\*\* Penultimate Draft \*\*

Please see published version with final corrections in *The Philosophical Quarterly*

**The Subjective Authority of Intention**

**Lilian O’Brien**

While much has been written about the functional profile of intentions, and about their normative or rational status, comparatively little has been said about the *subjective authority* of intention. What is it about intending that explains the “hold” that an intention has on an agent – a hold that is palpable from her first-person perspective? I argue that several *prima facie* appealing explanations are not promising. Instead, I maintain that the subjective authority of intention can be explained in terms of the inner structure of intention. In adopting an intention the agent comes to see herself as criticizable depending on whether she executes the intention. This allows us to explain in first-personal terms why the agent becomes disposed to act and deliberate in ways that are characteristic of intention. As intention-formation involves profound changes to reflexive evaluative attitudes, this is the “Self-Evaluation” view of the subjective authority of intention.

**Keywords:** intention; subjective authority; planning agency; diachronic will; reflexive attitude

Suppose that today you intend to work on co-ordinating exams for the department, you intend to make arrangements for your holidays, and you intend to do a number of other things besides. You are in the midst of this very busy day and must-thoughts clamour for your attention. You must make that phone call, you must attend that meeting, and so on. Your must-thoughts have a *subjective authority* – in having them you accept what you take to be legitimate demands on your behaviour. What is it about these thoughts – their inner structure or their relations to other things - that explains this feature of them?

As these must-thoughts flow from prior intentions, the intentions seem to be the source of the subjective authority of the must-thoughts. But what, then, explains the fact that these prior intentions have subjective authority? How should we explain the fact that you are not just pushed or pulled by them, as you might be by a desire – your intentions seem instead to you to make legitimate demands on you.

While much has been written about the functional profile of intentions, and about their normative or rational status, the aim of this paper is to shed light specifically on the psychology of the “hold” that an intention (and ensuing must-thoughts) has on an agent from her first-person perspective. How does the intention exert its rational – as opposed to non-rational – pressure?[[1]](#footnote-1) Does the agent take the intention to be a reason for action so that the subjective authority of the intention is of the same kind as that which an agent feels when she thinks that she has sufficient or conclusive reason to do something? Or is it that she values some feature of the intention, such as its role in co-ordination, and it is this psychological state of valuing that explains the subjective authority of the intention? It is widely accepted that intentions exert rational pressure on the agents who form them and that this is palpable to the agent from her first-person perspective. As long as we reject the view that intentions are like ‘lit fuses’ operating independently of the conscious will, there is a question about what it is like for the agent to form and have an intention.[[2]](#footnote-2)

Although there are important points of contact between the psychological question and questions concerning the normative or rational status of intention, the question should not be treated primarily as an issue in the theory of rationality or in value theory: the question concerns why intentions *feel* authoritative, it does not centrally concern why they are authoritative or if they are.[[3]](#footnote-3) Answering this question will allow us to better understand the psychology of rational planning agents. I believe that it will also pay dividends in our understanding of intentional actions, practical knowledge, rationalizing action explanation, and the differences between intentional actions and side-effects. Its potential to help us with these questions is unsurprising given that this issue is central to the philosophy of action: it concerns the agent’s first-person perspective in the exercise of rational control over her behaviour.

I will assume here that an intention’s subjective authority is not a matter of the agent’s thinking that the originating intention is itselfa conclusive reason to act. Among reasons to doubt this, such as worries about an agent’s own sensitivity to the seeming absurdity of boot-strapping, this proposal does not accurately represent the phenomenology of intending some course of action.[[4]](#footnote-4) I will argue that the subjective authority also does not stem from the agent’s *thinking that she has* *sufficient or conclusive practical reasons to intend and act*. This is a bit more surprising, I think. I will also argue that the *many benefits to the agent of having intentions* cannot be mined to explain the hold that her must-thoughts have on her. In short, none of the *prima facie* appealing accounts of the subjective authority of intentions and the must-thoughts flowing from them is adequate.

I will argue that the subjective authority of intentions and their must-thoughts should be explained in terms of the inner structure of intention. Roughly put, adopting an intention involves coming to see yourself as criticizable depending on whether you execute the intention. In effect, you become disposed to adopt negative or positive attitudes towards yourself depending on whether or not you act as you intend, and you also become disposed to deliberate and act in certain ways. Because of the rather profound changes to one’s reflexive evaluative attitudes when one forms an intention, I will call this the Self-Evaluation view of the subjective authority of intention.

**1. Competing Ways to Explain Subjective Authority**

**1.1 Reasons**

At least since Michael Bratman’s seminal discussion in *Intention, Plans, and Practical Reasoning* (1989), many philosophers take it that intentions are by their nature conduct-controlling. I accept this view of intentions here. Unlike when one merely desires to perform an action, when one intends to perform action A in circumstances C, one feels rational pressure to stop deliberating about competing ends, toengage in means-end reasoning where necessary, and to perform A once C obtains:

**Subjective Authority (SA):**

If an intention to A has subjective authority over its bearer, S, about what to do in circumstances, C, S *experiences normative or rational pressure* to (i) constrain deliberation about alternative ends in C, (ii) to engage in relevant means-end deliberation, (iii) to perform A once C obtains or (iv) to re-assess A-ing in the light of reasons if she comes to think that A may not be the thing to do.

The central question of this paper can be sharpened by filtering it through SA. Why, *from the agent’s perspective*, does the formation of an intention to A in C give rise to her feeling that she must do things along the lines of (i) to (iv)? Of course, a functional explanation of these features is already available – roughly, if intention is to play its role in facilitating intrapersonal and interpersonal coordination, it must involve (i) to (iv). But this doesn’t tell us how forming an intention changes things from the agent’s perspective. At a minimum we should note that it does not follow from the functional explanation that the agent feels bound to act in ways (i) to (iv) because she values or cares about the function - the co-ordinating role - of her intention. From her perspective, intrapersonal and interpersonal co-ordination may be, if she is much aware of it at all, a happy byproduct of her intention. In fact, I will argue below that we should look somewhere other than her awareness of the value of co-ordination to account for her dispositions to act in line with (i) to (iv).

