Abstract: Cognitivism or doxasticism (henceforth cognitivism) about intentions is the view that some agent's intention to Φ entails her belief that she will Φ, where Φ stands for the action the agent intends to perform. This paper argues that noncognitivism about intentions is [1] implausible and makes a positive case for cognitivism based mainly on the claim that the formation of intentions must be directly voluntary.

Keywords: Cognitivism • Doxasticism • Intentions • Action Theory

[Title Redacted]

Strong cognitivism about intentions is the view that an agent's belief that she will Φ is a proper part of the structure of her intention to Φ, where Φ stands for the action that the agent intends to perform. In this paper, I make a negative argument for strong cognitivism by rejecting potential motivations for the adoption of competing views. First, I rule out the possibility of noncognitivism, the view that an agent's intention to Φ does not entail her belief that she will Φ, by raising objections to proposed noncognitivist views. Next, I make a case against weak [2] cognitivism, which maintains that an agent's intention to Φ entails her belief that she will Φ, but allows that the beliefs entailed by agents' intentions are mental states that are distinct from the intentions that entail them. To do so, I first show that the formation of intentions must be subject to the direct voluntary control of the agents who form them if intentions really play the functional roles that we grant them as mental states. Second, I argue that if the formation of intentions is directly voluntary, then if an agent's belief that she will Φ is entailed by her intention to Φ, then that belief must be a proper part of her intention. Because the formation of intentions is directly voluntary, then, we should be strong cognitivists about intentions if we are cognitivists at all.

I. Noncognitivism about Intentions

This section argues against noncognitivists, or those who hold that an agent's intention to Φ does not entail her belief that she will Φ. One barrier to the evaluation of noncognitivist views [3] is that few have been fully presented, although several writers have defended noncognitivism negatively by rejecting arguments for cognitivism. I address two noncognitivist arguments against cognitivism here: one from a rejection of the alleged parity of expression between beliefs and intentions and another from considerations about whether a statement of intention can generate a sentence of the kind that features in a Moorean paradox.[4]

[1] Defenders of the cognitivist or doxasticist view include: Anscombe 1957; Clark 2020; Davis 1997; Harman 1997; Marušić 2012; Paul 2012; Setiya 2007, 2014; Velleman 2000.

[2] Namely, I address the noncognitivist proposals offered in Levy 2017.

[3] Clark 2020 identifies Bratman 1987, 2009 and Paul 2009 as noncognitivists about intention. Levy 2017 apparently endorses a noncognitivist view about intentions, or at least rejects a cognitivist view.

[4] Two such arguments can be found in Levy 2017.

One datum that has been cited in defense of cognitivism since Anscombe is the point that there is a parity of expression between beliefs and intentions. Beliefs about one's future actions are canonically expressed by sentences of the form <I will Φ,> where Φ stands for the action one intends to perform. Cognitivists argue that it speaks in favor of their view that canonical [5] expressions of intentions are likewise accomplished via sentences of the same form: i.e. sentences of the form <I will Φ.> Noncognitivists about intentions are therefore hard pressed to come up with an alternative explanation for the parity of expression between beliefs and intentions that does not invoke an entailment from one's intentions to perform actions to one's beliefs about the actions one will perform.

Levy 2017 offers two potential noncognitivist explanations for this linguistic fact, neither of which, I argue, should count as a refutation of cognitivism about intentions. The first is expressivism about avowals. Avowals are a certain class of sentences that make first-person mentalistic attributions. Expressivists about avowals hold that these sentences are direct expressions rather than assertoric reports of the mental states they concern. Expressivists often also hold that because of the particular way in which avowals are non-assertoric, they are non truth evaluable. Their function in our language is not to represent a statement about the world, or even something nonexistent, as true or false, but to be an outward indication of one's mental life in the same way as an exclamation, a command, or an audible response to pain.

According to avowal expressivists, statements of the form <I will Φ> count as avowals when such statements are used to directly express intentions. This explanation of how <I will Φ> functions as a canonical expression of one's intentions allows the noncognitivist to offer an underlying explanation for expressions of intention via the sentential form <I will Φ> that does not put these sentences on par, at least semantically, with assertions of beliefs about what one will do. Such statements, when used to express intentions, do not represent as true a statement about the future, but are rather a conventional way of using language to outwardly indicate the mental state one is in when one has an intention to Φ.

