An Assignment Theory of Desire (20180731)

Alonzo Fyfe

1. Introduction

> *The right act is the act that a person with good malleable desires and lacking bad desires would have done in those circumstances.*[1]

Ultimately, this is the thesis that I would like to defend.

I am not going to get there in this paper. I will only be making the first step. However, it may help to understand this step to have some idea of the destination.

I expect that many potential readers will immediately reject this. "He has *right act* on the left side of the equation and *good desire* on the right. No doubt I am going to find all sorts of question-begging and viciously circular claims as he inserts into *good desires* exactly the ideas he needs for *right action*."

That is a legitimate concern. It is true, I need an account of *good malleable desire* that does not beg any questions.

Of course, an account of *good malleable desire* must begin by answering the question, "What is a desire?" Then we can look about the ways in which desires are malleable and what makes them good (or bad).

So, this is that first step. What is a desire?

I suggest the following hypothesis:

---

[1] This formulation is non-accidentally similar to that which Rosalind Hursthouse presents: "An action is right iff it is what a virtuous agent would characteristically (i.e. acting in character) do in the circumstances." Hursthouse, 1999. This account of "good desire" will differ significantly from Hursthouse's Aristotelian concept of "virtue", but many of Hursthouse's arguments are applicable here*, mutatis mutandis*.

> *A desire is a propositional attitude - a mental state that that can be expressed in the form 'Agent desires that p', where 'p' is a proposition capable of being true or false, which assigns a value V to the proposition p representing the importance to Agent of p being made or kept true.*[2]

What does this mean?

It means that, if you want to go back to bed and get some sleep then you have a mental attitude to the proposition, "I am in bed and sleeping," that assigns to this proposition a certain value V representing how important it is to you that "I am in bed sleeping" be made or kept true. The more important this is to you, the more power the desire has to override other concerns such as getting the house cleaned up for company this evening or getting that philosophy paper written.

In order to explain this idea a bit more clearly, please allow me an opportunity to present a short glossary of terms and phrases that I will be using describing how I intend to use them. I also want to include some assumptions that I will be making – since a paper such as this cannot be expected to cover everything under the sun.

2. Glossary

I have two reasons for wanting to put some effort into more precisely defining my terms.

The first is that I find much of the discussion on this topic confused, and that this confusion is due to some degree on the confusing language that is being used. I think that much of this language came about under an assumption that the universe contained some sort of what J.L. Mackie (1977) called "objective, intrinsic prescriptivity." That is, it contains an assumption of

---

[2] I have toyed with the idea that both beliefs and desires assign a value to a proposition being true. The value that it assigns to a proposition that is the object of a belief - V(B) - represents the credence of that belief - the likelihood given whatever else the agent knows that the proposition is, in fact true. Whereas, as I said, the value it assigns to a proposition that is the object of a desires - V(D) - represents the importance to that agent of that proposition being made or kept true. However, I have not given the issue of belief in this regard a lot of thought. I toss that out here simply as something to think about.

intrinsic values and external reasons that, at best, is questionable. I would argue that this is false, but I wish to save that argument for later. At the very least, I want to warn about a language that seems to beg certain questions in favor of such a thesis and recommend, in its place, a language that will allow us to address this as an open question.

When Mackie presented his “error theory,” Simon Blackburn (1985) noted that Mackie went on to make substantive moral claims. Blackburn argued that Mackie must either be claiming that we have no choice but to make these false moral claims, or he needed to provide us with a new “schmoral” vocabulary that would allow us to make error-free claims. I hold that Blackburn missed the fact that Mackie did exactly this, providing an error-free definition of “morality” that allowed for the creation of true moral claims analogous to scientists offering an error-free definition of “atom” once they came to suspect that atoms did, indeed, have parts. (The original definition of “atom” was, literally, “without parts”.) Be that as it may, the requirement to offer a less question-begging vocabulary stands.

Furthermore, morality has a rare feature. Whatever it is, morality is something that virtually everybody within a community, regardless of education or background, must be able to grasp on some level. The history of philosophy contains references to the hypothetical “man on the Clapham omnibus” to represent your average reasonable individual and what he or she would have understood in a given situation. Yet, even this standard is too high. This is an “average” person, meaning that a full fifty percent of the population would fail to meet this standard. Morality has to be something that more than fifty percent of the population can use.

This does not imply that morality itself must be simple. The vast majority of people in a community learn to speak a sophisticated language at a very young age. This is an observed fact. It is, in part, in virtue and of this fact that we can dismiss any theory of language that implies that an agent could not determine the meanings of common words without years of advanced study. Similarly, we can dismiss any theory of morality that would require advanced cognitive capacities to understand that hitting one’s brother is wrong.

Put simply, my goal in this section is to state more precisely what I will be talking about and put it in terms that, I hope, will be easier to understand.

Humean or Non-Humean

Many of those who are familiar with the philosophical debate concerning desires may try to understand this in terms of a classic debate between “Humeans” and “Non-Humeans”.

Labels such as these are often useful. When we identify somebody as Utilitarian or Kantian, Muslim or Atheist, Marxist or Randian, it tells us a lot about a person in a phrase. These terms allow people to quickly identify a person with a set of beliefs in a large set of claims and counter-claims.

Yet, these heuristics sometimes fail, and this is an example.

A Humean theory of desire is generally one that conforms to Hume’s claim, “reason is the slave of the passions” and that it is "not contrary to reason to prefer the destruction of the whole world to the scratching of my finger" (Hume, 1896). Basically, beliefs and desires are two different kinds of things. Reason or rationality apply to beliefs. There is no rationality of desires.

In the philosophy of desires, the dominant view among those who identify as “Humeans” propose a “dispositional” or action-based theory of desires. Action-based theories hold that desires are dispositional states – dispositions to act so as to realize certain ends or goals or to engage in certain expressions or to experience certain emotions (Alvarez, 2017).

