



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

How Do You Know That You Settled a Question?

Citation for published version:

Vierkant, T 2015, 'How Do You Know That You Settled a Question?', *Philosophical explorations*, vol. 18, no. 2, pp. 199-211. <https://doi.org/10.1080/13869795.2015.1032330>

Digital Object Identifier (DOI):

[10.1080/13869795.2015.1032330](https://doi.org/10.1080/13869795.2015.1032330)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Philosophical explorations

Publisher Rights Statement:

©Vierkant, T. (2015). How Do You Know That You Settled a Question?. *Philosophical explorations*. / This is an Accepted Manuscript of an article published by Taylor & Francis Group in *Africa Review* on 17/04/2014, available online: <http://www.tandfonline.com/10.1080/13869795.2015.1032330>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



How Do You Know That You Settled a Question?

Tillmann Vierkant

School of Philosophy, Psychology and Language Sciences

Dugald Stewart Building

3 Charles Street

Edinburgh¹

Abstract:

It is commonly assumed in the philosophical literature that in order to acquire an intention the agent has to settle a question of what to do in practical deliberation.

Peter Carruthers (2007) has recently used this to argue that the acquisition of intentions can never be conscious even in cases where the agent asserts having the intention in inner speech. Because of that Carruthers also believes that knowledge of intentions even in first person cases is observational. This paper explores the challenge Carruthers' argument throws up for accounts that also rely on the notion of settling a question for intention acquisition, but who also want to maintain at the same time that knowledge of intentions in the first person case is not observational.

Keywords:

Intentions, self knowledge, dual systems, Carruthers, self other gap

¹ Email: T.vierkant@ed.ac.uk

It seems obvious that I know about my own intentionsⁱ in a different way than I know about your intentions. When I want to know your intentions it seems I need to observe your behaviour and then infer the invisible mental state that explains it.ⁱⁱ This does not seem to be something I have to do in order to know about my own intentions. In my own case I simply make up my mind what I am going to do and once I have done that I have acquired the intention in question and at the same time know that I have it. There is no need for any observation of behaviour or any inference. This asymmetry is pre-theoretically very intuitive and it is one of the main reasons why many scientists and even more philosophers are sceptical about the so called theory theory in the mindreadingⁱⁱⁱ debate that holds that the asymmetry is an illusion^{iv}. But while it is clear that the asymmetry claim has a lot of intuitive pull it is far less clear how to give a convincing account of it. The literature on this topic is enormous and very sophisticated and this paper could not attempt to even begin to do it justice.

Instead it will focus on a very specific heuristic that philosophers have used in order to clarify what it means to have an intention and ask whether a better understanding of this heuristic could help us to disentangle the controversy about how it is that we know our own intentions. The heuristic in question is the idea that in order to acquire an intention to x one must settle the question whether to x. This heuristic has been employed widely. It originates in the philosophy of Elisabeth Anscombe (1963) and is very much alive in the tradition that builds on her work, but the heuristic has also recently been used by different thinkers to clarify the idea of what it is to have an intention in the context of the cognitive sciences. Al Mele for example has used this idea for an argument that is supposed to show that in the Libet experiments there is no unconscious intention preceding the conscious intention^v (e.g. Mele 2009) and crucially Peter Carruthers has used the idea of settling a question in the context of defending his interpretative model of the mind (Carruthers, 2007, 2011). Carruthers' work is of specific interest here because he uses the idea of settling a question to argue for a model of the mind in the theory theory tradition (i.e. the tradition that is most sceptical of the self-other gap), while traditionally the idea of settling a question has been employed to give an explanation why it is that we know of our own intentions in a different way than we know of the intentions of others. It is this interesting use of settling a question for two diametrically opposed positions on the self-other asymmetry question that is the focus of this paper.

After a short outline of the way in which both positions use the idea of settling a question to make their point the paper will proceed to ask whether it can explain why it is that the two positions can use the same idea to come to such different conclusions. Options here are that maybe the two do not mean the same by settling a question. This would be an important result, because it would mean that we always would have to check which sense it is that we are using when making an argument. It could obviously also be the case that one of these understandings is simply the more appropriate one. Another possibility would be that what we mean by settling a question is a hybrid of the two notions.

Finally, it could also be the case that there is only one notion. In which case it will be interesting to see, whether it really is the case that one and the same notion can be

used to arrive at these very different positions or whether one of the positions uses the notion for a bad argument.