It is tempting to suppose that the subjective authority of an intention to A in C is explained by the agent’s sensitivity to the practical reasons, sufficient or conclusive, that she takes herself to have for A-ing in C:

**Reasons:** The subjective authority of S’s intention to A in C is explained by the fact that in thus intending S takes herself to have sufficient or conclusive reason to A in C.

Perhaps the agent is somehow - via her intention - made sensitive to the reasons that speak in favour of the intended action. This suggestion is problematic. If intention had its subjective authority because in having the intention the agent somehow had access to the reasons that seem to her to favour her action, then the intention would not have the reasons-independent subjective authority that it is typically taken to have. Intention is widely understood to operate as a kind of stand-in for deliberation. Intentions relieve the agent of the cognitive burden of being aware of, or of weighing the reasons that seem to her to favour, her intended action. This reasons-independent nature of the subjective authority of intention is reflected in the first-person experience of executing intentions. When an agent recalls something that she intends to do, the agent usually thinks that she *must* do the relevant thing. This must-thought has a peremptory quality - it does not *recommend* the course of action, given certain reasons. Rather, the must-thought seems to present the action as simply *required*. In fact, when the agent recalls what she must do, she does not even seem to be in a deliberative frame of mind in which she might be sensitive to the reasons that favour her doing the thing in question.

Given this, a more plausible way to interpret Reasons is that the agent takes the intention as a sign that she has sufficient or conclusive reasons to act. The agent may trust that her prior deliberating self found sufficient or conclusive reason to A, and she may think that this is why she formed the intention to A in C.[[5]](#footnote-5)

**Deliberative Trust:** The subjective authority of S’s intention to A in C is explained by the fact that S takes herself to have adopted the intention because she had discovered sufficient or conclusive reasons to A in C.

But serious problems face Deliberative Trust.[[6]](#footnote-6) It seems possible that an agent could intend to A in C and that she could accept the demands of must-thoughts flowing from her intention even when she thinks that she has insufficient reason to perform the intended action. The agent who decides to go to the pub at lunchtime, although she also judges that she shouldn’t do this, may think on her way to the pub that she *must* go to the ATM as she will need cash at the pub. This must-thought seems to have subjective authority. But going to the pub is not, the agent thinks, supported by reasons, nor does she think that she was trustworthy when she arrived at the intention to do this.

I am not claiming that reasons that an agent takes to favour an end play no role in the end’s having a kind of subjective authority for the agent. Such reasons may seem to the agent to make some kind of demand on her in the deliberative context, or in a context where the agent defends the pursuit of the end to a sceptic. Nevertheless, the foregoing suggests that we should not rush to the judgment that the subjective authority of downstream must-thoughts flows from the agent’s appreciation – direct or indirect - of practical reasons favouring (or seeming to her to favour) her intended course of action. This caution is strongly supported by akrasia cases. But it is also supported by reflecting on those busy days – there is little time for sensitivity to reasons and energy is devoted to “must” “need to” “have to” thoughts that seem to impose strict requirements to act rather than serve as recommendations for acting. We have good reasons, I think, to reject Reasons or Deliberative Trust.

**1.2. Co-Ordination**

If it weren’t for intentions it is unlikely that you would have many of the things in your life that you value, such as your education or your job. Securing such things usually requires that you organize the pursuit of your complex, competing, and long-range goals, which is to say that they require you to form intentions. Perhaps, it is the invaluable role in organizing and structuring the pursuit of your goals that explains the subjective authority of one’s intentions.

**Co-Ordination:** The subjective authority of S’s intention to A in C is explained by the fact that, in general, intentions allow S to co-ordinate the achievement of goals that compete for her attention, and S appreciates this, and is motivated to preserve the co-ordinating role of intention.[[7]](#footnote-7)

Co-Ordination also faces serious problems. First, it requires that the agent is herself sensitive to the valuable co-ordinating role that intention plays. This requires meta-psychological reflection on the role of intentions in the agent’s life, a positive evaluation of their role, and, because of this, acceptance of the intention as authoritative. But this implies that agents who do not engage in such reflection either do not have intentions or have intentions that lack subjective authority. If it is constitutive of intentions that they have subjective authority, the second option should be rejected. But the first is also unpalatable: even though planning agents are sophisticated in certain ways, it seems possible that they can have intentions without having reflected upon the role of intentions in co-ordination. It has, after all, taken considerable philosophical reflection to come to appreciate the role of intentions in co-ordination. Furthermore, although intentions help you, in general, to achieve valuable things, this fact does not obviously support executing any particular intention. It is not obvious that one can transition rationally from the general belief that intentions are immensely valuable, to a belief about the authority of some particular intention.[[8]](#footnote-8) But if Co-Ordination were correct, we would have to believe that all agents who have intentions are making, or are at least capable of making, this transition.

**1.3. Investment**

Once you form some particular intention, you create a psychic structure for co-ordinating deliberation and action, and this imposes costs on you. If you were not to act in accordance with an intention, then the impact that it had on limiting your deliberation about alternative ends, provoking deliberation about means, and executing sub-plans, would have been for nothing. Perhaps, the intention’s authority is explained by the agent’s aversion to such wasted efforts.

**Investment:** The subjective authority of S’s intention to A in C is explained by the fact that once an agent adopts an intention she makes deliberative, actional, emotional, or imaginative investments, and the agent does not want to waste these efforts, which she would do were she not to execute the intention.

This also faces a number of problems, I will mention just one serious one: it does not explain the authority of an intention for which no costs have been incurred. One forms the intention to do a PhD, and it seems to exert authority over one immediately, before any significant cost, psychic or otherwise, is incurred – intention’s subjective authority is a constitutive feature of it.