Standing up against action-based theories, we see a diverse sent of non-Humean philosophers who tend to hold that there is something to be said for the rationality of desires. These are roughly known as “evaluativists” or evaluation-based theories. The paradigm example of an evaluation-based theory holds that a desire is a form of belief. Specifically, a “desire that p” is “a belief that p is good”. Other versions say that a desire that p is to be understood as judging or perceiving (as opposed to believing) it to be the case that p is good. Still other versions suggest that the agent believes or perceives that p “ought to be” or “ought to be done”.

Since beliefs and judgments can be held to be reasonable or unreasonable, and perceptions can be accurate or inaccurate, desires can also be reasonable or accurate or fitting.

I will side with the Humeans insofar as I think that "reason is the slave of the passions" and, in a sense, it is "not contrary to reason to prefer the destruction of the whole world to the scratching of my finger."

At the same time, the hypothesis that I gave above is not an action-based theory of desire. It explains the desire that p in terms of p having a value. Such a theory appears to be evaluation-based.

What reconciles these two positions is that the value assigned to p in virtue of a desire that p is not subject to rational requirements. The desire that p is not a belief or judgment about what is good or ought to be, nor is it an accurate (or inaccurate) perception, nor is there a requirement that it be fitting. It is just a basic fact.

It is not contrary to reason for p to have a higher negative value where p = "there being a scratch on one's finger" than to q where q = "the destruction of the whole world" any more than that it is contrary to reason that rock A weighs more than rock B. Each rock simply weighs what it weighs and each end simply has the value that it has.

I will, of course, be discussing this subject in more detail later. My point here is simply to warn about potential confusion if one starts off trying to categorize this as Humean/action-based or non-Humean/good- or ought-based. It would be better categorized as Humean/evaluation-based. This itself may be confusing to somebody who does not see these as being compatible, but that is a major part of what I hope to address in the pages that follow.

Desire that p, Desire to φ

There is a convention in this discipline that says that desires can be expressed in the form "agent desires that p" or "agent desires to φ."

The symbol 'p' in the first phrase, as mentioned above, represents a proposition – the meaning of a sentence, which is capable of being true or false. Examples may include, "I desire that I be accepted into the PhD program," or "I desire that my favorite sports team win the championship," or "I desire that my obnoxious neighbor comes down with a dreadful disease so that I can watch him suffer."

The symbol '*φ*' in the second phrase represents some action. "I desire to stay home and cuddle with my wife," or "I desire to clean up the house before company arrives this evening," or "I desire to get this philosophy paper written by the end of the month."

That is not to say that these are the only types of desire claims we see in the English Language. We may see expressions such as "I want a beer" or "I would like to have somebody honest run for public office for a change." These do not directly fit into the two models given above. However, they can be rewritten so as to fit the model. "I want a beer" becomes "I desire to drink a beer" or "I desire that I am drinking a beer." Similarly, "I would like to have somebody honest run for public office for a change" becomes "I desire that somebody honest runs for public office." (Oddie, 2017)

In fact, the convention states that, in the end, even "desire that *φ*" statements can be translated into "desire that p" statements. I have already demonstrated this where I equated "I desire to have a beer" with "I desire that I am drinking a beer." It is in virtue of this possibility that I expressed the desire hypothesis at the start of this paper exclusively in terms of a "desire that p". We can rewrite all other types of desire into an expression having this form.

Realizing P

Another issue that will be important in what follows concerns the question of what desires aim at.

I will be building on the suggestion that a desire that p aims at realizing p. By "realizing" I am saying that the desire assigns a value to it being the case that 'p' is real – meaning that it is true. If p is false and the agent desires that p, then the agent will seek to make it the case that 'p' is true. If 'p' is already true and the agent desires that p, then the agent has a reason to preserve p – to keep it the case that 'p' is true.

On this account, the object of a desire is the correspondence between the proposition and a state of affairs. The desire is fulfilled where the proposition is true within that state of affairs.

One way to illustrate what this means is by looking at Robert Nozick's (1974) experience machine thought experiment. Nozick asked readers to imagine a machine that will feed the brain with images – nerve stimulations – identical to those that would come from the apparent

fulfillment of an agent's desires. So, if one wanted to aid the poor and starving in the most poverty-stricken parts of the world, the experience machine, like some virtual reality device, will give the subject a perfectly believable (and believed) experience of aiding the poor and starving in the most poverty-stricken parts of the world.

Nozick then asked us to consider whether we would want to live our lives inside of the machine. We would, of course, lose all memory of our decision to enter the machine, so that it would be as if we were living the life that the machine caused us to experience.

A refusal to enter the machine implies a preference for real-world experiences. We would rather risk the failure of perceiving our desires as being fulfilled in the real world, then to live with the illusion of them being fulfilled in the experience machine.

The idea that desires aim at realizing that which is desired implies that the value that an agent will find in the experiences generated by the experience machine is found in whether the machine can make true that which the agent desires.

If the agent desires that she have the experience of aiding the poor and starving in the most poverty-stricken parts of the world, then the machine has the capacity to make this true. It can make true the proposition, "I am having an experience of aiding the poor and starving in the most poverty-stricken parts of the world." However, if she desires that she actually aid the poor and starving in the most poverty-stricken parts of the world, then the machine has nothing to offer her – unless, perhaps, the owner of the experience machine was to say, "For every hour you spend in the experience machine I will donate $1 million to aiding the global poor."

Both agents are seeking to make true the propositions that are the objects of their desire. The difference between them is that they want different things. One of them wants the state of affairs of having an experience, the other wants the state of affairs of people being helped.

Intrinsic vs. Instrumental Desires

Another convention among those who write in this field is to distinguish what are called "intrinsic desires" from "instrumental desires".

Another conventional way of talking about these two concepts is to talk about ends (or goals, or objectives, or that which is valued for its own sake) and means.

I fear that the term “intrinsic desires” is an invitation to making a serious mistake. The term invites us to relate “intrinsic desires” to “intrinsic values” – a type of “objective, intrinsic prescriptivity” that J.L. Mackie (1977) and Sharon Smith (2006) warn us against. In Section 4, I will be arguing that this type of claim is mistaken, which means that the term "intrinsic desires" seduces people into making a serious mistake.