1) Why settling a question helps to explain the self-other asymmetry

In his seminal monograph 'Authority and Estrangement' Richard Moran develops an account that gives a very interesting explanation of the self-other asymmetry. One critical feature of this account is an emphasis on the agency that is involved in knowing about one's own mental states. When attempting to find out whether I really believe that the sun never shines in Scotland, I will not normally work like a psychologist and try to find evidence that a belief is in my head and it would seem odd to go to a psychotherapist or a neuroscientist to find the relevant beliefs. Instead, I will simply deliberate and collect evidence for the relevant proposition. Looking out of the window I might find that indeed the sun is not shining at the moment, but as soon as I go back a couple of months in my memory I will presumably find one of the rare instances when the sun was actually shining. I will then have found out that the proposition in question is false. Sometimes the sun does shine in Scotland. The critical point now is that it seems as if by settling this first order question in deliberation about whether or not the proposition in question is actually true I have also settled the second order question of whether I believe it to be true. Moran talks here about the transparency of the belief question. The reasons for ascribing the belief that p to me are transparent to the reasons that I have for p .

As Moore famously pointed out it would be very odd to sincerely assert a proposition p (the sun sometimes does shine in Scotland), while at the same time asserting sincerely that I don't believe that p .

It is worth emphasising however that while this does sound like a nonsensical thing to say in the case of oneself, this is clearly not because the sentence is contradictory.

Obviously, it is quite possible that someone believes that p , when in fact $\neg p$ is the truth of the matter. Given the abundance of rain in Scotland it would not be surprising if someone held the belief that the sun never shines in Scotland, despite the fact that there is relatively convincing meteorological evidence that this belief is false.

The strangeness of Moore's paradox only arises in the case of self. The reason for this asymmetry according to Moran is that only in the case of self am I responsible for coming up with the first order judgment. When observing someone else forming beliefs I am in a purely spectatorial role, but when thinking about my own beliefs this is not the case. I am not only observing myself I am at the same time the agent that is acquiring the beliefs. And I am acquiring the beliefs by deliberating about the subject matter and settling the question of whether or not p obtains. Obviously, this is not the whole story here. Settling the question of whether p does not automatically entail that I know that I now judge that p . Moran has an intriguing and contested story here about how to close that gap, but for now it is enough for us to understand the role that settling a first order question plays in the acquiring of beliefs.

Moran's main target states are beliefs, but it is easy to see how he can produce a very similar argument for the case of intentions. While for beliefs we are settling the question of whether a certain p is the case, in the case of intentions we are settling the question of what to do. Again, we are not 'looking under the hood' to use Moran's expression, to see whether we can find an intention but instead we are deliberating which of the available courses of action is the best one (see e.g. Moran p.105 for a

discussion of intention formation along those lines).^{vi} A nice way to illustrate the difference between thinking about one's future actions in a theoretical way and the normal deliberative way is the case of the gambler. A person who is addicted to gambling might well know in a theoretical psychological (looking under the hood) way that she is very likely to gamble again tonight, but having this knowledge does not mean that she by knowing this automatically acquires the intention to gamble. In contrast, she does acquire an intention when she after deliberation decides that she will not gamble tonight, because she feels that not gambling is the best thing to do over all. The reasons for acquiring the intention here are again transparent to the reasons for not gambling. Obviously by the way, this does not mean that she will not end up gambling after all, but for now she has settled the question about what to do tonight. It is important to stress this last point. Intentions obviously do not always lead to actions, because otherwise it would seem that phenomena like akrasia^{vii} or re-consideration would not be possible once one has an intention. That would obviously be absurd. So critically, all that is required for the acquisition of an intention is that for the moment practical deliberation has come to an end and the person has settled the question.

2) Why settling a question shows that we know of our own intentions in the same way we know of the intentions of others

Carruthers (2007) bases his main argument for the idea that we know about our own intentions in the same way that we know about the intentions of others^{viii} on two main arguments. Firstly there is the incredibly rich empirical literature that seems to show that subjects are surprisingly badly informed about their own motives for acting. This has been dubbed the automaticity juggernaut (Kihlstrom, 2008) or zombie challenge (Vierkant et al, 2013) and refers to findings that show amongst other things that it is possible to manipulate the decision making of subjects from cases as trivial as the speed with which they walk to the elevator to moral decision making without subjects realising that their decision had been modulated by these influences. It could also be shown that subjects fail to realise which decision they made and that they are willing to justify why they made decisions that they never made.^{ix}

In addition famously, Daniel Wegner (e.g. Wegner 1999) showed that he can induce in subjects the belief that they intended to do something when in fact they had no such intention.

Secondly Carruthers thinks that it seems hard to see what the evolutionary plausibility of a self-monitoring system would have been. Humans like all other animals needed to be able to cope with their external environment not monitor their own mentality in order to survive. The focus of early humans would have been outward rather than inward. (see also Prinz 2003 for a similar argument).^x

These two arguments together according to Carruthers make it very implausible that humans have direct introspective access to their intentions. It seems evolutionarily unnecessary that they do and if they did then the rich data from the empirical self-blindness literature seem very hard to explain.