**1.4. Knowledge, Self-Trust**

Intentions bear intimate and systematic relationships to things of value and it is appealing to suppose that intentions can gain their subjective authority by their relationship to these things. When one intends to A in C, one arguably thereby acquires beliefs about what one will be doing in C. In fact, a long-range intention of considerable complexity gives one lots of beliefs about where one will be and what one will be doing in the future, and it has the potential to yield lots of knowledge (or true beliefs with a high epistemic status, say) should one act in accordance with the intention and its sub-plans:

**Epistemic Windfall:** The subjective authority of S’s intention to A in C is explained by the fact that once an agent adopts an intention she acquires beliefs that she can make true by acting on her intention. This epistemic windfall would be lost if she didn’t act on the intention and she is motivated to benefit from it.[[9]](#footnote-9)

Suppose that you abandon your intention without re-deliberation, or suppose that you keep doubting whether it is the right thing to do, don’t you treat your deliberating self with an unhealthy disrespect? Won’t this, in turn, undermine your confidence in your ability to follow through on future deliberations?

**Self-Trust:** The subjective authority of S’s intention to A in C is explained by the fact that once an agent adopts an intention, persistent groundless doubt about it, re-deliberating without motivation, or simply reneging, threatens the agent’s self-trust, and she is motivated to preserve her trust in herself.[[10]](#footnote-10)

I think that Epistemic Windfall and Self-Trust face serious problems. First, they require that any agent who has an intention must be able to find their intention authoritative on the basis of the intention’s relationship to knowledge or self-trust. We are trying, after all, to explain the *subjective* authority of the intention – why the agent finds her intentions authoritative. But it seems possible that there are simple and thoughtless planning agents who never realize that there is a connection between intention and such things as knowledge or self-trust. Second, even if agents systematically came to appreciate the connection, this would still leave too much to the contingencies of reflection and deliberation. It seems possible that agents could deem the value of knowledge or self-trust irrelevant to whether to continue to abide by some particular intention. If so, these agents shouldn’t be able to continue to have such intentions (on the assumption that intention’s subjective authority is a constitutive feature of it). But couldn’t one still feel that one must stick to one’s intentions even if one doesn’t value knowledge or self-trust? Finally, such views characterize the subjective authority of intention in “deliberative” terms – that is, gaining knowledge or self-trust are reasons that favour acting on the intention. But this doesn’t fit easily with the peremptory nature of must-thoughts, which impose strict demands that one act in a certain way, and do not seem to merely recommend courses of action because they will promote things of value, such as knowledge and self-trust.

**1.5. Dilemma**

The discussion so far suggests a dilemma. On one horn of the dilemma, we have Co-Ordination, Investment, Epistemic Windfall, Self-Trust, and on the other, Reasons and Deliberative Trust. On the first horn, the authority of intention is “thick” and the agent who accepts the demands of an intention does so because she is responsive to what she takes to be the value of things such as co-ordination, self-trust, or knowledge. Intending taken alone and independent of its content exerts what seems to her to be a distinctively normative pressure. But this requires sophistication in reflection and deliberation that may be beyond many planning agents, it doesn’t sit easily with the constitutive nature of the subjective authority of intention, nor with the peremptory nature of must-thoughts.

Reasons and Deliberative Trust seem to avoid these failings by attempting to explain the subjective authority wholly in terms of the agent’s sensitivity to the reasons that (she takes it) favour the intended action. On such approaches the subjective authority of intention is “thin”, wholly explained by the normative authority of reasons that seem to the agent to favour the intended course of action. But this seems to strip intention of its content-independent pressure on the agent. This is seen quite clearly in akrasia cases: the fact of the persisting intention seems to bring some kind of force or pressure with it irrespective of the fact that the agent takes there to be insufficient reason to pursue the intended course of action. Where Co-Ordination, Investment, Epistemic Windfall, and Self-trust make the distinctive subjective authority of intention too robust, Reasons and Deliberative Trust do not make it robust enough. In the next section, Section 2, I will develop a view that, I will argue in Section 3, avoids both horns of the dilemma.

**2. The Structure of Intention**

Suppose that an agent, let’s call her Concerned, intends to go to the cinema with her friend on Thursday evening. She later plans with her sister to go for dinner on the same evening. When she realizes that she has made conflicting plans she chides herself for her carelessness and she sets about cancelling one of them. Suppose that Concerned also intends to make a Christmas cake. But she gets busy with other things and doesn’t get around to figuring out how to do this until mid-December. At this point, she learns that Christmas cakes have to mature, and so, must be made in November. She thinks ruefully that she should have figured out how to make the cake earlier. Suppose that Concerned also intends to read War and Peace by year’s end. Most evenings she has an opportunity to read, but she usually watches old episodes of Arrested Development instead. It’s year’s end, she hasn’t read more than the first chapter, she should have followed through, she thinks. In a contrite mood, she cancels her streaming subscriptions.

Now consider another agent, let’s call her Unconcerned. She also makes inconsistent plans for Thursday evening, she also intends to make the Christmas cake but fails to engage in effective means-end reasoning, she also intends to read War and Peace but repeatedly fails to take the means. But the difference between Concerned and Unconcerned is that on each occasion where Unconcerned doesn’t act on an intention, she feels no inclination whatsoever to chide herself. In addition, she doesn’t take compensatory courses of action. She doesn’t change conflicting plans, she doesn’t deliberate about how to redress her mistakes, she doesn’t try to make the means more attainable, and so on.

Where the first agent is a mundane example of a flawed planning agent – the failures, the inclination to self-criticism, and the changes of plan are unsurprising, intelligible, natural - the second agent is not just unusual, but very puzzling. It is hard to make sense of someone who says that they intend to do something, but who feels no inclination whatsoever to regard her not doing it (where it remained within her control and no new reasons came to light against doing it) as some sort of failure of hers - shouldn’t she be at least in some cases a little bit disappointed or irritated with herself? And shouldn’t she make further plans on the basis of her failures? In fact, where it seems clear that Concerned is capable of genuinely intending to do things, it is not clear that Unconcerned is. It is not just that she is insufficiently self-critical, it’s that the absence of self-criticism combined with the absenceof a practical response indicates a failure to intend.