Philosophers don’t tend to have "intrinsic values" in mind when they talk about "intrinsic desires". They simply mean this term to refer to that which is desired for its own sake and not as a means to some further end. Though, of course, something can be valued both for its own sake and as a means to an end, but these are still two different types of value sometimes found together. On this matter, my fear is that philosophical jargon makes the subject confusing to the non-philosophical reader. For that reason, I prefer the term “final desires” or “end desires” to “intrinsic desires.” This specifies that the desires are concerned with the agent's ends, goals, or ultimate purposes, regardless of what it is that gives those ends its value.

Within this same set of terms, I also want to endorse the tradition that an instrumental desire is not some different kind of thing distinct from final desires and means-ends beliefs. Rather, it is a term that refers to bundles of final desires and means-ends beliefs. "I want some aspirin" states "I want to get rid of this pain and I believe that if I take some aspirin than the pain will go away." A person who does not want to go to work as an end or goal may still want to go to work insofar as they want a comfortable and secure place to sleep at night and food to eat combined with a means-ends belief that going to work will result in receiving money that one can use to purchase such things.

It will be particularly important to keep in mind when we discuss whether a desire is a type of belief. Since instrumental desires contain a belief component, they are a type of desire that functions, in some ways, like a belief. For example, a claim about instrumental desires can be mistaken precisely because the beliefs built into a particular instrumental desire are mistaken (or false). A person who “wants an 8mm wrench to tighten a bolt” might actually want a 7mm

wrench. However, the possibility of error in the realm of instrumental desires does not imply the possibility of error in the realm of final desires. Drawing such an implication would be a mistake.

For the sake of convenience and illustration, I will follow the practice of referring to this distinction between instrumental and final ("intrinsic") desires. However, as others have noted, these are not the only two ways in which an object of evaluation can relate to a final end.

It is also possible for the object of evaluation to be a part of a whole. The peanut butter in a peanut butter cup does not have value for its own sake (or not solely for its own sake) or in virtue of being a means to the end of having a peanut butter cup. It has value in virtue of being a part of the whole. This is even more true of the cocoa, sugar, or salt. These have value in virtue of being ingredients in a valuable whole.

Just as is the case between means and ends, people can make mistakes about the relationship between parts and wholes. A person can be mistaken about whether that which is being evaluated is a part of the whole or can be mistaken about whether the whole would continue to have its value with or without this part.

Nomy Arpaly and Timothy Schroeder (2014) make a case for realizer values. In a few pages, I will talk about a person who wishes to buy a Picasso who doubts that the painting being offered to him is a Picasso. The relationship of this painting to what the individual desires is not that of means to end, or part to whole, but as something that would realize the agent's desire to own a Picasso painting. Consequently, it has what Arpaly and Schroeder will call "realizer value".

Yet, usually, for illustrative purposes, we will focus on instrumental desires and final desires.

Desires Narrowly vs. Broadly Defined

If a person wishes that he had lived during the Enlightenment, does he desire that he had lived during the enlightenment?

In some of the things written on the subject of desire, the author uses a broad conception that Donald Davidson (1963) calls "pro-attitudes". On this conception, hopes are types that f desires.

On another conception, hopes are not desires. To have a desire that p one has to believe that there is something one can do to realize p. There is nothing one can do to have made it the case that one had lived during the enlightenment. Consequently, one can wish to have lived during the enlightenment, but could not desire it. (Doring and Eker, 2017)

In this paper, I will be looking at "desires" in the broad sense. For example, I will not be distinguishing "desiring that p" from "wishing that p."

My two main reasons for wanting to use this definition includes the convention of speaking in terms of the belief-desire theory of intentional action. Authors tend to distinguish between beliefs and desires in this sense in a broad sense. For example, Elizabeth Anscombe (1963) distinguishes them based on "direction of fit", where beliefs are those things that must conform to the world and desires are those things that motivate the agent or give the agent a reason to change the world. Even the person who hopes that p has an attitude that is best understood as a hope that the world had been different. That is, it has a world to mind direction of fit.

The other main reason is because I am ultimately interested in a theory of value and I think that a broad definition of desire links up better with the concept of value. After all, the person who hopes that p still values that p, even if she cannot think of a way to realize that which has value.

"There Is" vs. "She Has" a Reason

Another area where talk about desires is unnecessarily confusing is exemplified by the Stanford Encyclopedia of Philosophy entry on "Reasons for Action: Internal vs. External."

The entry provides this account of a revised version of what the authors call the Humean Theory of Reasons:

> *[Humean Theory of Reasons] (revised): If there is a reason for someone to do something, then she must have some desire that would be served by her doing it, which is the source of her reason.*[3]

On a plausible interpretation of this, it is easily proved false, even within a Humean framework of desires.

Imagine a community made up of two people, Alph and Bet. Each person has an aversion to their own pain, and that's it. Because of this aversion, each person has a reason to avoid their own pain.

In this community, two reasons exist. There is Alph's reason for avoiding his own pain, and Bet's reason for avoiding her own pain.

In this situation, there are conditions where the antecedent is true, but the consequent is false.

Specifically, *there is a reason* for Bet to refrain from putting Alph in a state of pain - a reason that exists. That reason is Alph's reason to avoid being in pain, which is grounded on Alph's aversion to his own pain. That is to say, the antecedent, "There is a reason" is true.

However, the consequent is false. That is to say, it is not the case that Bet has a desire that would be served by refraining from causing Alph pain. The only reason that Bet has is her aversion to her own pain, and we may assume that the act that causes Alph pain will not also cause her pain.

This true antecedent (there is a reason; Alph's aversion to his own pain) and false consequent (Bet has a desire that would be served) shows that the conditional is false.

The likely way out of this trap would be to give a different interpretation to "there is a reason for someone to do something." "There is a reason" cannot refer to Alph's reason for avoiding

---

[3] Stanford Encyclopedia of Philosophy, "Reasons for Action: Internal vs. External", First published Thu Sep 4, 2008; substantive revision Fri Aug 18, 2017, 7/8/2018, https://plato.stanford.edu/entries/reasons-internal-external/.

pain. It must be a "reason for Bet to refrain from causing pain." Bet only has a reason to refrain from causing pain if Bet has a desire that would be served by refraining from causing pain.