Note that the account proposed above offers no explanation for the admitted syntactic parity of expression between intentions and beliefs about one's future actions. The account in Levy 2017 also does not offer any explanation of why we should think that a semantic dissimilarity exists between statements of beliefs about one's future actions and expressions of one's intentions, notwithstanding the syntactic parity between these two kinds of utterances. For this reason, I will proceed to evaluate the plausibility of avowal expressivism about statements of intentions.

It may help in this project to consider whether some of the motivations for adopting expressivism about other domains should be thought to apply to the case of statements expressing intentions. I consider two of the considerations that have been cited as rationales for

[5] Notably, there have been some dissenters on this point, and Anscombe raised doubts about the parity of expression of beliefs and intentions using her directions of fit metaphor. Thanks to [anonymous] for

highlighting this.

expressivism about our moral discourse. The first is that moral language is imperative or recommending rather than descriptive, and the second is that moral judgements are non assertoric when an utterer of a moral claim intends to convey that she holds some noncognitive attitude by her utterance. I argue that neither of these considerations suggest that statements of intentions expressed via sentences of the form <I will Φ> should be understood as expressions of noncognitive attitudes rather than beliefs about what one will do.

The first motivation for expressivism I wish to consider is the imperatival motivation. Some expressivists about morality consider judgments of actions as right or wrong to be commands or exhortations to behave in certain ways rather than others. It seems like we may [6] abandon this proposal as it concerns statements of intentions, as an utterance expressing an intention via a sentence of the form <I will Φ> is neither obviously a command nor a recommendation to oneself or others to perform the action one intends to perform. Arguing that a statement of intention has this function would be an interpretive stretch.

Consider whom the addressee of the command would have to be if statements of intention were conceived as commands or recommendations. On the most plausible imperatival expressivist account of statements of intentions, a statement of some intention of the form <I will Φ> would express a command to oneself to Φ. Because the commands on such an account would be self-directed, an imperatival approach to expressivism about statements of intentions would raise the question of why one would need to utter the command to oneself aloud or state it in the presence of others. The addressee could not obviously be interpreted as a different audience, either, as one does not always judge the actions one intends to perform to be the best actions for others to perform.

Second, expressivists about morality sometimes describe their view in terms of the communicative intent of moral statements. According to these expressivists, people make moral judgments expressing noncognitive attitudes with the intention of conveying to others that they hold those noncognitive attitudes. The question to raise about the analogue of this kind of [7] expressivist account for statements of intention is whether a statement of a prospective intention can really be analyzed as the intended expression of a noncognitive attitude. It seems rational to hold accountable someone who says “I will meet you at 1:00” to express an intention to meet you at 1:00 if they do not succeed at meeting you then. It is further unclear how a speaker could [8] rationally intend to convey a noncognitive attitude of hers by uttering a sentence qualitatively identical to one that she could utter in the same context to convey a prediction about her future action. If one’s intending to Φ requires one to have some degree of confidence that one will Φ (or even some degree of confidence that Φ-ing is possible), an intention to convey a noncognitive attitude by the kind of statement just described should not count as a genuine intention.

[6] Stevenson 1937.

Schroeder 2010. [7]

Thanks to [anonymous]. [8]

It is therefore clear from these remarks that the kinds of considerations that motivate expressivism about moral discourse do not apply in the case of statements of intentions. Another proposal introduced by Levy is that a belief that one will Φ is conversationally implicated, rather than entailed by one's expression of an intention to Φ via a statement of the form <I will Φ.> Do we have any reason to think that the speaker's belief that she will Φ is conversationally implicated when she utters a statement of the form <I will Φ>?

If we think of implicating something beyond the literal meaning of a sentence as an illocutionary speech act, the claim that a statement of an intention to Φ via an utterance of the sentence <I will Φ> implicates, but does not literally mean, that one believes one will Φ nonetheless implies that one intends by uttering that sentence to convey what is implicated as part of its meaning. Of the view that a belief that one will Φ is conversationally implicated by a [9] statement of intention, then, we may ask two things: (1) how we should think of the literal semantic content of the sentence implicating the belief and (2) why the explanation of the speaker's intent in implicating the belief would not involve an appeal to the claim that the speaker actually holds the belief. It is not clear that a satisfactory answer to either of these questions exists.