It might be objected that Concerned is self-critical just because she places a high value on Christmas cakes and reading War and Peace, and Unconcerned just doesn’t value them all that much, and consequently, is not self-critical and makes no practical response.[[11]](#footnote-11) The presence or absence of self-criticism tells us, then, only about what the agents value highly and tells us nothing about the nature of the subjective authority of their intentions. But we can imagine Concerned casting her eye over the weeks running up to the holidays and enumerating her failures – she sets aside the loss of the cake and the fact that her copy of War and Peace is gathering dust under the couch, and ruminates instead on the fact that she has not been very good in general at bringing her plans to fruition. This disappointment with herself seems to be disappointment with herself just qua agent.[[12]](#footnote-12) And suppose that Concerned comes to know that even if she had made the cake it wouldn’t have turned out well and no-one would have eaten it. Her reaction might be to think that she was just lucky, then, that she hadn’t made it, but that, still, she did not act quite as she should have. This shows, I think, that the objection is almost certainly too quick. Why is it so natural to engage in self-criticism of the agential sort and why is it natural to make an alternative practical arrangement, such as suspending one’s streaming subscriptions? And, why does lack of self-criticism and lack of practical response suggest lack of intention?

A promising explanation is that *in intending* to do the things that Concerned intended to do, she regarded herself as evaluable as a failure if she did not do them as long as she hadn’t changed her mind and as long as they remained within her power.[[13]](#footnote-13) In fact, the naturalness and intelligibility of her self-criticism, the fact that some of the self-criticism is directed at herself just *qua agent*, and the fact that Unconcerned doesn’t even seem to intend, suggests that it is *constitutive* of intention to regard oneself as criticizable as an agent for not doing the intended thing or for failing to deliberate about means, take means, etc. (as long as one hasn’t changed one’s mind and it still seems to be within one’s control to do the intended thing). More specifically, in forming an intention one holds *oneself* to a standard: one must A in C where it remains within one’s power. The standard is a practical standard – a standard of action that applies to one qua agent. One must token the relevant act-type or one must token it in a certain way at a certain time and place, and so on. Holding oneself to such standards plausibly involves regarding oneself as criticizable in light of this standard – one regards oneself as a failure as an agent if one does not meet the standard and a success if one does. We might capture this as follows:

**Structure (of intention):** Intention is a psychological state of holding oneself to standards in light of which one regards oneself as evaluable as an agent.

On the basis of this, we could characterize the structure of intention in somewhat more specific terms as follows: when an agent forms an intention to A in circumstances C she adopts a reflexive evaluative attitude – an attitude of regarding herself as a failure if she doesn’t A in C (as long as A-ing seems to be within her power and as long as no countervailing reasons have come to light) and a success if she does:

**Self-Evaluation (Proto):**

An intention to A in C has subjective authority for its bearer, S, in virtue of the fact that in intending to A S sets herself a standard by reference to which

1. she regards herself as a success as an agent if she As intentionally in C
2. she regards herself as a failure as an agent if she does not A intentionally in C (where A-ing is within her control and S is aware of this and S has not re-deliberated about whether to A and discovered that she has insufficient reason to A).

Although Self-Evaluation (Proto) seems to get something right, it falls short. Concerned makes key practical responses to her failures: she makes compensatory adjustments to her plans to rectify what she regards as her mistakes, such as engaging in further deliberation, and shaping the means to make it easier for her to achieve her end. As it stands, Self-Evaluation (Proto) focuses exclusively on the agent’s disposition to take reflexive evaluative attitudes, but this disposition is not clearly connected to the agent’s practical response to her failure. Intentions, be they distal or proximal, are executive states - they *guide* the agent in the *execution* of her plans. We should amend Self-Evaluation (Proto) to reflect this:

**Self-Evaluation:**

In a context in which (1) S intends to A in C, (2) where A-ing is within S’s control and she is aware of this, and (3) where S has not re-deliberated about whether to A and discovered that she has insufficient reason to A, S’s intention to A in C has subjective authority for S, in virtue of the fact that in intending to A S sets herself a standard by reference to which,

1. she regards herself as a success as an agent if she As intentionally in C,
2. she regards herself as a failure as an agent if she does not A intentionally in C,
3. she regards herself as a failure if she does not consider or make compensatory adjustments to her behaviour where there is opportunity to rectify what she regards as her failure to A in C.

A few things are important to note about Self-Evaluation before explaining its relationship to Subjective Authority (SA). First, Self-Evaluation does not entail that the agent actually takes a reflexive evaluative attitude. The reference to an agent’s *regarding* herself in a particular way implies only that the agent has a disposition to take such an attitude. An agent may, for example, be so defensive about her failure or so shy about her success that her disposition is masked. When the disposition is manifest, taking an evaluative attitude might involve a wide array of things - mild to severe disappointment with oneself as an agent, criticism of oneself as an agent, feeling proud of oneself as an agent, chiding oneself as an agent, and so on.

Second, as with the taking of evaluative attitudes, whether or not the agent makes compensatory adjustments to her behaviour depends on many factors. It is perfectly possible that the agent will not make the adjustments, but again, there must be something that masks the manifestation of the disposition to do so.

Third, the reference to “consider or make compensatory adjustments to her behaviour” is deliberately open-ended. It does not, for example, require that the agent try again to perform the very action that was intended, and it may require as little as a brief deliberation to determine that no further action need be taken. The important point is that in virtue of having an intention to A in C the agent has a disposition to treat what she regards as her failure to A in C as having implications for what she must do next.

Fourth, viewing oneself as a failure as an agent involves viewing oneself as the agent who is tasked with the tokening of the act-type specified in the content of the intention. Being a success is to bring it about intentionally and to be a failure is to fail to do so in spite of the fact that it remained within one’s control to do so. According to the Self-Evaluation view, the agent has at least a limited conception of herself as the one who is planning and executing the plans and it is at least this limited conception that is at work in this viewing of oneself as a failure or a success as the tasked agent.

Finally, the agent’s willingness to engage in reflexive evaluative attitudes may also involve a willingness to accept the critical or approving attitudes of others. She may, in fact, share the details of her intentions with others precisely so that they may remind her of her commitments and help her to keep them. Although the focus in this paper is on the manner in which intentions shape the psychology of the lone agent, there are potentially fruitful connections between these reflexive evaluative attitudes and the role that they play in shaping the agent’s relations to other agents.