If this is the preferred interpretation, we can make the theory easier to understand by simply being explicit about this.

> *[AHTR1] If someone has a reason to do something, then she must have some desire that would be served by her doing it, which is the source of her reason.*[4]

What I have done is changed "if there is a reason" in the antecedent to "if someone has a reason" in the antecedent. This eliminates the interpretation that lead to the counter example.

Plus, we can also say the following:

> *[AHTR2] If there is a reason for someone to do something, then there must be some desire that would be served by her doing it, which is the source of that reason*

Within the context of AHTR2, "there is a reason" to refrain from causing Alph pain based on Alph's aversion to being in pain, and it is true that "there must be some desire that would be served" – this, of course, being Alph's aversion to his own pain. But this does not imply that Bet has a reason to refrain from causing Alph pain.

The question remains whether the Humean Theory of Reasons (revised) contains something more than what is contained in AHTR1 and AHTR2. Is there also a sense in which Alph having an aversion to pain mean that Bet has a reason not to put Alph in a state of pain? I am going to leave that question open for the moment, though I think that the counter-example I provided above stands. The Humean Theory of Reasons implies that, under the conditions specified here, Bet has no reason to refrain from causing Alph pain.

---

[4] Stanford Encyclopedia of Philosophy, "Reasons for Action: Internal vs. External", First published Thu Sep 4, 2008; substantive revision Fri Aug 18, 2017, 7/8/2018, https://plato.stanford.edu/entries/reasons-internal-external/.

In short, I want the option of denying, "There is a reason" implies "she has a desire" (which I take to be false), while still preserving the option of saying, "There is a reason" implies "there is a desire" and "she has a reason" implies "she has a desire" (which I take to be true).

What would Hume say?

The question of whether Hume would say that Bet has a reason to refrain from causing Alph pain, when Bet only has an aversion to her own pain, is an interesting question for historical reasons. However, I fear that this would be a long and unnecessary digression. Without passing judgment as to which position Hume will take, I will be calling as Humean the version that is the combination of AHTR1 and AHTR2.

Summary

In this section, I sought to specify what I want to talk about in talking about desires.

I want to question the tendency to associate action-based theories of desire with Humean theories, and classifying evaluation-based theories as non-Humean. The view that I seek to defend is both Humean and evaluation-based.

I wish to use the convention of expressing desires in the form “desire that p” as a desire concerning a state of affairs and “desire to φ” to express a desire to perform some action. Ultimately, praises of the form “desire to φ” can be restated as “desire that I φ”.

Ultimately, I am going to seek to show that a “desire that p” is a desire that the proposition p be made or kept true.p, which I will sometimes express as the realization (making real) of p.

While authors in the field like to write about “intrinsic desires” and “instrumental desires,” I find the term “intrinsic desires” to invite a mistaken inference to “intrinsic values”. Consequently, I prefer to use the terms “final desires” or “ends” to refer to that which is valued for its own sake.

I will be using the term “desire” in its broad sense, to refer to any pro attitude that has a “world to mind” direction of fit.

I will be using the phrases “there is a reason”and “there is a desire” to mean that a reason or desire exists, and “he/she has a reason” and “he/she has a desire” to refer to a reasons or desire that the agent has. This means that it will be possible for “there is a reason or desire” to be true when “she has a reason or desire” to be false when the former phrase refers to somebody else’s reason or desire.

With these concepts in mind, I want to look at how this may influence talk about desires.

Applications of These Terms

In the previous section, I mentioned how I want to use certain terms that I hope will make this paper a little more comprehensible. To illustrate these recommendations, I would like to apply them to a set of claims about desires.

Michael Smith (1987) defends a dispositional theory of value. In defending his theory, he presented three counter-examples. At this point, I do not wish to defend or criticize his theory. I simply want to show how the terms I mentioned above might make it easier to understand the issues.

Stepping on Toes

One counter-example went as follows:

> *Suppose that I am standing on someone's foot so causing him pain, and that I know that this is what I am doing. Surely, we can imagine its being appropriate for an outsider to say that I have a reason to get off his foot even though I lacked the relevant desire, and, indeed, even if I desired to cause him pain.*

I need a name for Smith’s victim here, so I will call her Jones.

I am not at all eager to accept that Smith has a reason to get off of Jones’ foot. Indeed, if Smith lacks the relevant desire, then Smith lacks the relevant reason. At the very least, I want to suspend judgment until I see an argument in defense of this claim. Furthermore, I will be surprised if somebody can provide such an argument.

A lot depends on this phrase, “lacked the relevant desire”. This has to include not only lacking an aversion to stepping on toes and an aversion to causing pain. Smith must also lack a desire to get that job that Jones will be interviewing him for in an hour, or for his classmate to see him as a kind person. These are not cases of Smith having a reason though he lacks the relevant desire. These are cases where Smith has a reason because he has the relevant desire.

Please note that Jones has a reason to cause Smith to have a reason to get off of his foot. Jones has a reason to threaten or bribe Smith, thereby creating a reason for Smith to get off of her foot. However, the reason that Jones creates is tied to a desire that Smith has. A threat is a promise to thwart some desire that Smith has if he does not get off of her foot. A bribe is a promise to fulfill some desire that Smith has if he would get off of her foot. These reasons that Jones can create still depends on th3 desires Smith has.

We can also say that Jones has a reason to cause Smith to have an aversion to stepping on her toe or causing her pain. If he could – through a twitch of her nose – cause Smith to have an aversion, she would have reason to twitch her note. Similarly, if she could create such an aversion in others by praising those who refrain from stepping on toes and condemning those who do not refrain, she would have reason to offer praise and condemnation. In this sense, we could say that Smith *should* have an aversion to stepping on her toes – in the sense that there are reasons to cause Smith to have such an aversion. However, saying that he should have such an aversion (such a reason) is not the same as saying he does have such a reason.

At this point, I am simply not so willing to grant Smith’s claim that he has a reason to get off of Jones’ foot even though he lacks the relevant desire. This is something we need to discuss.