But Carruthers thinks that there is a powerful objection to his argument and this objection involves two systems theory. According to this immensely popular^{xi} theory humans operate with two distinct cognitive systems. System 1 is normally described

as fast, unconscious, parallel, and evolutionarily old, while system 2 is normally understood to be slow, conscious, serial, linguistic and evolutionarily new (e.g. Kahnemann 2011, Carruthers 2009, Evans 2007).

Now if it is true that humans possess these two systems then this allows for an argument that accounts for the empirical data, takes the evolutionary considerations into account and at the same time allows to uphold our very strong folk intuition that we do know in a direct way when we have formed an intention. This argument, which looks like very bad news for Carruthers' claim that self-knowledge of propositional attitudes is always interpretative, goes like this. According to two systems theory much of our behaviour is controlled by system 1 and we do not have direct conscious access to what happens in this system. So all the behaviour that is controlled by this system will have been produced by mental states of which we were not directly aware. This explains the empirical data, we are not directly aware of the behaviour produced by this system and need to self-interpret. This then leads to confabulations if the interpreting mechanisms are tricked. It is also consistent with the evolutionary speculation, because system 1 is the old system and it is - as the speculation expects - purely world directed.

But system 2, one might naturally think, also produces intentions and these we are aware of. Obviously, the data do not show that we are never aware of our intentions, but only that we sometimes get it wrong. But that is not surprising if we are only aware of some of our intentions in a direct way. System 2 is supposed to only control a fraction of our behaviour, so it is not surprising that we can find so many instances of 'self-blindness'. But, it is not surprising either that we find it so intuitive that we are directly aware of our intentions, because system 2 is the one that makes the important decisions and we are always aware of those intentions. Again this is also consistent with the evolutionary considerations because system 2 is supposedly evolutionarily very young and systematic scrutiny of the thought process seems much more useful for a being on the verge of becoming a creature with a scientific mind. ^{xii} Taken together this separation of systems seems to provide a very powerful case against Carruthers claim that all our self-knowledge of propositional attitudes needs interpretation. The argument is very powerful, because it concedes that knowledge of some propositional attitudes does require interpretation (the system 1 attitudes) and because of that concession it can fully agree with the two arguments that Carruthers had for the interpretative model. At the same time it seems to show that the claim fails for system 2 propositional attitudes and it seems as if Carruthers here can't use the same arguments again, because system 2 is not evolutionarily old and having direct knowledge of system 2 propositional attitudes is not incompatible with the empirical literature that shows that we do not have direct access to a large part of our mental life.

It is at this point in the dialectic that Carruthers uses the idea of settling a question in order to disarm the objection. It is undeniable that in system 2 there are conscious inner speech events of the form 'I will do P', and it seems very likely that there is a causal connection between these inner speech events and doings of p by the agent, but Carruthers worries that the causal connection between the two events is not the right causal connection for the inner speech act to count as an intention. Why not? Because the inner speech act 'I will do P' does not settle the question as to whether an agent will in fact do P. Rather according to Carruthers all the conscious speech act will do is to prompt some further system 1 activity as to whether to interpret the speech act as a decision or not. Because of this, the inner speech act, while part of the causal chain

that leads to the action, is not the point where deliberation about whether or not to perform the action stops, but that is exactly what an intention is supposed to do. So the intention only really is formed after the unconscious interpretation of the speech act. But as this processing is unconscious we are not directly aware of it and need to interpret ourselves in order to know what our intentions are. In other words we never are directly aware of our intentions. Our conscious knowledge of our intentions is always theoretical and is based on us interpreting ourselves as having them.^{xiii}

3) Is Carruthers' account at all plausible?

On the face of it Carruthers' argument seems to provide us with a problem for Moran like accounts. If it is true that we find out about our intentions by observing the consequences of system 1 decisions then the self-other asymmetry becomes very shallow. It is obviously still true that I have more access to all kinds of indicators that somebody else would not have normally (e.g. feedback from the body that it is in a state of action readiness), but obviously this kind of advantage is merely a difference in quantity rather than a deep qualitative difference. Our conscious self-knowledge of intentions is on this picture clearly observational in exactly the same way as our knowledge of the intentions of other people. On Carruthers' picture it makes perfect sense to say that a neuroscientist might know about your intentions far more reliably than you do. In fact, one might think that this is not science fiction any longer. There is already an ever growing literature on claims by neuroscientists that they know the intentions of their subjects before the subject (famously Libet 1985, but also very interestingly Soon 2008).^{xiv} Additionally, if the Carruthers picture is right it looks as if our knowledge of our intentions is also metacognitive. This is bad for Moran, because metacognitive processes do not obey transparency. I.e. for metacognitive processes it is not true that in order to find out whether one has an attitude one simply looks at the reasons for the first order proposition. They target the mental processes themselves rather than trying to settle a first order question.^{xv} If a neuroscientist tells me that I have the belief that p that normally will be good evidence that I have the belief, but it might not be good evidence for p itself. The reasons for the first order proposition and the metacognitive reasons for thinking that I am holding the belief come apart and that means that transparency fails in these cases.