**3. Self-Evaluation, Subjective Authority, and First-Personal Explanations**

Self-Evaluation is motivated by the comparison of Concerned and Unconcerned, but it can also be recommended on the basis that it promises an explanation of elements of Subjective Authority (SA) in distinctively first-personalterms.

We may wonder why an agent intending to A in C feels pressure to find a means to A-ing. Why does an agent intending to A feel pressure to stop deliberating about competing ends? Why does she feel that she must try to perform the action, A, when she is in C? Why does she feel pressure to deliberate about whether there is sufficient reason to A if she comes to worry that she should not, after all, A in C? The Self-Evaluation view implies that the answers to these questions turn, not on the fact that the agent is sensitive to the balance of reasons regarding the planned course of action, or because she values and does not want to jeopardize things of value such as intrapersonal co-ordination or knowledge. Rather, the answers to these questions turn on the fact that in forming an intention, *the agent has come to view herself as criticizable, as at fault as an agent, if she does not A in C*. In intending to A in C, she regards herself as at fault if she does not find a means to A-ing in C, and she is at fault if she doesn’t attempt to A in C where A-ing is, she thinks, within her control. She is, by her own lights, at fault for jeopardizing her A-ing in C by needless deliberation about competing ends, and if she doubts that A-ing in C is the right thing to do, she faces the burden of establishing that it would be unreasonable of her to regard herself as being at fault if she didn’t A in C. Consequently, she feels pressure to re-deliberate about, rather than simply abandon, her intention.

The Self-Evaluation view is not the view that the agent values such things as her successful execution of adopted plans or her effective practical agency, and because she values such things, is motivated to comply with her plans, thereby serving what she values. A rather different thought process is involved – in forming the intention, the agent comes to regard herself as a failure as an agent if she doesn’t do the kinds of thing that, she thinks, will allow her to successfully execute her plan. As long as one plays the role of agent, and as long as one doesn’t question the appropriateness of the plan, one’s willingness to regard oneself as a success or failure in light of the plan amounts to one’s acceptance of the plan as authoritative for one. This is a point to which I will return in the next section. The psychological mechanism is also unlike the kind of mechanisms that may be involved in desire. Things desired do not seem to have acquired the status of a legitimate demand on an agent’s deliberative and actional resources. By contrast, the undertaken tasks of the planning agent are regarded by her as legitimate demands that must be accommodated in action and deliberation on pain of her failure.

As the Self-Evaluation view does not attempt to explain the subjective authority of intention by appealing to the agent’s aversion to wasted effort (as Investment does), but by appealing only to the inner structure of intention, we can see why there is subjective authority just as soon as the intention is formed and before any effort is expended. And because the subjective authority does not flow from the agent’s appreciation of things of value that are extrinsic to the intention, such as self-trust or knowledge, it does not require that planning agents are responsive to such things. And as the view does not explain subjective authority by appeal to the practical reasons that seem to the agent to favour the intended course of action, room is left for intention and ensuing must-thoughts to have a subjective authority that is independent of such reasons. Consequently, the view copes better with akrasia cases and with the puzzle of busy days when weighing reasons is beyond the agent’s capacities. The Self-Evaluation view avoids, I believe, the horns of the dilemma described at the end of 1.5.

**4. Objections**

As I emphasized in the introduction, the main aim of this paper is to shed light on the psychology of planning, and more specifically, to shed light on the psychology of the *felt* or *subjective* authority of intentions. As there is already a large literature on the normative status of intention and on questions concerning when it is rational to act on an intention and when it is rational to re-deliberate, it seems appropriate to devote this discussion to the first-personal psychology of intending.

It should be noted that the Self-Evaluation view is consistent with the claim that planning agency involves a kind of normative or evaluative illusion. If the Self-Evaluation view is correct, intending agents cannot help but view themselves as criticizable as agents when they intend. And yet, although failing to execute an undertaken task that remained within her control may deprive the agent of the good that she was pursuing, it could be that there is, appearances to the contrary, no further objective loss in terms of “agential failure”.

But this would not be a sufficient reason to reject the Self-Evaluation view. We may instead lament the fact that agents cannot help viewing themselves as criticizable qua agents when they form intentions, and cannot help feeling motivated to avoid such failure. And we may need an error theory to explain why planning agents do this. But if my arguments are sound, we should still accept the Self-Evaluation view as a correct claim about the first-personal psychology of intention-formation and -execution.

It is clear that clarifying the normative status of intentions is an important task for any complete theory of planning agency. We need to know what, if anything, explains the seeming normative heft of intentions over and above the normative authority of ends pursued. It may be that things - some briefly discussed earlier - such as the value of knowledge, self-trust, co-ordination, self-government[[14]](#footnote-14), self-constitution, or the value of achievements[[15]](#footnote-15) over and above the value of ends attained by those achievements, might be called on to answer this question – let’s call it the Normative Question for brevity. We might draw, for example, on a view such as Korsgaard’s constitutivism to address the question. Let’s suppose that we can provide a compelling account of the link between intention-formation and -execution, on one hand, and self-constitution, on the other. Roughly, on such an approach, executing previously undertaken tasks would be a route to constituting oneself. Were the agent to fail to execute tasks that she had earlier undertaken, she would fail to constitute herself, and fail to be the kind of being that has a unified will and is subject to requirements of instrumental rationality.[[16]](#footnote-16)

I have assumed that we can address the psychological question without addressing the Normative Question. But, couldn’t the agent pose the Normative Question from within her executive perspective – couldn’t she ask whether being an agential failure if she doesn’t execute some intention has any real authority? And if she found a positive answer to the Normative Question, couldn’t she draw on it to bolster her commitment to acting as she intends? If so, isn’t answering the Normative Question a priority if we are to develop an account of the first-personal psychology of intention – an account of the first-personal psychology of the agent’s executive perspective? I will argue that this is not so.[[17]](#footnote-17)

Suppose that an agent, S, who intends to A in C, wonders whether or why it matters if she is a failure as an agent. When she reflects on this, she steps back from and casts doubt on this view of herself – she wonders whether the critical reflexive attitude is warranted. If it is constitutive of intending that one accepts that one is a failure or success as an agent in certain circumstances, S seems to at least temporarily abandon or bracket the mindset of the agent who intends to A in C. Given this, we have reason to think that the Normative Question is not one that can arise and be answered from within the executive mindset of the agent – the mindset she is in when she has adopted an intention and is ready to act on it when circumstances demand it.