The Gin and Tonic Case

Smith presented a second case that the language of desire makes unnecessarily confusing.

In this case, he desires to drink a gin and tonic. Believing that the glass in front of him contains gin, he says that he wants to pour some tonic into it and drink it. However, the glass actually contains petrol.

Remember, Smith is discussing this case as an objection to his dispositional account of desire. In that context, Smith is said to have a desire to pour tonic into the glass in front of him and drink the contents. Furthermore, he has the requisite beliefs. However, contrary to Smith's own dispositional theory, he lacks a reason to tonic into the glass and drink its contents.

I am not looking at objections to Smith's dispositional theory yet. I simply want to look at the way this case is described. Smith concedes that it makes sense to say that he desires to pour gin into the glass containing petrol and drink the contents. I would deny this. Smith believes that he has such a desire. However, Smith's belief – just like his belief that the glass contains gin – is mistaken. He has no such desire.

Imagine that Smith is with a companion, Jones. Jones sees Smith pour tonic into the glass and begin to raise it to his lips to drink the contents. She puts her hand over the top of the glass to stop him and says, "You don't want to do that."

I would assert that Jones' statement is literally true.

Consistent with this, Smith is unlikely to say, "Jones, your statement is so clearly false that you obviously have no understanding of the English Language or of what it means to say that I want to drink what is in this glass." On the contrary, Smith would likely ask, "Why not?" In asking this question, Smith is admitting to the possibility that "I desire to drink the contents of this glass" is false – and that Jones knows a truth that he is unaware of. This is why he is asking Jones to provide evidence for that truth.

When Jones provides Smith with convincing evidence, Smith could be expected to say, "You're right. I don't want to do that." This statement, "I don't want to do that," did not become true the instant he acquired the true beliefs about what was in the glass. It was true from the start. The corresponding proposition, "I want to pour tonic in this glass and drink its contents" was false from the start.

Why does this matter?

I want to get away from the (false) implication that when Smith pours tonic in the glass and drinks the content that he is getting something that he wants. This is simply not the case. No

desire of his is fulfilled when he drinks the contents of the glass. He has no desire that is to be understood as a desire to pour gin into the glass and drink its contents. His only desire is to have a gin and tonic, which happens to be combined with the (false) belief that the glass contains gin.

For the sake of this discussion, I will be using "Agent has a desire that p" or "Agent has a desire to φ" to refer to the desires that Agent actually has. If the agent does not want p or to φ, then he has no desire that p or desire to φ.

The Picasso Case

Smith's Picasso case is similar to the gin and tonic case in that it concerns an agent with false beliefs. In this case, instead of falsely believing that p will fulfill his desire, Smith falsely believes it will not.

In this case, Smith wishes to purchase a Picasso painting. Michael offers him an opportunity to purchase a Picasso at a quite reasonable price. However, Smith is suspicious that the painting is not a genuine Picasso, so he refuses to buy it.

Smith suggests that he does not want to purchase the painting, even though, by purchasing the painting, he would be fulfilling his desire to own a Picasso.

As with the gin and tonic case, I would say that Smith does want to buy the painting, he just doesn't know it. He believes (falsely) that he doesn't want to purchase the painting.

Once again, we can appeal to Jones, who takes Smith by the arm and says, "Smith, buddy, you're making a mistake. You want to buy that painting. Trust me." What Jones says is true. Smith's claim that he doesn't want the painting is false.

Here, too, the debate is not a debate over whether a given proposition is true or false. The debate is over what terms we are going to use in talking about a given state of affairs. We agree that Smith has a desire to own a Picasso, that he can fulfill this desire by buying the painting, and a belief that the painting is not a Picasso. There is no disagreement over the facts. There is only a disagreement over how we are going to talk about the facts.

I want to avoid saying that Smith does not want to buy the painting. This is because I believe that claims such as this lead to confusion. It suggests that buying the painting would fulfill no desire that Smith has. As we have already established, this is false. With this false inference, we must choose between risking error and confusion or going through a great deal of work to prevent that error and confusion.

On Smith's way of dealing with the issue goes as follows:

> *[T]he reason that I have to buy the painting in front of me is a normative reason. For it suffices for the truth of the claim that I have such a reason, that there is a requirement-in this case, in the broad sense, a requirement of rationality"-that I buy the painting in front of me.*

More generally, Smith seeks to distinguish between motivating reasons (those that cause action) from normative reasons (those that justify action). On Smith's account, it makes no sense to say that he has a motivating reason to purchase the painting, since he wasn't motivated to purchase the painting. This is the sense in which the statement that he did not want the painting is true – he lacks a reason that would motivate him from purchasing the painting.

However, Smith would argue that he does have a normative reason to purchase the painting. This normative reason is grounded on the fact that purchasing the painting would have fulfilled his desire to own a Picasso. It is what a person who wanted to purchase a Picasso ought to have done.

Remember, my objection is not that Smith is wrong, or that he fails to capture normal ways of speaking. My objection is that this way of speaking leads to confusion and error. Smith is saying that there is a rational requirement that calls for buying the painting. However, in refusing to purchase the painting, Smith may be acting in a perfectly rational manner. His desire to purchase a Picasso is not irrational, and his belief that this art seller is offering him a forgery may be perfectly rational as well, given his evidence. Indeed, it may be more rational than believing that the painting is an original Picasso.

On the account that I offer, what explains the fact that Smith did not buy the painting is not that he lacked a motivating reason. It is that he had a false belief that the painting was a forgery. Because of this false belief, he had a false belief that buying the painting was not something he had a reason to do.

Smith would say quite emphatically that he does not want to purchase the painting. However, non-coincidentally, he would also say, just as emphatically, that the painting is not a Picasso. Indeed, the strength of his emphasis that he does not want to purchase the painting is proportional to the strength of his conviction that he believes the painting to be a forgery. He is wrong about the painting. We can capture the fact that his unwillingness to purchase the painting is proportional to the credence he assigns to the belief that it is a forgery by saying that the belief is responsible for his refusal.