From what I have said, it seems like we have little reason to accept the argument against cognitivism from the rejection of the alleged parity of expression between statements of intention and statements of belief about the future. In any case, it would not be clear if this argument succeeded that cognitivism about intentions should stand or fall with the parity of expression between statements of intention and those of belief. Even if the canonical statement of a prospective intention had a different sentential form, this would not speak in favor of or against cognitivism as a metaphysical thesis about what is involved in intending. As a general rule, "B is entailed by A implies that expressions of A are semantically and syntactically on par with expressions of B" does not hold.

The second argument for noncognitivism about intentions that I will consider is motivated by the claim that "I intend to Φ, but I do not believe that I will Φ" is not a sentence of the kind that would generate a Moorean paradox. The original Moorean paradox involves the claim that the sentence "It is raining, but I do not believe it is raining" sounds paradoxical despite the fact that the sentence does not state an overt contradiction. People have appealed to various pragmatic norms to explain the paradox at the level of speech, as well as analogies between propositional thought and assertion to oneself to explain the paradox at the level of thought.

The argument for noncognitivism from the claim that "I intend to Φ, but I do not believe that I will Φ" fails to generate a Moorean paradox is unconvincing precisely because it begs the question. That is, it presupposes the conclusion that an intention to Φ does not entail a belief that one will Φ. If one accepted that an intention to Φ entailed a belief that one would Φ, the sentence

in question would arguably generate a paradox of the kind that is characteristic of the Moorean cases.

Austin 1962. [9]

II. The Formation of Intentions is Directly Voluntary

We have just seen the implausibility of taking a noncognitivist approach to the structure of intentions. In the last section, I argued that two of the major points cited in favor of noncognitivism are not cogent. These include the endorsement of expressivism about statements of intention and the denial that one's statement of intention conjoined with the statement that one does not believe that one will do what one intends should count as a Moorean paradox.

Therefore, it seems that the view that one's intention to Φ entails one's belief that one will Φ is comparatively more plausible than the alternative view. Indeed, many mainstream accounts of intentions have been committed to some kind of belief constraint on intentions involving the belief that one will act as one intends. However, we may still ask in virtue of what one's intention to perform an action should entail a belief that one will perform it. [10]

Writers on intentions have traditionally located the need for a belief constraint on what counts as an intention in considerations that do not have to do with the structure of intentions.[11] They have sometimes gone on to give more and less specific answers to the question of how it is that one forms the belief that is entailed by one's intention. There is one main division in the types of answers that have been given to this latter question, and the two opposing positions to which it has given rise have been discussed under the headings of strong cognitivism and weak cognitivism.

Whereas ***strong cognitivism*** says that the belief that one will do what one intends is a proper part of whatever mental phenomenon we are calling an intention, ***weak cognitivism*** is the position that the belief involved in an intention is merely entailed by, but is not a part of, the structure of the intention. Defenders of the strong cognitivist view include Harman 1997 and Hieronymi 2006, who hold that an intention consists, to some extent, in a belief about what one will do. Defenses of weak cognitivism can be found in Davis 1997 and Grice 1971, which argue that the belief entailed by the intention is formed on the basis of one's awareness of the conative attitude that comprises the intention.

Some writers, including Paul 2009 and Clark 2020, have utilized a further distinction dividing weak cognitivist views into ones on which the belief that one will do what one intends is formed inferentially on the basis of the mental structure that comprises the intention (***inferential weak cognitivism***) and ones on which the belief is not formed via such an inference (***noninferential weak cognitivism***).

To proceed with my argument for the strong cognitivist thesis, I show that the formation of intentions must be directly voluntary in two ways. First, I argue that this point is supported by

Davis 1997; Paul 2012; Harman 1997; Hieronymi 2006. [10]

The rationales for these constraints have included the unification of theoretical and practical reasoning [11] and the syntactic parity of expression between statements of intentions and those of beliefs about one's future actions.

the widely noted point that what is distinctive of intentional actions that are performed from intentions is our direct voluntary control over over them in contradistinction to our apparent lack of direct voluntary control over our attitudes like beliefs and emotions. I claim that it would be [12] hard to preserve this explanation of what is distinctive about the performance of intentional actions performed from intentions if we were to grant that the formation of intentions is not directly voluntary. Second, I argue that if we were to reject the claim that the formation of intentions is directly voluntary, we would have no means for distinguishing volitive from appetitive desires in cases where doing this would be salient to the discrimination of an intention from a mere prediction inferred from the presence of a non-rational desire. For both of these reasons, giving intentions the explanatory power we often presume them to have requires granting that their formation is directly voluntary on the part of the agents who form them.