One way out of this problem is a flat denial of the plausibility of Carruthers story. This is what Al Mele opts for in his book *Effective Intentions*. Mele here claims that there are clearly cases where conscious intentions lead directly to the sending of the appropriate motor signals (2011, page16). Mele does not give a positive reason why this 'clearly' is the case. Instead he provides us with an argument why the specifics of Carruthers' account are implausible. Carruthers suggests a specific way in which an inner speech act causally contributes to an action. On this account agents normally have a standing desire to act in accordance with conscious inner speech events. When an agent says to herself in inner speech "I will ϕ now", then this can lead to the unconscious belief that one has decided to ϕ and because of the standing desire the agent will then settle on ϕ ing. Mele does not elaborate why this account is implausible apart from the note that it would contradict his architecture of how an intention is supposed to work^{xvi}. I will not attempt here to adjudicate, whether or not it is obvious that Mele's account of the architecture of intention is reason enough to reject Carruthers account. It is certainly true that the specifics of the Carruthers story are

highly speculative and provide us with a rather roundabout story.^{xvii} However, even if Mele is right about the implausibility of the specifics of Carruthers' account, it seems perfectly possible to maintain Carruthers' general point without buying into the standing desire story. All that Carruthers needs for his argument to go through is that some form of unconscious practical deliberation always follows the relevant inner speech act. It seems very plausible that something like this happens at least sometimes. When an agent tells herself to 'get up now' in the morning only to find that instead she has turned round to snooze for another 5 minutes it clearly seems to be the case that the inner speech act that sounded like a decision was in fact no such thing, because the matter was not settled or at least it was not settled in such a way that it conformed with the content of the inner speech. But obviously cases like this are unusual and it is not clear at all that we should think that all conscious inner speech is so devoid of the power of causing the agent to move directly.^{xviii} Carruthers can also invoke Wegner in his defence. Wegner's 'I spy' experiments seem to indicate that it is at least sometimes possible that we do mistake episodes of conscious inner speech for intentions. In these experiments subjects rated an action that they performed involuntarily as voluntary, if they had heard words consciously that were consistent with the movement. According to Wegner they misinterpreted the words they had heard consciously as intentions that produced their movement. But all of this hardly amounts to a knock down argument against the claim that at least sometimes practical deliberation does end with the forming of a conscious intention to which we might have direct access^{xix} and in this case Carruthers would be wrong.

But then neither did anyone provide a clear cut argument that would show that Carruthers is wrong. Rather it seems to be a standoff between intuitions built around folk psychological intuitions and speculations extrapolated from the scientific data. To a degree then, whether or not Mele or Carruthers is right here seems like a simple empirical question. As such we can only await the outcome of the empirical tribunal. However, there is one thorny conceptual issue here. The question is whether there is unconscious practical deliberation. At the moment we don't even know what processes there are and so we clearly lack some empirical data, but even if we did have the data we still would need to answer the question whether these processes really count as practical deliberation that settle the question whether or not to ϕ . In any case at the moment both alternatives seem live possibilities.^{xx} If Mele is right and Carruthers is wrong then this worry about the nature of the self-other asymmetry is answered, but as long as we don't yet know whether it is true that conscious inner speech is always followed by unconscious processes that count as practical deliberation it seems worthwhile to consider the option that Carruthers is right.

Before moving on it seems worth discussing shortly one thought by Shepherd (2013 p.28). Perhaps it makes sense to call the conscious inner speech act a decision that leads to the acquisition of an intention even if it is followed by more unconscious practical deliberation. That seems like an interesting possibility. After all the conscious event still does play an important role in producing the action and it does seem likely that folk psychology will have referred to a conscious event. However, as Shepherd admits this proposal is slightly revisionary and it is also not in line with Mele's definition of a decision. This is because, as we have discussed already, on the standard definition a decision leads to an intention and an intention is only an

intention if it settles the question. But if Carruthers is right then this is exactly what the conscious inner speech act does not achieve.

So at the end of this section the state of play is that Carruthers account is at least one plausible contender for explaining how we acquire intentions and as long as this possibility is still alive and well, it also continues to pose a significant challenge to accounts of self-knowledge that rely on settling a question in order to acquire an intention. It still seems to be a possibility that we acquire intentions unconsciously and only find out about them indirectly by observational means.

4) Does Carruthers account have anything to do with the Moran argument?