At this point, one can object that I am arguing for a significant change in our use of language. As evidence that this is not such a radical change, let us change the example slightly. The art seller convinces Smith that this is a genuine Picasso, so now Smith believes this. The dealer then invites Smith to play a game. The seller takes a black stone and a white stone, puts his hands behind his back, and puts a stone in each hand. Then, holding his closed hands out, asks Smith to choose one. If Smith chooses the hand with the white stone, he gets the painting without paying for it. Otherwise, he pays the price of the painting and gets nothing.

Smith clearly wants to choose the hand with the white stone. He does not know which hand it is, but that is beside the point. The clear answer to the question, “Which hand do you want?" is “The one with the white stone.” This answer does not look at what Smith believes, it looks at what will fulfill his desire to have a Picasso painting without paying for it. Here, the relationship between desire or want and what actually fulfills that desire becomes a bit more obvious.

The Pluto Proposal

Because this is a dispute over what words to use, I am not able to argue that Smith and the others who use these terms are making a factual error. It is a dispute about what language to use, not what is true.

One way to settle such disputes is to some conceptual analysis and look at how the terms are regularly used in common speech. I did a little of that in the sections above where I said that it made sense for Jones to tell Smith that he didn't want to drink out of the glass and that he did want to buy the painting, However, even if I were to lose this argument, there is another argument for me to make.

This is the argument that our language would be a cleaner and less confusing if we adopt this convention.

In 2005, when astronomers discovered more objects in the Kuyper Belt about the same size as Pluto, they addressed the question of whether to call Pluto a planet. Specifically, they debated giving a new definition to the term 'planet'.

The question of what the definition of 'planet' should be is not a question that can be answered through a scientific experiment. What to call Pluto is a different kind of question compared to the question of its mass or how long it takes to orbit the sun.

Furthermore, it was not considered to be a question to be answered using conceptual analysis. A look at how the concepts were traditionally used would have shown that "Pluto is a planet" was as fixed in language as "Socrates was a human."

The argument against calling Pluto a planet was grounded on the assumption that our language and our science is cleaner if we classify like objects with like objects. When astronomers discovered Eris in the Kuyper belt, which was the same size as Pluto, and it appeared likely that astronomers would find other like objects, some astronomers said that Pluto should be reclassified to fit in with those things that were like Pluto. Pluto had more in common with these other large Kuyper belt objects than it did with the eight other planets. So, classifying like with like, Pluto would no longer be a planet.

We do not have an International Philosophical Union that can settle the definition of a desire for us. Instead, the best we can do is provide reasons for believing that a particular usage would be better than alternatives and to see if people are willing to adopt it. I have provided some of those reasons above.

The same principle applied to the term “atom”. (Mackie, 1977). Once it was discovered that the smallest bits of carbon, sulphur, and iron had parts, it made sense to ask whether they were still to be called “atoms”. After all, “atom” literally meant “without parts”. This was not a question to be answered in a laboratory, or through conceptual analysis. It was to be settled according to principles of convenience, clarity, and precision – on practical concerns.

I actually think that the issue with the confusing language surrounding desire has something in common with the story of atoms. The language of desire was built on an assumption of external reasons in the same way that the language of atoms was built on an assumption of indivisible smallest bits of matter. Conceptual analysis begs the question in favor of these assumptions. When those assumptions come under dispute, it is often not wise to continue to use the traditional language in traditional ways. We must make changes.

I will be rejecting the thesis that there are external reasons. But I will not reject the thesis that the language of desire assumes that they exist. An examination of the changes I propose will show that, for the most part, they seek to remove assumptions of external reasons and intrinsic value from that language.

Hopefully, I also avoid begging questions in that other direction.

I suggested “final desire” instead of “intrinsic desire” since the latter tempts people into the false belief that such desires take objects in virtue of the object’s intrinsic value.

I suggested against using “there is a reason for somebody to do something” in a way that is equivalent to “somebody has a reason to do something” because it assumes that a reason that exists is something that everybody automatically has.

I suggested saying that Smith does not have a desire to pour tonic into the glass sitting in front of him and drink its contents and does have a desire to buy the painting he thinks is a forgery because this respects the fact that the act will not or will fulfill a final desire of his respectively, regardless of his false beliefs.

If one wants to insist that I am seeking to revise time-honored ways of talking about desire, so be it. In that case, I invoke the Pluto Principle.

I think that this will be easier to understand.

3. Dispositional Theories of Desire

I have specified how I wish to talk about desires, so now it is time to start talking about them.

If I am right in what I wrote in the previous section, readers should find what follows easier to understand than they otherwise would have, unless the reader has become familiar with conventional philosophical jargon. Let us see how this works.

I first want to take on dispositional theories of value.

The Stanford Encyclopedia of Philosophy calls these action-based theories and describes them as follows:

> *For an organism to desire p is for the organism to be disposed to take whatever actions it believes are likely to bring about p.*[5]

Michael Smith (1987) probably provides the default account of dispositionalism.

> *[A]ccording to this conception, we should think of the desire to φ as that state of a subject that grounds all sorts of his dispositions: like the disposition to φ in conditions C, the disposition to φ' in conditions C', and so on.*

Sabine Doring and Bahadir Eker (2017) express this position as follows:

> *Necessarily, for any agent a, any proposition p, any time t, and any act type φ, if, at t, a desires that p, then a is disposed at t to φ in circumstances where a takes her φ-ing to be conducive to p's being the case.*

These accounts, which relate having a desire with having a disposition to act in a way that the agent believes will bring about a particular state of affairs, is one that we can test in two

---

[5] Stanford Encyclopedia of Philosophy, "Desire", First published Wed Dec 9, 2009; substantive revision Thu Apr 9, 2015, 7/8/2018, https://plato.stanford.edu/entries/desire/.

different ways. We can test it by seeing whether we can find any dispositions to act in a way that the agent believes will bring about a state of affairs that is not desired. And we can test it by seeing if there are things that are desired where the agent is not disposed to act in a particular way.