We apparently have direct voluntary control over the intentional actions of ours that are formed from intentions in a way we do not over our formation of attitudes like beliefs and emotions. On a traditional picture, part of the reason why we have such control over our [13] intentional actions is because their direction of fit is from mind to world. Whereas beliefs are sensitive to evidential inputs and emotions are reactive attitudes, actions can be productive. They are supposed to be our means of shaping the world to our will. We get a preliminary sense from these remarks that, while attitudes like beliefs and emotions, by definition, comprise a kind of programmatic range of responses to our awareness of fixed domains of facts about the world, no straightforward function can be defined in a non-arbitrary way from inputs consisting of facts about the world (or even facts about the world of which agents are aware) before the formation of an intention to perform some action to outputs conceived as the performance of particular actions, or even instances of tryings to perform them.

If intentions were formed involuntarily and according to non-probabilistic laws of nature, I claim that we would be able to define this kind of straightforward function from facts about the world prior to agents' formation of intentions to instances of tryings to perform particular actions from those intentions. We may understand a trying as a mental event that causes the event of the

Desires are complicated attitudes to include in such generalizations about the formation of attitudes, [12] because, as we will later see, there may be multiple kinds of desires that are formed in different ways.

One might point out that intentional actions need only arise from intentionalistic elements, and not from [13] intentions. However, my paper is concerned with the set of actions performed from intentions, which is a subset of the set of intentional actions. Answering the question of whether we should think of intentional actions that are not performed from an intention as voluntary in the sense I discuss here is beyond the scope of this project.

action in the case where it succeeds. The definition of this kind of function would be [141516] possible because in a world where the formation of intentions were involuntary, their formation would have to be explicable in terms of the laws of nature that held in that world.

Moreover, a view on which intentions were involuntarily formed according to strict universal laws would imply that one could not change one's mind about what one intended without a change in the facts about the external world. However, it seems like one should be able to change one's mind about what one intends without any changes being made to the relevant facts about the world between the formation of one's initial intention and the changing of one's mind. It also does not seem like we can define a non-arbitrary and straightforward function [17] from facts about the world prior to agents' formation of intentions to instances of tryings to perform particular actions. Therefore, it cannot be the case that intentions are formed involuntarily and according to non-probabilistic laws.

If intentions were formed involuntarily and according to probabilistic laws of nature, we would be able to define a straightforward function from facts about the world prior to agents' formation of intentions to probabilities of occurrences of tryings. Again, this is the case because if involuntary, the formation of intentions in a given world would have to occur in conformity with the laws of nature of that world. However, it is not clear that we can define a function from facts about some world prior to agents' formation of intentions there to probabilities of occurrences of tryings, either. It is also not clear what would count as evidence for the predictive accuracy of such a function if we were to try to formulate one. We can therefore reject this approach as implausible.

According to the only available alternative on which intentions are formed involuntarily, then, they are formed involuntarily and completely randomly, or in a way that bears no obvious

relation to the facts prior to their formation. However, this does not seem right either. Agents' beliefs, desires, and goals all seem to correlate with their formation of intentions to some degree, albeit not a straightforward statistical one. Having now ruled out the logically possible ways in

This definition of a trying is a disjunction of those of O'Shaughnessy, Adams, and Mele (Mele 1997, [14] 15).

Tryings rather than actions themselves would be the output of this function because one's mental trying [15] to perform a particular action does not have to lead to one's performance of that action (e.g. in cases where one's wide abilities are in some way compromised).

Strictly speaking, it does not matter whether we allow for the possibility of one such function or many. [16] The point is that determinate pairings of facts about the world prior to the formation of intentions with intentions formed should be possible for the formation of each intention on an account on which the formation of intentions is involuntary. Thanks to [anonymous] for pressing this point.

One could counter that this would be an implication of any view of the formation of intentions, since [17] facts about the passage of time are facts about the external world whose truth makes possible the changing of one's mind. However, the relevant facts about the world I am speaking of here are those facts that would have a causal influence on the production of an agent's motivational state, which facts about the progress of time arguably are not.

which intentions might be involuntarily formed, we may conclude that the formation of intentions is voluntary.