So far we have assumed that it is uncontroversial what we mean by ‘settling a question’. But this might be unwise. Carruthers argument relies on a very specific psychological architecture, where conscious ‘decisions’ are system 2 inner speech acts and the claim that these inner speech acts never settle a question of whether to ϕ , where this is understood as the claim that they never directly set the system on a course to ϕ without further system 1 processing as to whether to ϕ .

One first hint that it might be problematic to understand the ‘notion of settling a question’ in such a way can be found in the work of Pamela Hieronymi who clarifies what she means by settling a question:

As should become clear, the appeal to "settling a question" is not meant to introduce an additional psychological state or event. Rather, the claim "to intend to ϕ is to settle positively the question of whether to ϕ " simply notes the uncontroversial conceptual connection between an intention and a positive answer to a certain question (whether to ϕ). (Hieronymi, 2008 footnote 5 p.360)

Hieronymi here clearly urges us to understand the notion of settling a question not as a psychological state at all, never mind a conscious inner speech act. Instead the phrase is used to provide us with a conceptual tool to ascertain whether a subject intends to ϕ . This is the case if the subject would give a positive answer to the question whether she indeed intends to ϕ . Interestingly, she also claims that this tool simply makes use of an “**uncontroversial conceptual connection** (emphasis added) between an intention and a positive answer” to the question whether someone intends to ϕ . For her this link is crucial for understanding intentions, because the stating of intentions is essentially about providing reasons for why one performs or did perform an action. In answering the why question the agent provides her self-understanding of why she is performing the action. She claims that she is performing the action, because of the reason provided by settling the question. Without this self-understanding the action would not be an intentional action and therefore the answer to the why question is the critical element that distinguishes a mere behaviour from an intentional action.

On Hieronymi's understanding the relevant sense of 'settling a question' allows us to ascribe intentions by means of a conceptual link. On this account it is simply true by definition that a sincere positive answer to the question whether to ϕ settles the question in the relevant sense and implies an intention to ϕ .^{xxi} However, there is a glitch here. It is not obvious from the Hieronymi text exactly what a sincere positive answer to the why question looks like. The most natural way to understand this is to understand it as an assertion that a subject would actually make if asked. But we could also understand this asking of a question more metaphorically. It could be that a positive answer to the why question can be provided by any deliberative process that can settle a question.

If the first understanding is the right interpretation of how we should understand the notion of 'settling a question' in the context of self-knowledge of intentions, then it is obvious that Carruthers argument is wrongheaded. On this account it simply does not matter that there might be further unconscious deliberation, which might in the end lead to the person not ϕ ing. It does not even matter that the person at the time of the sincere assertion was psychologically not settled on acting. All that matters is that she would have sincerely asserted that she had settled the matter.

But while this account clearly takes care of Carruthers argument it does so at a high cost. Making the link between assertion and intention conceptual makes it impossible that one could ever be wrong about having an intention. This is very much not in the spirit of Moran's project (to which Hieronymi is normally close) who wants his account to be not merely logical. It also disables conversations between philosophers and cognitive scientists on the nature of intentions, because intentions are now defined in such a way that they clearly do not belong in the realm of the sciences. For example, the large literature discussing the Libet experiments would be far more obviously misguided than assumed, because it would simply be impossible that there could be an unconscious intention.

Secondly, and more importantly it might well rob intentions of their most important distinguishing feature. Intentions in contrast to beliefs and desires are supposed to be executive states^{xxii} (e.g. Bratman 1987, Mele 2009), but if we insist that there has to be a conceptual link between sincere assertion and intention then it might well turn out that they never are. If Carruthers psychological account turned out to be right and it really is the case that a person only ever is ready to ϕ after postconscious system1 processing, then the necessarily conscious intentions never have any executive dimension.

If the second interpretation is right, however, then it looks as if the Carruthers argument still works and Hieronymi would have to accept that if the empirical data work out the way Carruthers predicts we have to accept that our conscious knowledge of our intentions must be observational.^{xxiii}

The way out of this predicament might be, if we simply accept that we are talking about two different senses of settling the question. In the one sense we are talking about consciously settling a question and it seems right that a sincere assertion that one has settled the question whether to ϕ is conceptually linked to the fact that one consciously understands oneself as having settled on ϕ ing. However, understanding oneself in this way does not imply that one is really ready to ϕ , because if Carruthers is right that might require further processing.

But then the question becomes which of the two senses of settling a question is relevant for intention formation?

The answer here surely is both:

It seems right to say, that there is a conceptual connection between understanding ourselves consciously as having settled the question whether to ϕ and the conscious intention to ϕ , but we also seem to need the settling of a question in the Carruthers's sense in order to account for the executive dimension of intention.