On example of the first type of test is that which Warren Quinn (1993) describes involving a man with a disposition to turn on radios even though turning on radios does not serve as a means to some valued end and is not something the agent values for its own sake. Without it being the case that the end is valued, according to Quinn, it is not desired.

Examples of the latter include cases where the agent desires something he could not be at all disposed to change because it is – and is known to be - beyond his power, such as wishing to have never been born, that pi was an irrational number, or anything that is already the case (since the agent cannot bring about that which has already been brought about) (Schroeder, 2004; Lauria, 2017).

I will discuss these types of cases shortly, but I want to say a quick word about what the dispositional theory gets right.

What Dispositional Theories Get Right

There is much to be said in defense of dispositional theories of desire. First among these is its close association to a Humean theory of intentional action. David Hume provided a fundamentally intuitive account of action as a combination of certain desires (e.g., wanting to have some chocolate cake) and beliefs (a belief that there is some chocolate cake in the kitchen, how to get to get to the kitchen, and the like). Michael Smith (2012) describes this as the standard story.

The key component of this standard story is the sharp distinction between beliefs, which are cognitive states or states of understanding about what is true of the world, and conative states or interests in the world being different (or in making sure it remains the same). Later in this paper, we are going to be looking at theories that blur this distinction. However, I am going to

assert that this distinction between beliefs and desires is one of the things that the standard story gets right.

Another thing that the standard story gets right is that desires provide the motivation for intentional action. Desire, as I said, is an interest in the world being different, or in remaining the same. Changing the world, or, in some cases, preventing change sometimes involves the agent doing something. Towards these ends, desires provide motivation to act. Beliefs, then, guide those actions towards the realization of the desired end, at least when those beliefs are true and relevant.

In saying that there is a connection between desires and motivation to act, I mean something like the following:

> *Necessarily, for any agent a, any proposition p, any time t, if, at t, a desires that p, then, if there is an act type φ such that, at t, a thinks that her φ-ing would be conducive to p's being the case, then a would be motivated to φ.*

Note that this is like the dispositional account we get from Doering and Ekert (2017):

> *Necessarily, for any agent a, any proposition p, any time t, if, at t, a desires that p, then there is at least one act type φ such that, at t, a does not think her φ-ing not to be conducive to p's being the case.*

The two are similar, but they have a significant difference. Doering and Ekert are saying that their account, "merely rules out that you can desire that p, while thinking that nothing you could do in any possible situation would be conducive to p's being the case." In other words, you must think that there is at least a hypothetical situation in which you can do something to realize p. Without this, you cannot desire that p.

In contrast, the account I presented says that an agent can desire that p even though the agent knows as an indisputable fact that there is nothing she could do to realize that which she desires. She can have a desire that she was a mediaeval Eurpean queen or that she had become a famous novelist while still in her early 20's. These are desires, according to the sense in which I am using the term, but not in the way Doering and Ekert use the term.

At the same time, Doering and Ekert are among those who use a narrow concept of desire than I do. Specifically, they distinguish desiring from wishing. Whenever somebody wants something where there is no way to bring it about – e.g., that one had never been born or that one had taken the risk and asked Sam to the prom 20 years ago, this is not a desire that p, this is a "wishes that p".

This makes the dispute the what Doering and Ekert present and what I present merely semantic. The question becomes, "Is a desire a species of the genus 'wish' where "there is at least one act type φ such that, at t, a does not think her φ-ing not to be conducive to p's being the case." Or is it a case that a wish is a species of the genus 'desire' where there may be no act type φ such that, at t, a does not think her φ-ing to be conducive to p's being the case?

This is comparable to the dispute over whether to call Pluto a planet. We may try to settle this dispute by means of conceptual analysis and look at the way people conventionally use the terms. Yet. Even that will not settle the argument if one can make the case that a revised set of definitions would be easier to work with.

Without settling this issue, I could simply say that I am offering a theory of what would be called 'wish' (as opposed to a theory of desire) in the Doering and Ekert terminology. Furthermore, Doering and Ekert need a theory of wish. The hypothesis that a desire is a wish combined with the conceivability of action tells us very little about what a desire is without a theory of wish to go along with it.

Doering and Ekert could take a different route – claiming that desires and wishes are entirely different things. This is suggested in the subtitle to their article, "Why we need not value what we want." very title of their article: The thesis that a desire is a disposition to act suggests that there is no wishing involved in desire. Unfortunately, defending the thesis that the possibility of action is necessary for desire does not prove that it is sufficient, or that the valuing of what we desire is not also necessary.

A thesis that desiring and wishing are different types of things would have to handle questions such as: What happens when a person has a desire that p (in the Donner and Ekert sense) comes to believe that there is no act of φ-ing that is conducive to p. For example, what happens

to his desire to become Partner by the time he was 35 years old when, on his 35th birthday, he is still an Associate? Or what happens to the wish when Agent comes to realize, after all, there IS something he can do to realize that which he was wishing for?

Fortunately, we do not need to address those questions for the point that I want to make here. That point is that there is something that the dispositionalists get right about desire. Desires produce dispositions to act. With a desire that p, the instant Agent comes to believe that *φ*-ing has a chance of realizing p, Agent will be disposed to *φ.* In criticizing dispositional theories, I will not be criticizing this.

What I will be criticizing is the idea that you can have desires without value. A person with a "desire that p" is disposed to bring about p because he values p being the case.

Radioman

Warren Quinn (1993) provides a classic counterexample to the thesis that desire is a disposition to act. In his example, Radioman is an alleged a person with a disposition to act, but who has no desire.

Quinn describes Radioman as somebody who is simply disposed to turn on radios if there is one within reach that he discovers is not on. He has no interest in listening to music or in hearing any type of broadcast. In fact, he would prefer silence. He doesn't value having radios on in any way or the act of turning radios on. It is simply that, as he passes buy such a radio, he will reach out and turn it on. He has, in this case, a disposition to act so as to turn on radios. However, he has no desire to do so. Consequently, having a disposition to act cannot be identified with having a desire – not if there are dispositions to act that are not desires.

Quinn offers this case against theories of desire, like those of Michael Smith (1987, 2012) that take desires to simply be dispositions to act.