My next argument for the claim that the formation of intentions must be voluntary is based on one of the explanatory virtues of allowing that intentions to act may only involve conative states that are voluntarily formed by agents. Scholars since Plato have distinguished [18] between appetitive desires and rational ones, and such distinctions have often given normative authority to second order desires in determining which desires should count as rational. In other words, many of the accounts that draw the distinction in question have in some form or another involved the normative claim that we should desire to have rational desires.[19]

Indeed, talk of "rational desires" has often appeared in discussions of moral motivation, since both a cognitive and conative component are required for action on the basis of practical reasons according to a widely adopted Humean theory of practical rationality. In the context of this kind of discussion, figures including Kant and Aristotle have arguably relied on a distinction between appetitive and rational desires in order to distinguish between a kind of non-moral motivation that is merely coincidentally also motivation to perform a right action and a kind of distinctively moral motivation that should genuinely count as motivation to do the right thing.[20] For the purposes of this paper, we can think of a rational desire as any first order desire that is consistent with our second order desires. In other words, we can think of a rational desire as the kind of desire we would want to have given our values, projects, and goals. An appetitive desire can be rational insofar as it fits this description. Notably, this picture assumes that our second order desires—i.e. the instrumental desires we in point of fact want to have given our projects,

values and goals—are generally our best epistemic guide to the instrumental desires we would want to have given our projects, values and goals. I am understanding an ***instrumental desire*** [2122] to be a desire that one has as a means to the attainment of one's ***intrinsic desires***, which are desires that are determined by the relevant projects, values and goals.

Davis 1997. [18]

Davis 1997 does not say this, but this claim is suggested by passages in Gibbard 1990 and Railton [19] 1986.

Aristotle 340 BC; Kant 1785; Audi 1995; Korsgaard 2008. [20]

Thanks to [anonymous] and [anonymous] for pressing this point. [21]

The distinction between having a disposition to desire and having a desire in a dispositional form may [22] be relevant here. Thanks to [anonymous] for this suggestion. A dispositional desire is a desire toward whose formation you are already disposed, whereas a disposition to desire is a desire toward whose formation you would be disposed if you encountered some kind of stimulus that disposed you to be disposed toward it, e.g. being questioned on the matter. The second order endorsement that is required for the kind of motivational state I am calling rational is better conceived in terms of a dispositional desire than a disposition to desire. We are assuming that the agent's motivational set does not change based on her occurrent consideration of what she would want to desire.

I contend that there are explanatory benefits to conceiving of an intention as the kind of mental structure that allows us to act on our rational desires thus construed, rather than on our non-rational desires, which, defined negatively, we may think of as those first-order desires we have that are not of the kind that we would want to have given our current values, projects, and goals. Other writers on intentions have picked up on the usefulness of a similar distinction in order to account for cases in which intuitions lean toward saying that the agent in question does not have an intention to undertake a particular course of action, but in which that agent predicts that she will undertake that course of action and reports a belief that she will do so inferred from her awareness of her desire to do so.

One example of this kind might involve my belief that I will buy a cookie at the supermarket based on a prediction that I will do this. I make this prediction even though I am currently on a diet because I know that I often fail to resist the temptation of buying a cookie at the supermarket, even when I am on a diet. It seems like the intuition in this case is to say that I do not have an intention to buy a cookie, even though I believe that I will do so based on a motivational state I am in. Davis 1997 would account for my not having an intention to buy a cookie in this case by saying that I have the wrong kind of desire to count as one of the kind that could feature in an intention. On this point, he appeals to a distinction between appetitive and volitive desires. Relying only on the intuition that an agent's belief-desire pair in a relevant case would not count as an intention in order to substantiate this claim, he appears to hold that only

volitive desires can play a role in the structure of intentions.

My addition to this picture is that volitive desires are what I have been calling rational desires, whereas we may think of Davis's appetitive desires as being the desires of agents to perform actions that they would not rationally want to have. Because I do not want myself to want to buy a cookie given my project of losing weight in the earlier case, my desire to buy a cookie is not a rational desire in the terminology I have introduced. I am arguing that this feature of that desire explains why we would not be justified in ascribing to me an intention to buy a cookie despite my prediction that I will do so on the basis of my awareness of my desire to buy a cookie and my past failures to resist the temptation of buying one. If intentions could involve desires of the non-rational kind, we could not provide an equally persuasive account of why it is that one would not be justified in ascribing an intention to me in this case.