But once we accept that there are these two different relevant senses of settling a question, we can now go back to the solution suggested by Shepherd that the conscious decision does not have to directly cause the movement in order to count as having produced an intention, because it is still true that it has a significant influence on the likelihood of the action happening.

We rejected this solution earlier, because it was in direct conflict with the central notion that an intention requires the settling of a question and it is clear that on his solution a "conscious decision" does not settle the question in the Carruthers sense. As we have established now that there are two senses of settling a question involved in intention formation, we have two options that both use the Shepherd move.

On the one hand, we could say that consciously understanding oneself as having settled a question is necessary and sufficient for an intention. This would mean that sometimes intentions do not have the executive dimension that we ordinarily associate with them, but this does not mean that they do not have this dimension most of the time.

This does not seem implausible and is in line with moves that have been made for belief. Pacherie/Bayne (2005) have argued quite convincingly that the strict conditions for belief are too strong.^{xxiv} Discussing monothematic delusions, they argue that in line with intuition we should allow that there are delusional beliefs, despite the fact that patients suffering from the delusion do normally not show a complete pattern of behaviour that would fit with the delusional belief. For example, patients suffering from Capgras syndrome who believe that their partner has been replaced by an impostor nevertheless normally do not call the police or take similar dramatic action which clearly would be called for in the case that there really was an impostor. Pacherie/Bayne argue that we should nevertheless think of these delusions as beliefs, because ordinary beliefs also tend to be context dependent and to not fulfil the high philosophical ideal of being in a fully coherent web that also controls behaviour in a coherent way.

Alternatively, one could argue that the acquisition of an intention always requires the settling of the question in the Carruthers sense. But this does not rule out that we also allow that the conscious settling of the question is also highly significant. It might well be the case that it is necessary to settle a question consciously to acquire a conscious intention, but that while settling the question consciously makes it highly likely that an intention will be acquired it remains possible that one could be wrong about this. This would, obviously, be the case if the question got settled in a different way in unconscious processing. This route takes the executive dimension of intention as seriously as possible and this is obviously a big advantage, because as mentioned

before one of the main arguments why we cannot do philosophy of action with only beliefs and desires was hinging on the point that only intentions are genuinely executive states.

5) Conclusion:

The stated aim of this paper was to get a better grip on the notion of ‘settling a question’ in the context of intention acquisition. The hope was that this would help us to better understand how it is that we know about our intentions and to help understand the intuitive self-other gap. It now is time to take stock and to see what we achieved. As suspected it turned out that there are two different notions of ‘settling a question’ that are not clearly kept distinct in the literature. The paper argued that both notions clearly contribute to our concept of intention and offered two different ways in which these contributions work. There was no space to work out whether one of these two options should be preferred and it is not obvious to me that this could be done. ^{xxv}

However, there are obvious consequences from adopting either of the two models. If we adopt the first model that is good news for Moran like accounts. In that case it seems that the conscious settling of a question is all that is required for intention acquisition. While this still leaves the question of how the acquisition of an intention in that way leads to privileged knowledge there at least does not seem a problem with the fact that sometimes the acquisition of an intention does not settle the question in the Carruthers sense. We can use the Shepherd, Bayne story to make sure that it nevertheless seems ok to say that intentions do have an executive dimension.

If we adopt the second model on the other hand, then it does seem to be the case that we need some modifications for Moran like accounts. In this case, while it would still be true that I could know about my conscious ‘decision’ in a special way, knowing about my intentions would become observational.

Literature:

Anscombe, G. E. M., (1963), *Intention*. Oxford: Blackwell.

Bayne, T (2006). Phenomenology and the Feeling of Doing: Wegner on the Conscious Will. In *Does Consciousness Cause Behaviour* edited by Susan Pockett, William Banks, et al. Cambridge MA: MIT Press.

Bayne, T. (2013). Agency as a Marker of Consciousness. In *Decomposing the Will* edited by Andy Clark, Julian Kiverstein, Tillmann Vierkant. Oxford: Oxford University Press.

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge MA: Harvard University Press.

Carruthers, P. (2007). The Illusion of Conscious Will. *Synthese*, 159: 197–213.

Carruthers, P. (2009). An Architecture for Dual Reasoning. In *Two Minds: Dual Processes and Beyond* edited by Jonathan Evans & Keith Frankish. Oxford: Oxford University Press, 109-27.

