-269.  
  
Recanati, François (2012). *Mental Files*. Oxford University Press. Oxford.

Salmon, Nathan (1986). *Frege’s Puzzle.* Ridgeview press. Atascadero, CA.

1. This kind of problem goes back to Mach (Mach, 1914, p. 4, n1). [↑](#footnote-ref-2)
2. These names, in taking the de se role, do not thereby act like indexicals. For they do not have their referent determined by speaker context. E.g. it is still the case that Tarzan understands “Tarzan” in the mouths of others to refer to him. [↑](#footnote-ref-3)
3. This is clearer using the symbols we introduce later in the paper. Consider the corresponding third-person claim: “He! hoped that he¡ would return to his¡ life of philosophy.” To make this claim about one’s former self, one must change “he” to “I” and “his” to “my”. But, if this is genuinely to be the same claim made about one’s former self, the “!” and “¡” assignments had better not change. [↑](#footnote-ref-4)
4. You can build similar examples in which the agent speaks about his future self using the term “I” in a non-de se way. It is interesting, however, that you can’t get the phenomenon to appear in the present tense, though we suspect that this is a pragmatic accident (in the sense of Moore’s pragmatic contradictions and tautologies). Just as Lewis can utter that “Castañeda falsely believes that it is raining” whereas Castañeda cannot (honestly, coherently) assert “I falsely believe that it is raining”, so it is that Lewis can claim that “Castañeda hopes de rethat he is alive and well, but he does not, indeed could not, hope it de se”; while Castañeda cannot (honestly, coherently) assert “I hope de rethat I am alive and well, but I do not hope it de se”. In such cases, if Lewis”‘s assertion is a true one, then Castañeda really does hope it de re but not de se; but Castañeda is pragmatically excluded from asserting it. The difference turns on the fact that when Lewis speaks, the speaker is distinct from the person spoken about, while when Castañeda expresses the same thought, the speaker and person spoken about coincide. [↑](#footnote-ref-5)
5. Our use of “!”corresponds to Casteneda’s “\*” (1968, p. 441). We prefer the “!” as it supports the contrasting “¡” of which we will make extensive use. [↑](#footnote-ref-6)
6. We have expressed this point with the neutral term “understanding”; but do we here have a linguistic (that is, syntactic or semantic) ambiguity, or some looser, pragmatic, distinction? There is some evidence that the ambiguity is linguistic, since the difference in readings seems to be grammatically encoded in certain sentence types:

   1. John wants to leave the party.

   ...can only by understood in the ! sense - i.e. as meaning that John wants himself! to leave the party. This contrasts with “John wants himself to leave the party,” which, just like cases in which “himself” is within a that-clause, admits of a ¡ reading. For example, we can imagine John wanting the tallest man to leave the party, so that he will be the tallest, without knowing that in fact he is already the tallest man at the party. Again we imagine a perfectly knowledgeable Lewis on the sidelines truly saying “John(unknowingly) wants himself to leave the party.” Clearly Lewis means “himself¡.” But there is no way Lewis could utter (1) to the same effect. In (1), the de se reading appears to be grammatically forced. This effect occurs generally with both infinitival and gerundive, so-called “PRO” constructions.