The Self-Evaluation view starts from the observation that the agent in the executive mindset has must-, need to-, have to- thoughts that flow from her intentions. In having these thoughts the agent seems to accept as authoritative *strict* requirements on her to act. Her acceptance of these requirements as authoritative is not, that is, *conditional* on her valuing her success as an agent, nor is it conditional on anything else. If we characterize the executive mindset as one in which the agent can address and resolve the Normative Question, we assume that the executive mindset is one in which the planning agent regards the authority of requirements as conditional. On this view, the agent in the executive mindset is open to the idea that she would not fail if she didn’t comply with her must-thoughts, and that the normative authority of her intentions is conditional on certain other things. But this characterization does not honour the peremptory nature of must-thoughts – when the agent has the thought “I must A”, she thereby grants that she will be a failure if she does not A. The executive mindset of the planning agent is not even clearly a *deliberative* mindset - the agent does not weigh reasons that speak conditionally for or against the value of being a success as an agent or for or against intended courses of action. Because pursuing the Normative Question requires *considering whether* requirements are authoritative and what their authority is *conditional* upon, it requires a departure from the executive mindset, which is characterized by *experiencing requirements as authoritative and strict*.

For a possible parallel, consider that occupying the role of parent or soldier or teacher arguably does not require that the agent takes being a good parent or soldier or teacher as a reason to act in accordance with the requirements of the role.[[18]](#footnote-18) It is not implausible to suppose that agents may embrace or identify with such roles just by accepting the requirements of the role as strict requirements on them to act, and by accepting that they would fail were they not to comply with those requirements (where it remains within their power, etc.). Similarly, it seems that one can occupy the executive role of the planning agent by entering a mindset where requirements for action that flow from one’s intentions are regarded as authoritative and strict. As granted earlier, answering the Normative Question is extremely important for a complete theory of planning agency. A planning agent may address this question herself, thereby reflecting on herself qua rational planning agent. And if she successfully answers the Normative Question, she thereby comes to a richer understanding of her own agency and its ordinary working. And this may even aid her success as a rational planning agent. She may benefit from enhanced capacities for self-control or autonomy, say, by being able to remind herself of the value of commitment or achievement or self-constitution over and above the value of particular outcomes. Nevertheless, this kind of reflective understanding can only serve as a psychological backdrop for the executive mindset. Although she can draw on it in moments of doubt or fear or confusion, it is not a proper part of the executive mindset – it does not provide her with a reason that can be appreciated from within that mindset for executing all of her intentions or some particular intention.[[19]](#footnote-19)

An objector may worry that the picture of planning agency being presented here is one in which the agent occupying the executive mindset is implausibly insensitive to questions about the normative and rational status of her intentions. But my limited claim is that when the agent is in the executive mindset - when she has already undertaken a task and is ready to execute it - she has put questions like the Normative Question firmly aside. But it is obvious that a planning agent can step aside from the executive mindset to engage in such enquiries. So, the planning agent is not implausibly insulated from the Normative Question.

It should be noted that it is widely accepted that planning agency allows for psychological compartmentalization in which the agent is freed from the burdensome task of deliberating about ends, or deliberating about what, more generally is good or right or valuable. This compartmentalization facilitates the swift and skillful performance of actions and it frees the agent to look for appropriate means and to execute her plans. The concern of the objector arguably rests on the undefended assumption that there is uniformity throughout all facets of planning agency – the agent is always ready to deliberate about ends, and to review and be sensitive to what is good or right, and so on. But if the objector also accepts that there is compartmentalization in planning agency, it must be explained how this is possible on a uniformly “intellectual” picture of the planning agent.

Let us turn to a different kind of worry about the Self-Evaluation view. It is assumed here that it is constitutive of intentions, even akratic intentions, that an agent feels bound by the practical demands that flow from the intention. The Self-Evaluation view tries to capture the agent’s feeling boundin terms of her considering herself a failure as an agent if she does not abide by the practical demands of her intention. But isn’t it implausible to suppose that the akratic agent regards herself as a failure as an agent for not abiding by an intention that she knows she shouldn’t have? If this is implausible, then we should not spell out the subjective authority of intention in terms of regarding oneself as a failure if one doesn’t execute.

This is a difficult issue and a complete reply is not possible because it requires a substantive treatment of akrasia. But it is worth noting some related points that help to show that the Self-Evaluation view is not as seriously challenged by this worry as it may first appear. First, the kind of failure or success that is involved in intention is failure as an agent, rather than failure as, say, a friend or soldier or neighbour. As the agent can pull conflicting standards apart in thought, the akratic agent may be particularly sensitive to her failure or success as an agent while ignoring the fact that she is thereby a failure as a friend. By focusing on issues of execution and ignoring issues concerning things that she values, this may facilitate the otherwise unlikely transition from believing that she should not A to forming the intention to A, and the otherwise unlikely willingness to accede to the authority of the task of A-ing in deliberation and action. Second, as mentioned earlier, intention is widely regarded as facilitating a kind of psychological compartmentalization: it allows the agent to set aside the arduous process of deliberation about ends, so that she can find means and focus on the swift and skillful performance of actions. As the search for means may be arduous, and as executing action-plans may be challenging and absorbing, this further facilitates the focus of the agent’s attention on the standards by reference to which she is an agential failure or success, rather than on other conflicting standards that she holds herself to. Rather than being refuted by akrasia cases, I am inclined to think that the Self-Evaluation view may help us to better understand them. Finally, an answer to the Normative Question might allow us to explain why the felt authority of an akratic intention, unlike a non-akratic one, lacks objective authority. If the normative authority of intentions is bound up with self-government or self-constitution, say, then the akratic intention lacks authority to the extent that akrasia undermines these things. But as we have seen, the executive mindset enjoys a kind of psychic insulation from such facts, and this facilitates the agent’s prior motivation to ignore the poor normative status of her intention. Although there is no way to simply and briefly answer the challenge, it does not, I think, provide compelling reasons to abandon the Self-Evaluation view.