Carruthers, P. (2011). *The Opacity of Mind*. Oxford: Oxford University Press

- Evans, J. (2008) Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology* Vol. 59: 255-278
- Gallagher, S. (2008a). Direct Perception in an Intersubjective Context. *Consciousness and Cognition*, 17, 535-543.
- Gallagher, S., & Varga, S. (2014). Social Constraints on the Direct Perception of Emotions and Intentions. *Topoi*, 33, 185 - 199.
- Gopnik, A. and Wellman, H. (1992) Why the Child's Theory of Mind Really Is a Theory, *Mind and Language* 7: 145-171.
- Hieronymi, P. (2008). Responsibility for believing. *Synthese*, 161: 357-373
- Libet, B. (1985). Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action. *The Behavioral and Brain Sciences* 8: 529-566.
- Kahneman, Daniel (2011). *Thinking, Fast and Slow*. New York
- Keren, G. & Schul, Y. (2009) Two Is Not Always Better Than One. A Critical Evaluation of Two-System Theories. *Perspectives on Psychological Science*. 4(6):533-550
- Kihlstrom, J. F. (2008). The Automaticity Juggernaut—or, Are We Automatons After All. In *Are We Free? Psychology and Free Will* edited by John Baer, James Kaufman & Roy Baumeister. Oxford: Oxford University Press
- Mele, A. (2009). *Effective intentions: The Power of Conscious Will*. Oxford: Oxford University Press.
- Moran, R. (2001). *Authority and Estrangement: An Essay on Self-knowledge*. Princeton: Princeton University Press.
- Pacherie, E. (2006). Towards a Dynamic Theory of Intentions. *Does Consciousness Cause Behaviour* edited by Susan Pickett, William Banks, et al MIT Press, pp.145-167
- Pacherie, E. Bayne, T. (2005). In Defence of the Doxastic Conception of Delusions. *Mind & Language*, 20 (2):163-188.
- Pettit, P. and McGeer, V. (2002). The Self-regulating Mind. *Language and Communication*, 22, 281 – 99.
- Prinz, W. (2003). Emerging selves. Representational Foundations of Subjectivity. *Consciousness and Cognition*, 12 :515 – 28.
- Roth, G. (2003). The Interaction of Cortex and Basal Ganglia in the Control of Voluntary Action. In *Voluntary Action. Brains, Minds and Sociality* edited by Sabine Maasen, Wolfgang Prinz & Gerhard Roth, G Oxford: OUP
- Schwitzgebel, E. (2011), Knowing Your Own Beliefs. *Canadian Journal of Philosophy*, 35 supplement 41-62.
- Shepherd, J. (2014): Deciding as Intentional Action: Control over Decisions, *Australasian Journal of Philosophy*, Published online: 17 Oct 2014
- Shepherd, J (2013) The Apparent Illusion of Conscious Deciding, *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 16(1): 18-30
- Soon, C. S., M. Brass, et al. (2008). Unconscious Determinants of Free Decisions in the Human Brain. *Nature Neuroscience* 11(5): 543 - 545
- Vierkant, T., Kiverstein, J., & Clark, A. (2013). Decomposing the Will: Meeting the Zombie Challenge. In *Decomposing the Will* edited by Andy Clark, Julian Kiverstein, Tillmann Vierkant, Oxford: Oxford University Press.
- Wegner, D.M., and T.P. Wheatley. 1999. Apparent Mental Causation: Sources of the Experience of Will. *American Psychologist* 54: 480–92.
- Zawidzki, T. (2008). The Function of Folk Psychology: Mind Reading or Mind Shaping? *Philosophical Explorations*, 11 (3), 193 – 210.

ⁱ And other propositional attitudes, but there is a large and specific literature on intentions. For this reason this paper focuses on intentions, but similar arguments could be made for e.g. beliefs.

ⁱⁱ This seems very intuitive, but it has been challenged in the philosophical literature. Some philosophers argue that we can directly perceive the mental states of others (e.g. Gallagher 2008, Gallagher & Varga 2014).

ⁱⁱⁱ Even the name reflects the worry. The debate used to be known as the theory of mind debate, but people were worried that this seemed to prejudge that mindreading really is a theoretical activity.

^{iv} Or at least very shallow. Theory theorist Alison Gopnik (1992) e.g. explains the intuition of the self-other gap with the fact that we are experts about our own mentality. Like chess experts might feel that they can directly see mate in three, when we know that this seeing is built on the intense practice of the relevant inferences it seems to us that we directly know which mental states we are in, when in reality we know this because of highly automatised inferences.

^v Mele argues that this is because in Libet's veto experiments subjects are clearly settled on not moving, so the conscious urge to move they have before they abort the movement cannot be an intention to move, because subjects were always settled on not moving.

^{vi} Hieronymi discusses the structural similarity at length in her 2008 Reason for Believing

^{vii} Thanks to Mike Ridge for reminding me to clarify the akrasia point.

^{viii} To be clear: I am emphasising here the point that Carruthers does argue that self-knowledge of intentions is observational just as the knowledge of the attitudes of others and that is the critical point under discussion here. However, he does argue as well that in the case of self-knowledge the system has access to more material to make the inference, because we are e.g. introspectively aware of our inner speech.