   For present purposes we do not need to decide the issue of whether all statements expressing the de se understanding involve, as in the case of PRO constructions, some strict syntactic or semantic contrast in reading. So henceforth we will use “reading” loosely and not worry about whether various readings (in our loose usage) correspond to variation in the linguistic features of the sentence or to some wider contrast. [↑](#footnote-ref-7)
7. Lewis takes properties to be sets of possibilia. Nothing we say here will turn on details of how one understands properties. [↑](#footnote-ref-8)
8. We will also be appealing to identity, as do both Recanati and Perry. We will discuss the relation of the roles that identity plays in these other approaches. [↑](#footnote-ref-9)
9. In (1998, p. 97) Perry characterizes the distinctive aspect of action as “reflexive action”. [↑](#footnote-ref-10)
10. Accounts that are more or less fine/course-grained can be expected to form a partial, not a linear, ordering. [↑](#footnote-ref-11)
11. Since we are using “object” when working in terms of propositions or proposition-like objects of belief, we will instead use “topic” to talk about the referent of the subject term in the that-clause used in a belief-state description. We note, as pointed out by Dennett (1983, p. 77), that all the issues surrounding any de re/de dicto distinction apply just as much to the predicate position: John may believe that Mary’s house is the same color as Peter’s without appreciating that that color is white, thereby believing de re, but not de dicto, that both houses are white. [↑](#footnote-ref-12)
12. This way of understanding the de re/de dicto distinction also explains the usual tests using a scope distinction and whether one can or cannot quantify into the belief context. In ⌜x is such that A believes that x is P⌝ the first occurrence of ⌜x⌝ provides the speaker’s, not A’s MOP. This then constitutes a de re description of A’s belief-state, while ⌜A believes that x is P⌝ can be understood as providing either the speakers or A’s MOP for ⌜x⌝. Similarly, in ⌜There is an x such that A believes that x is P⌝ the speaker is asserting the existence of an x, while ⌜A believes that there is an x that is P⌝ can be understood as either the speaker or A asserting/believing that there is an x that is P. [↑](#footnote-ref-13)
13. The speaker’s MOP may make a difference to the action *the speaker takes the agent to have performed,* and in this sense the de re/ de dicto distinction “can make a difference to action.” [↑](#footnote-ref-14)
14. In many places Recanati draws the same conclusion about the de re/de dicto distinction, e.g., (2012, p. 185 ff.). See also Ludlow (2000, section 2). [↑](#footnote-ref-15)
15. Compare Lewis (1979, p. 538): “I might ascribe espionage to a stranger under a description given by a visual image of his face.” [↑](#footnote-ref-16)
16. Less the extra-linguistic MOPs, dressed propositions are exactly what Perry (1979) calls (p. 6) propositions, and later (p. 17) de dicto propositions. In his (2011, pp. 389, 390, 392) Perry approaches this issue by making a distinction between conditions of truth and truth conditions. [↑](#footnote-ref-17)
17. Since propositional attitudes are action-guiding and MOPs figure in how action is guided, the MOPs must be included in the analysis of proposition attitudes one way or another. In his (1986, pp. 111ff.) Salmon explores how such an analysis might be set up characterizing the attitudes with a three-place relation, holding between an agent, a proposition understood as in the Russelian tradition, and the third place carrying information about the agent’s MOPs. [↑](#footnote-ref-18)
18. As we explain in the appendix, in more recent publications Perry develops a view very much like the one presented here. [↑](#footnote-ref-19)
19. See Recanati (2012, p. vii) for a history of this idea. [↑](#footnote-ref-20)
20. Are the labels identifying files part of the content or information on the file? Recanati (2009, pp. 262 ff.) argues not, for files set up on the strength of relations of acquaintance, such as perception. We would think that for files set up on the strength of definite descriptions, such as a file for ‘the strongest man in the world”, the answer would be yes. This issue is not ultimately relevant for the present discussion, but we will be careful to make our exposition consistent with Recanati’s views on this issue. [↑](#footnote-ref-21)
21. In what follows we will streamline exposition by discussing only cases in which the file heading is also in the belief sub-file, as a representation under which the agent refers to the referent of the file, and not mention this explicitly. [↑](#footnote-ref-22)
22. Perry puts this idea in terms of merging files (2010, p. 241). [↑](#footnote-ref-23)
23. Perry similarly on (2010, p 243). [↑](#footnote-ref-24)