It might be objected that simple and silly intentions pose a problem for the Self-Evaluation view. Quite arbitrarily formed intentions to have lunch in an hour or simple intentions to go out to buy milk in a few minutes or silly intentions to jump up and down for no reason may play little or no role in co-ordination. Such intentions don’t seem to involve taking a complex evaluative attitude to oneself, nor do they seem to involve regarding oneself as under a requirement to take compensatory measures should one not execute the intention in the absence of excusing conditions, and so on.

It should be noted that even if a simple or silly or spontaneously formed intention never plays a co-ordinating role, if it is an intention as that is understood here, then it is poised to do so. Consequently, it must be able to play a role in explicit planning, such as constraining further deliberation about conflicting ends, and guiding action. As long as simple and silly intentions are understood in this way, they must incorporate some mechanisms that are appropriate for playing such roles. Given this, either these “intentions” do not have these characteristic features of intentions, and there is no challenge to the Self-Evaluation view, or they are not all that simple, appearances to the contrary, and they do incorporate some mechanism by which they play the characteristic roles of intention. But, if the latter, is it so implausible that the Self-Evaluation view can give an account of how they would play their characteristic roles? And is the Self-Evaluation view describing a prohibitively more complex mechanism of control and co-ordination than an alternative view might provide? The objection relies on the assumption that it is clear that the Self-Evaluation view is too complex but this is a difficult claim to vindicate. An adequate theory of intention must capture the psychology of what is a complex state, and so, we should be wary of “too-much-complexity” objections.[[20]](#footnote-20)

**Final Remarks**

It may seem odd that the subjective authority of an intention does not stem from the reasons that the agent thinks that she has for acting in the way that she intends. And it may seem equally odd that such felt authority cannot be directly shaped by reflection on the value of things such as self-trust or self-government. But the appeal of these thoughts is illusory. In the hustle and bustle of a busy day, where what the agent intends must come to the fore to control conduct, and must do so swiftly and without inordinately taxing her, it is too difficult to review or be sensitive to the reasons that favour the intended action. And it is also too difficult to reflect on the large question of what normative status intentions may have. The agent needs to think peremptory must-, need to-, have to-thoughts if she is to be swift, skillful, effective in action. It makes sense that she needs to enter a mindset where she simply grants that she would be a failure if she doesn’t comply with the demands expressed by these thoughts.

Given the exigencies of execution, it is not unreasonable to think that there is a rather strict division of labour between, roughly put, non-executive and executive dimensions of planning agency. From this vantage point, the thought that the subjective authority of intentions and their associated must-thoughts derives either from the appreciation of practical reasons favouring the intended course of action or from the appreciation of things of value, such as co-ordination or self-trust, arguably runs two distinct facets of planning agency together: deliberating about ends or reflecting on what is good or valuable, on one hand, and undertaking and executing tasks, on the other. If the arguments of the paper are on track, the former activities will not explain why the agent experiences her must-thoughts as authoritative about what she must do. But then again, if the subjective authority of intention were rooted in such cumbersome activity, it would be hard to make sense of the peremptory nature of our must-thoughts on those very hectic days.[[21]](#footnote-21)

**References**

Bradford, G. (2015) *Achievement*. New York: Oxford University Press.

Bratman, M. (1989) *Intention, Plans, and Practical Reasoning*. Stanford: CSLI Publications.

\_\_\_\_\_ (2007) Structures of Agency. New York: Oxford University Press.

\_\_\_\_\_ (2012) ‘Time, Rationality, and Self-Governance’, *Philosophical Issues*,22/1: 73–88.

\_\_\_\_\_ (2014) ‘Temptation and the Agent’s Standpoint’, *Inquiry* 57: 293-310.

Broome, J. (2001) ‘Are Intentions Reasons and How should we cope with Incommensurable Values?’ in C. Morris and A. Ripstein (eds.) *Practical Rationality and Preference*, 98–120. Cambridge University Press.

Enoch, D. (2006) ‘Agency Schmagency: Why Normativity won’t Come from what is Constitutive of Action’, *The Philosophical Review*, 115/2: 169–198.

Ferrero, L. (2010) ‘Decisions, Diachronic Autonomy, and the Division of Deliberative Labor’, *Philosophers’ Imprint*,10/2: 1–23.

Hieronymi, P. (2009) ‘The Will as Reason’, *Philosophical Perspectives*, 23: 201–220.

Hinchman, E. (2003) ‘Trust and Diachronic Agency’, *Noûs*,37/1: 25–51.

Kiesewetter, B. (2017) *The Normativity of Rationality*. Oxford: Oxford University Press.

Kolodny, N. (2005) ‘Why be rational?’, *Mind*, 114/455: 509–563.

Korsgaard, C. (2008) *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. New York: Oxford University Press.

\_\_\_\_\_\_ (2009) *Self-Constitution: Agency, Identity, and Integrity*. New York: Oxford University Press.

Levy, Y. (2018) ‘Why Cognitivism?’, *Canadian Journal of Philosophy*, 48/2: 223–244.

Marušić, B. and J. Schwenkler (2018) ‘Intending is Believing: A Defense of Strong Cognitivism’, *Analytic Philosophy*, 59/3: 309–340.

Mele, A. (1995) *Autonomous Agents: From Self-Control to Autonomy.* New York: Oxford University Press.

Nefsky, J. and Tenenbaum, S. Unpublished MS. “Extended Agency and the Problem of Diachronic Autonomy”.

Paul, S. (2014) ‘Diachronic Incontinence is a Problem in Moral Philosophy’, *Inquiry,* 57/3:337–355.

Setiya, K. (2007) *Reasons without Rationalism.* Princeton: Princeton University Press.

Smith, M. N. (2015) “One Dogma of Philosophy of Action”, *Philosophical Studies*, 173/8:2249–2266.

Sosa, E. (2007) *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. New York: Oxford University Press.