^{ix} Shepherd (2013) runs the argument that confabulating the motives for your decision does do nothing to show that you were not aware of the decision. While the findings discussed under the zombie challenge might show that it is possible to modulate a decision with unconscious influences this does not show that the decision itself was not conscious. While it e.g. might be possible to influence a moral decision like whether to help a stranger or not by priming them unconsciously this does not mean that the decision of whether to help or not is unconscious (see Bayne 2013 for a similar argument). That is an important clarification, but these studies are not therefore unimportant, because they undermine the idea that these conscious states are real decisions, because decisions are supposed to cause action, but here it looks as if the state that the agent is not aware of produces the action (whether to help or not) and the conscious event seems to have little influence. Obviously, this is no knock down argument, because the conscious events could play an important role, but it puts the onus on the consciousness defender, if it looks as if the story could also work without it.

^x In the Opacity of Mind Carruthers makes the further evolutionary observation why this outward directedness also predicts that humans would have understood other minds first. Humans live in highly complex social environments, where one crucial element of the environment that the individual has to cope with is the other human.

Being able to predict the behavior of conspecifics is here clearly a great advantage (p.64-66)

^{xi} But not uncontroversial see e.g. Keren & Schul (2009)

^{xii} Shepherd (2014) e.g. argues that being able to form an intention on whether or not to continue to deliberate is an important advantage of conscious decision making.

^{xiii} Carruthers has a very elaborate account of what he sees as the right cognitive architecture for this self-interpretation process involving e.g. an unconscious mind reading module. But as far as I see the argument does not depend on accepting this architecture. As the argument is stronger anyway if it does not require a specific and highly contested architecture I do not discuss this architecture here.

^{xiv} Obviously, these results are controversial and even if we were to take them at face value they would still not show that neuroscientists could know of all intentions before the subject. Thanks to Leon de Bruin for pushing me on this point. Still it might seem worrying enough that neuroscientists know about a large class of intentions earlier than the conscious subject and at least for those intentions Carruthers claim seems to be borne out by empirical studies already. Carruthers argument is supposed to give us a reason to suspect that this holds not only for those limited cases but for all intentions.

^{xv} So if Carruthers is right, then the point that Kloosterboer (this issue) makes for emotions actually would hold for intentions as well. Knowing about our intentions does at least have a self-referential component.

^{xvi} For a very similar point see Shepherd (2013)

^{xvii} However, it might actually fit quite nicely with recently quite popular mind shaping accounts. The idea that our folk psychology provides us with a normative framework that orients our behaviour (Pettit & McGeer 2002, Zawidsky, 2008) fits quite nicely with Carruthers' implementation idea that our conscious inner speech assertion of having an intention leads to the relevant behaviour via a desire to act according to these assertions because they are norms.

^{xviii} There seems also to be neuro-scientific evidence to back up such a view. The neuroscientist Gerhard Roth (2003) for example claims that all voluntary actions go through unconscious censorship that is concerned with two questions, whether the action is better than any other action and whether the intended action is appropriate.

^{xix} see e.g. Bayne (2006) for a similar point

^{xx} Thanks to Dave Ward, Josh Shepherd and Suilin Lavelle who all raised this important issue.

^{xxi} Importantly, Hieronymi does not think that any assertions or other intentional actions can be what settles a question. Settling a question is according to Hieronymi an evaluative not an intentional activity. So asserting an answer to a question is a way of finding out whether a person has settled a question. It is not the question settling itself.

^{xxii} i.e. the crucial characteristic of intentions is that they are directly linked to the execution of the action, in contrast to other attitudes like beliefs and desires whose role in practical deliberation can be detached from any potential behaviour.

^{xxiii} It would however be still possible that we have special agential unconscious knowledge of our intentions. This would have interesting consequences for the importance of those unconscious states for moral agency. I explore this option in a forthcoming paper.

^{xxiv} Eric Schwitzgebel in fact makes this point in the context of self-knowledge of beliefs in a 2011 paper. While I am happy to accept his pluralist approach for belief, intentions with their executive dimension seem to demand a special treatment.

^{xxv} Obviously, there is also a different conclusion that one could draw from this paper. If it is the case that the notion of settling a question is less clear than we thought it was, then perhaps it was a mistake in the first place to rely too much on it. It might e.g. be the case that there are different types of intentions and perhaps the idea of settling a question applies only to some or even none of them. The idea of motor intentions as suggested by Elisabeth Pacherie (e.g. 2006) looks like a prime candidate here. I am happy to accept that that is a possibility. But given how important the notion is in the philosophical literature from Anscombe through to Bratman and Mele I think that would be also a very significant result. Thanks to Leon de Bruin for pressing me on this.