24. Detailed formulation of practical reasoning will not be relevant to our analysis. All that matters is that such reasoning transforms some beliefs and objectives into new objectives. In particular it will not be relevant to our analysis just how it is determined when an objective is transformed into an instruction. These shifts will depend on considerations such as cost-benefit estimates, whether “other things being equal” conditions have been met, and the like. Such considerations are not relevant to present concerns and so are left out of the model. [↑](#footnote-ref-25)
25. Perry similarly (2010, p. 244). [↑](#footnote-ref-26)
26. The action-generating “module” might reissue the instruction externally – asking someone else to carry out the action for the agent. [↑](#footnote-ref-27)
27. This on the assumption that the MOP “Designated Agent” counts as linguistic. If not, the third is not, strictly speaking, described de dicto, but is treated theoretically in a way that has important parallels with de dicto descriptions as traditionally understood. This was the point of including non-linguistic MOPs in section 1.5. [↑](#footnote-ref-28)
28. When all three files become linked, are there then three beliefs or only one belief? We hold that this provides a case in which Perry’s advice (1979, p. 18 ff.) to consider belief-states rather than beliefs is particularly germane, a case in which either answer threatens to mislead. One is better off describing the full relevant structure of the belief-state in terms of linked files. When the three files are linked the de re/de dicto/de se contrast also becomes muddied because all three MOPs are the agent’s MOPs for (what the agent takes to be) the same thing while, to succeed in describing the belief as de se, the speaker must communicate what amounts to one of these linked files being the Designated Agent file. [↑](#footnote-ref-29)
29. It is a further project to work out in detail how, and to what extent, what we are representing with the “!”, “¡” notation is marked and organized in the semantics, in particular its relation to PRO constructions. There may be much more to say here than is to be found in the current literature. [↑](#footnote-ref-30)
30. The “!” is essential here: Get The Messy Shopper to patch her¡/his¡ sugar bag is a different objective, corresponding to the distinction in kinds of acts that we have explained above. [↑](#footnote-ref-31)
31. This is consistent with the last footnote. Perry can want the Messy Shopper to patch his! own sugar bag, in the knowledge that it is his! (the Messy Shopper’s) sugar bag, while still not appreciating that he! (Perry) is the Messy Shopper. [↑](#footnote-ref-32)
32. Note that, pursuant to the last two footnotes, we need the “!” on “her/his” despite the “¡” on “Messy Shopper”: The speaker, Perry, wants the person known to him only as “the Messy Shopper” to patch the sugar bag that, Perry assumes, that person, the Messy Shopper, appreciates, or will appreciate, is her!/his! sugar bag. In terms of the modeling, the “!” on the “her/his” is to be understood as signaling a link, not to Perry’s own Designated Agent file, but a link that Perry takes to apply in his model of the Messy Shopper , to the modeled Messy Shopper’s Designated Agent file. A reading with “¡” on “her/his” is possible, but it would be a contorted story that would make that appropriate. [↑](#footnote-ref-33)
33. For the slide in the exposition here from “objective” to “instruction” see note 24 above. [↑](#footnote-ref-34)
34. See note 30. [↑](#footnote-ref-35)
35. Again, for the slide from “objective” to “instruction”, see note 24. [↑](#footnote-ref-36)
36. One might have formulated the question as, “How does an agent learn *who* is the agent designated for him/her?” But since the agent designated for an agent *is* that very agent, the question could be heard as “Who am I?” As explained in the last paragraph of section 2.5, we want to keep it clear that we are developing a functional notion that is supposed to be part, but only part, of the conception of self, so we want to stay clear of questions like that! The solution is to give the question a clearly delimited formulation: Each agent has a list of agents, one corresponding to each card that is marked as agentile. The clearly stated, delimited question then is: which agent on that list is the agent designated for the agent in question? [↑](#footnote-ref-37)
37. In the articles discussed in the appendix, Perry and Recanati frequently mention this epistemic functionality. Perry calls channels such as perception and introspection “normally self-informative methods”. Recanati calls such channels “epistemically rewarding relations”, giving perception, proprioception and kinaesthesis as examples. [↑](#footnote-ref-38)
38. Similarly considering how this way of approaching the idea of the self might be radically stretched, Perry considers the possibility that “my visual centers have been wired to someone else’s eyes” (Thinking about the self, p. 23). See also (2010, pp. 233-234). [↑](#footnote-ref-39)
39. On page 261 this is put: Where Rm is the general case of a relation on which a file may be based, Rego is the special case of the relation on which the self, or EGO, file is based, where Rego is the relation of identity . [↑](#footnote-ref-40)