Tenenbaum, S. (2014) ‘Minimalism about Intention: A Modest Defense’, *Inquiry,* 57/ 3:384-411.

\_\_\_\_ (2016) ‘Re-Considering Intentions’, Noûs, 52: 443-472.

Velleman, D. (1989) *Practical Reflection*. Princeton: Princeton University Press.

Williams, B. (1981) “Persons, Character, and Morality.” In *Moral Luck*, 1-19. Cambridge: Cambridge University Press.

1. The seminal discussion of intention as a state that exerts “rational pressure” is in Bratman (1989). How I frame the issues here is deeply indebted to Bratman’s work. [↑](#footnote-ref-1)
2. I assume here that intentions are not like lit fuses, but exert pressure via conscious rational thought. See the discussion of Broome’s (2001) view in Ferrero (2010), 5–6. [↑](#footnote-ref-2)
3. There is a large and growing literature on the rational or normative status of intention – some notable recent discussions include Bratman (2012), Ferrero (2010), Paul (2014), Tenenbaum (2014, 2016), Nefsky and Tenenbaum (unpublished manuscript). [↑](#footnote-ref-3)
4. See Bratman (1989) for a seminal discussion of bootstrapping and Smith (2015) for a review of relevant literature and discussion. [↑](#footnote-ref-4)
5. In the literature on the rationality of acting on intentions, Edward Hinchman (2003) and Luca Ferrero (2010) have developed views that appeal in different ways to an agent’s trust in earlier deliberation. [↑](#footnote-ref-5)
6. This kind of case also poses a problem for Reasons. [↑](#footnote-ref-6)
7. For the seminal discussion of intention’s role in co-ordination, see Bratman (1989). [↑](#footnote-ref-7)
8. This point is made eloquently in Bratman (1989). [↑](#footnote-ref-8)
9. See Velleman (1989) and Setiya (2007) for “cognitivist” views of intention. A more recent defense of a cognitivist view can be found in Schwenkler and Marušić (2018) and for a recent discussion of cognitivism, see Levy (2018). [↑](#footnote-ref-9)
10. See Paul (2014) for interesting treatment of the role of self-trust in the normative authority of intention. [↑](#footnote-ref-10)
11. Thanks to Antti Kauppinen for pressing this point. [↑](#footnote-ref-11)
12. Mele (1995) makes a closely related distinction between evaluative and executive commitments, pp. 71-73. [↑](#footnote-ref-12)
13. Although Hieronymi is not exclusively concerned with distinctively practical standards – standards stemming from action plans - as I will be here, and although her focus is not exclusively on the agent’s view of herself qua agent, she makes a related point when she says “… your decisions, intentions, and intentional actions open you, as a person, to certain characteristic sorts of evaluations. Your accepting of the mission might show you courageous, clever, malicious, dastardly, or vengeful.” P. 204, 2009 [↑](#footnote-ref-13)
14. See Bratman (2012, 2014) and for recent critical discussion, Kiesewetter (2017). [↑](#footnote-ref-14)
15. See for examples, Sosa (2007), Bradford (2015). [↑](#footnote-ref-15)
16. See Korsgaard (2008, 2009), and for critical discussion, Enoch (2006), Kolodny (2005), Kiesewetter (2017). [↑](#footnote-ref-16)
17. My thanks to an anonymous reviewer for this journal for pressing these questions. [↑](#footnote-ref-17)
18. And it might be that it is inimical to occupying certain roles that one consider and act on such reasons - this would be “one thought too many” Williams, p. 18, 1981. [↑](#footnote-ref-18)
19. Bratman maintains that although a course of action, C, may be rationally required by an intention (a “framework reason”), it doesn’t follow that the agent has sufficient reason to pursue C. (Bratman, 1989, 33-5; Bratman, 2007, 287-91). The agent may, however, take an “external” perspective on her intentions and question the reasonableness of the intention that led to the adoption of C, or even the value of being a planning agent, and so on. Bratman’s reference to an “external” perspective resonates to some extent with what I say about the need to bracket one’s intention if one is to address the Normative Question. But I do not think that he agrees with the extent of the isolation of the subjective authority of intention from such “external” evaluation that I accept here. A full discussion goes beyond my scope here. [↑](#footnote-ref-19)
20. It should be borne in mind that not just any rational agent counts as a planning agent, one who forms and is capable of hewing to intentions as they are understood here. I am also not concerned with psychological states that may be called intentions in existing literature, such as motor intentions – I am only concerned with states that conform to SA in Section 1. [↑](#footnote-ref-20)
21. This paper was made possible through the support of a generous grant from the John Templeton Foundation. The opinions expressed are my own and do not necessarily reflect the views of the John Templeton Foundation. I am grateful to Jaegwon Kim for guiding earlier work on this topic. Special thanks are due to Antti Kauppinen for discussions and for reading earlier drafts. I am very grateful to Philip Clark and Al Mele for giving insightful feedback on earlier drafts and to Facundo Alonso for serving as my commentator at SLACRR. Thanks to N. Emrah Aydinonat, Olle Blomberg, Michael Bratman, Vittorio Bufacchi, Stephen Butterfill, Jason D’Cruz, Mary Edwards, Marion Godman, Natalie Gold, Thor Grünbaum, Raul Hakli, Kathryn Lindeman, Uskali Mäki, Caterina Marchionni, Cara Nine, Julia Nefsky, Krisztina Orbán, Sarah Paul, Glenda Satne, Alessandro Salice, John Schwenkler, Thomas Smith, Ninni Suni, Till Vierkant, Joel Walmsley, Eric Wiland, and Hong Yu Wong for engaging with the ideas and giving helpful feedback. Thanks to audiences at Florida State University, SUNY-Albany, St. Louis Annual Conference on Reasons and Rationality (SLACRR), King’s College London, the Cork Annual Workshop on Social Agency (CAWSA), University of Copenhagen, University of St. Andrews, and TINT at the University of Helsinki. I am very grateful to my colleagues at University College Cork for feedback and for facilitating research leave that allowed me to complete this paper. The paper benefitted a lot from the insightful comments of two anonymous reviewers for this journal. [↑](#footnote-ref-21)