Someone could press the point that agents apparently do have the ability to form intentions to act against their current values, projects, and goals, and that this alleged ability would problematize the definition of a rational desire I give here as a second order desire, at least when this definition is combined with my claim that only rational desires can feature in the structure of intentions. However, my definition of a rational desire as one that we would want ourselves to have given our current values, projects, and goals can be interpreted loosely enough to accommodate some agent's having an intention in a seemingly borderline case, although the issue may not seem completely straightforward in cases in which my opponent would say we seem to have the ability to form intentions to act in ways in which we are compelled to act.

One such type of case involves an action that one performs under compulsion. If, for instance, I were forced at gunpoint to unlock a door, my current values, projects, and goals might be such as to support the prioritization of keeping myself alive. If this were the case, there would be no trouble in saying that for the sake of remaining alive in such a situation, I would want myself to want to unlock the door. There might be an additional question of whether I would still count as having an intention to unlock the door if I valued being the kind of person who would not perform this kind of task under compulsion. The intuitive answer would probably be no, if I performed the action from a motivational state that I, at the moment when I performed it, did not want myself to be in. If I did, upon reflection, realize that I wanted myself to be in the [23] motivational state I was in when I performed the action, despite allegedly being the kind of person who valued not performing this kind of action under compulsion, this would not suggest that an intention can involve a motivational state that an agent would not want herself to be in, but rather that my values were not what I thought they were. This assessment of what goes on in cases of compulsion successfully captures the sense in which the things done from an intention are supposed to be free.

A related example from the moral domain that might be thought to problematize the account I give here can be found in Joyce 2001. The case involves someone who acts against her moral values because she has a fascination with being, by her own lights, an evil person. It might

appear that for such a case to be possible, it would have to be possible for one to form an intention involving a desire that one would not want oneself to have. My reply to this is that it's not clear in this kind of case that the character involved would not want herself to want to perform actions she considered evil. Assessing whether she would or would not want to have a desire to perform actions she considered to be evil would require more information about her psyche and the precise nature of her fascination with becoming an evil person.

There is a different question in this case about whether the agent would rationally endorse her second order desire to act against her values based on some of model of means-ends rationality, and given her motivational set. It seems like she would have to rationally endorse her second order desire if, in her ranking of values, being an evil person according to her own value system were ranked higher than the other values of hers. The answer to the question of whether an agent would have to rationally endorse her second order desire in order for her first order desire to be rational is tied up with the question of whether, in order for some agent's first order

Whether or not I unlocked the door out of fear would not bear on whether I would endorse the [23]
motivational state I was in when I unlocked it in a higher order way. First of all, it is unclear whether we should think of an emotion like fear as being the kind of thing that can ground a normative reason for action, defined as a reason that could potentially motivate one to act in a certain way. It seems like an action performed from fear could be conceived as an action performed from a neural disposition rather than an action performed from an intention of the kind that would play a role in means-ends rationality.

However, if we assume that practical deliberation based on fear is possible, then, if, given my motivational set at the time of my action, I would have placed being the kind of person who would not unlock the door under compulsion higher in a ranking of values than being the kind of person who would unlock it out of fear, my desire to unlock the door out of fear at the time when I hypothetically did would not be a rational desire on my account.

motivational state to be considered rational, her higher order endorsement must penetrate to her second order desire, and her third order desire, and her fourth order desire, *ad infinitum*.

My reply is that an agent's higher order endorsement automatically penetrates to her higher order desires if we hold that forming an intention for an agent is a basic action of hers that proceeds via her recognition that performing a particular action would be consistent with one of her values, projects or goals, which are the items in her motivational set that fix her second order instrumental desires to perform certain actions. An agent forms a rational first order desire to act in a certain way when she performs the basic action of forming an intention to act in a way that would help her further her projects and attain her goals. Because any intention of hers to perform such a basic action of intending would be trivially consistent with the same motivational set, all of the desires involved in such higher order intentions would be rational.

The logical relationship between the voluntary formation of intentions and the distinction between rational and non-rational desires remains in need of explanation. It might seem like an open question how the claim that the formation of some agent's first order desire is voluntary follows from the fact that the first order desire in question is the object of some second order

desire of hers. However, it arguably does follow from this fact that the first order desire in question was not formed against the will of the person who formed it, as she holds a second order attitude endorsing its possession. If it follows from the fact that the desire was not involuntarily formed by the agent in question that it was formed voluntarily, I have shown that rational desires are formed voluntarily.

III. A Belief that One Will Do What One Intends is a Proper Part of an Intention

In the final section, I argue that (1) the cognitivist claim that someone's intention to perform some action entails a belief that she will perform the action and (2) the claim that the motivational state involved in an agent's intention is voluntarily formed by the agent jointly imply (3) the view that the belief that one will do what one intends that is entailed by one's intention is a proper part of that intention. The view that the formation of intentions is directly voluntary is a sufficient condition for the strong cognitivist thesis because in order for the formation of the conative state involved in the intention to be voluntary, the agent's belief that she will do what she intends must be formed prior to the conative state in the temporal order. There is no way to satisfy this requirement without granting that the belief that one will do what one intends is a proper part of one's intention.

As I mentioned earlier, those who endorse weaker forms of cognitivism than those that hold that a belief is actually a proper part of an intention tend to hold that the belief entailed by an intention is based either inferentially or non-inferentially on one's awareness of the conative state that comprises the intention. On this account, the intention *qua* conative state is taken to [24]

[24] The only people who do not hold this, maintain the cognitivist thesis, and deny that a belief that one will do as one intends is a proper part of an intention endorse a view called "noninferential weak cognitivism." According to Clark 2020, Clark is the only person who endorses this view.

have been formed before the belief in the temporal order. However, I argue that this account cannot be correct because if this were the case, the intention could not have been voluntarily formed.

Suppose that the motivational state that featured in an agent's intention were formed prior to her belief that she would do what she intended. This would indicate that at some time, the agent in question possessed a motivational state whose object were the performance of a particular action, and, that at that time, she did not believe she would perform that action. The agent's ignorance of the content of the motivational state into which she entered in this case would suggest that the first order motivational state into which she entered, whose content were the performance of the action in question, could not have been formed on the basis of a second order conative attitude endorsing that state when she entered it. If it had been formed on the basis of such a second order attitude, the agent would have to have been cognizant of the object of the first order attitude when she entered the corresponding conative motivational state. However, [25] we have seen that she was not. Therefore, the formation of the agent's first order motivational

state whose object were the performance of the action in question in this case must have been involuntary.

So, we have seen that if the motivational state involved in an agent's intention were formed prior to her belief that she would do what she intended, it would have to have been formed involuntarily. However, we have seen that in order for a structure to count as an [26] intention, the motivational state featuring in that structure must be one that was formed voluntarily, or on the basis of a second order conative attitude endorsing the holding of a first order state with that content. Therefore, the structure involved in the case under consideration is not an intention. Moreover, it follows that on an account involving intentions, either the formation of the motivational state featuring in an agent's intention must follow that agent's belief about what she will do or it must occur at the same time. Both options entail that the belief that one will perform the action one intends to perform is a proper part of one's intention. Therefore, the strong cognitivist thesis is entailed by any plausible account of intentions on which these include both a conative and a cognitive state and the formation of the conative state is voluntary.

[25] This is consistent with Sartorio's view that what is done voluntarily has a certain causal history that is sensitive to reasons.

[26] It might be noted that you could deliberate about whether you want to do something and conclude your deliberation, but not intend to perform the action on which you settled until you actually formed a belief that you would perform it. For example, you could conclude in favor of meeting someone and only later form the intention to meet them. My response is that it is not clear that this kind of case is in fact possible. If you have not formed an intention to act in a certain way, you have not concluded your practical deliberation with respect to the question of whether or not to act that way. Hieronymi 2006 proposes a view along these lines, arguing that only one's formation of an intention can fully settle the question of how one will act. Cases in which an agent "settles" what to do before forming an intention to do it might be appropriate places to consider the role of credences in practical deliberation rather than taking full belief to be what characterizes the agent's alleged "conclusion." Thanks to [anonymous] for pressing this point.

IV. Conclusion

In this paper, I first argued from the implausibility of noncognitivism about intentions for the claim that an agent's intention to perform an action implies her belief that she will perform it. I then demonstrated from the implausibility of the implications of the position that the formation of the motivational state that features in some agent's intention is involuntary that the formation of the motivational state to perform an action that is part of the structure of her intention must be voluntarily formed by her. Thirdly, I showed that positing a distinction between rational and non rational desires can help us explain a class of cases involving ascriptions of intention whose explanation would elude an account of the structure of intentions on which an agent's intention could include a first order motivational state that were not endorsed by a second order attitude of hers. Finally, I contended that this claim and the cognitivist thesis jointly entail strong cognitivism about intentions.

Works Cited

Anscombe, G.E.M. (1957). *Intention*. Cambridge, MA: Harvard University Press.

Aristotle, Ross, W. D., & Brown, L. (2009). *The Nicomachean Ethics*. Oxford, UK: Oxford University Press.

Audi, R. (1973). Intending. *The Journal of Philosophy,* 70(13): 387-403. Retrieved from https:// doi.org/10.2307/2024677.

Audi, R. (1979). Wants and intentions in the explanation of action. *Journal for the Theory of Social Behaviour,* 9(3): 227-49. Retrieved from https://doi.org/10.1111/

j.1468-5914.1979.tb00427.x.

Audi, R. (1979). Weakness of Will and Practical Judgment. *Nous,* 13(2):173-96. Retrieved from https://www.jstor.org/stable/2214396.

Audi, R. (1991). Intention, cognitive commitment, and planning. *Synthese*, 86:361-78.

Audi, R. (1995). Acting From Virtue. *Mind*, 104(415), new series, 449-471. Retrieved from www.jstor.org/stable/2254637.

Austin, J.L. (1962). *How To Do Things with Words.* Cambridge, MA: Harvard University Press.

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

Clark, P. (2020). Intentions, Intending, and Belief: Noninferential Weak Cognitivism. *Pacific Philosophical Quarterly,* 101(2): 308-27.

Davis, W. (1997). A Causal Theory of Intending. In Mele, A. (ed.), *The Philosophy of Action* (p. 131-48). Oxford, UK: Oxford University Press.

Frankfurt, H. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy,* 66(23):829-39.

Gibbard, A. (1990). *Wise Choices, Apt Feelings.* Cambridge, MA: Harvard University Press.

Goldman, A. (1970). *A Theory of Human Action.* Englewood Cliffs, NJ: Prentice-Hall Incorporated.

Harman, G. (1997). Practical Reasoning. In Mele, A. (ed.), *The Philosophy of Action* (p. 149-77). Oxford, UK: Oxford University Press.

Hieronymi, P. (2006). Controlling Attitudes. *Pacific Philosophical Quarterly,* 87(1): 45-74.

Hume, D. (1986). *A Treatise of Human Nature.* London, UK: Penguin Books. Joyce, R.

(2001). *The Myth of Morality.* Cambridge, UK: Cambridge University Press.

Kant, I., Gregor, M., & Timmermann J. (2012). *Groundwork of the Metaphysics of Morals.* Cambridge, UK: Cambridge University Press.

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

Levy, Y. (2017). Why cognitivism? *Canadian Journal of Philosophy,* Online. Retrieved from https://doi.org/10.1080/00455091.2017.1345207.

Marušić, B. (2012). Belief and difficult action. *Philosophers' Imprint,* 12: 1-30.

McCann, H. (1997). Settled Objectives and Rational Constraints. In Mele, A. (ed.), *The Philosophy of Action* (p. 1-26). Oxford, UK: Oxford University Press.

Mele, A. (1997). Introduction. In Mele, A. (ed.), *The Philosophy of Action* (p. 1-26). Oxford, UK: Oxford University Press.

Paul, S. (2012). How We Know What We Intend. *Philosophical Studies,* 161(2):

327-46. Railton, P. (1986). Moral Realism. *Philosophical Review,* 95(2): 163-207.

Rinard, S. (2018). Believing for Practical Reasons. *Nous*. Retrieved from https://www.susannarinard.com/s/Believing-For-Practical-Reasons-For-Online.docx .

Sartorio, C. (2016). The Grounds of Freedom. In *Causation and Free Will* (p. 7-44). Oxford, UK: Oxford University Press.

Schroeder, M. (2010). *Noncognitivism in Ethics.* London, UK: Routledge. Setiya, K.

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

K. (2014). What is a Reason to Act? *Philosophical Studies,* 167(2): 221-35. Stampe, D.

(1987). The Authority of Desire. *The Philosophical Review,* 96(3): 335-81. Stevenson,

C.L. (1937). The Emotive Meaning of Ethical Terms. *Mind*, 46(181):14-31.

Velleman, D. (2000). *The Possibility of Practical Reason.* Oxford, UK: Oxford University Press.

Vihvelin, K. (2013). *Causes, Laws, and Free Will: Why Determinism Doesn't Matter.* Oxford, UK: Oxford University Press.

Williams, B. (1979). Internal and External Reasons. In *Moral Luck* (79-89). Cambridge, UK: Cambridge University Press.

Thanks to [various anonymous readers] for helpful comments.