> *[T]he standard story, told in the way in which I have told it, conceives of the desires that move us as simply dispositions to be moved in certain ways, depending on what means-end beliefs we have . . . What it is to be motivated*

> *just is to be disposed to move in certain ways, depending on your means-end beliefs.*

What Smith gets wrong is that desires are more than dispositions to be moved. Desires recognize some value in that which is desired, and the desire motivates the agent to realize this end.

Quinn's Radioman example illustrates this point by describing an agent disposed to bring about a state he does not care about. Without the caring, there is no desire.

Smith responds by arguing that the case of Radioman is either one that defies explanation or even comprehension, or that Radioman has a state that is at least a species of desire.

The first option asks us to find a way of explaining Radioman's behavior if it is not a desire. Why is he turning on the radios? It is not a muscular twitch or the effect of a drug or some other external cause since turning on radios must be some sort of action. It must be something that Radioman does. Without some sort of explanation to work with, we can do nothing but throw up our hands and give up any attempt to answer the question. It simply happens.

On the second option, Smith notes that Radioman's behavior is under the influence both of beliefs (knowing that a radio is present, how to reach the radio, and how to turn it on) and desires (threatening to shoot Radioman if he turns on another radio may prevent him from turning on another radio). This makes it appear that Radioman's impulse to turn on radios is, in fact, another desire. Perhaps it is triggered by the mere sight of a radio. Perhaps it is a particularly strong desire. However, it is a desire nonetheless and, thus, provides no counter-example to Smith's thesis.

One of the reasons why I dislike thought experiments such as this is because it is so difficult to get a grasp of what is going on. It is easy to manipulate the results and build in the elements that serve the interests of the author.

However, Nomy Arpaly and Timothy Schroeder (2014) provide us with a more mundane example of dispositions without desire. It is something much like Radioman's behavior, but

something we should all be able to understand because it is something we have probably all experienced: the force of habit.

For 12 years, when I left work, I would turn right upon exiting the building I work in to get to the bus station to catch the bus home. Then, the city built a new bus station in the opposite direction. When they were done with construction, they closed the old bus station and opened the new one.

For several days, when I left the building to go home, I turned right.

Like Radioman, I had a disposition to turn right when I left the left the building. I did not value turning right for its own sake. Nor was it a means to some other end, not with the new bus station being in the opposite direction.

It would be a mistake to attribute the disposition to act to a desire to get to the bus and a false means-ends belief. At any moment, somebody could have asked me for directions to the bus stop and I could have given a true response. My beliefs were in order, but I was still disposed to turn right.

Learning to play the piano or learning to type basically comes down to learning scripts. The desire to play or type is the original cause of the repetition. Through practice, the awkward movements motivated by belief and desire become the rapid and precise actions of habit.

We can easily create these features of acting on habit. Under the assumption that you are a competent typist, reprogram your computer keyboard to switch the ‘r’ and the ‘t’ keys. After doing so, you will probably discover that, when you desire to type ‘the’, you will instead type ‘rhe’ more often than not.

Your act of typing ‘rhe’ is just as intentional as the act of typing ‘the’ was before you changed the keyboard. One can scarcely argue that you were thwarted by the false belief that the ‘r’ and ‘h’ keys were in their original position because you know that this is not the case. After all, we are assuming you are the one who reprogrammed the keyboard. Still, you are disposed to type ‘rhe’, but you desire to type ‘the’. That is to say, you are disposed to act in such a way that,

given your beliefs, you will type ‘rhe’. But we have no reason to type ‘rhe’. It is not something that you desire.

One might object that habits do not have the same responsiveness to beliefs that desires do. With a desire, the disposition to turn right or left would be dependent on my beliefs about where to catch the bus to go home. With a habit, the very example I use show that this is not the case.

We cannot “fix” habit and make it a desire by making it more belief – dependent.

Assume that our agent can make his habits more responsive to beliefs. All it takes is putting himself in the right stare of mind, and he can reprogram his habit on a single instance. So, when the old bus station closes, he puts himself into this state, turns left as he leaves the building, and instantly acquires a new habit he can use almost without thinking from that moment forward.

Similarly, our agent can put himself in a state, type “the” using the new characters, and acquire the new habit of using the new key for ‘t’ to type “the” from that moment on. Or, by putting himself in this state and playing a song once while reading the music, our agent can learn to play the song with the same habitual ease as any practiced professional.

We have made his dispositions to act more dependent on beliefs, but we have not transformed the habits into desires. These dispositions to act are still missing something.

I would suggest that what they are missing is the fact that I did not value turning right as I left work for its own sake. Nor do I value the set of motions that go into typing ‘the’ on a conventional keyboard as an end in itself. I have these dispositions to act, but these dispositions serve other ends. They serve my desire to get home, or they serve my desire to write a particular philosophy paper. Yet, they are not consciously done in the service to those ends. This is shown to be the case by the fact that if conditions remove them from being in service to those ends, then, even if the agent knows this, the agent performs the act anyway.

At this point, I simply want to suggest this diagnosis. I will have more to say about it later. For the moment, I want to look at another claim made about dispositional theories of value. This is the claim that somebody cannot desire something that they know to be the case.

The Death of Desire Principle

If a desire is a disposition to act, then what happens when there is nothing one can do? How can there be a disposition to act when there is no act that one is disposed to perform?

Dispositional theories of desire have a problem when there is nothing the agent can do to bring about that which, according to the dispositional theory, they are disposed to bring about. It seems to imply that it is impossible to desire something that it is impossible to bring about – or, at least, the agent thinks that it is impossible to bring about. It is impossible to desire something that cannot produce dispositions.

Chief among these is the notion that a person cannot be disposed to bring about that which the agent believes is already true. Consequently, when somebody obtains that which they desire, when there is nothing they can do to bring about such a state, the dispositional theory says that the desire disappears.

It is important to note that one can have a desire that something continue to be the case when it is the case. That is to say, somebody can be disposed to preserve a situation that one knows to obtain and, in virtue of this desire, be disposed to act in ways that will preserve it.

Fredrico Lauria (2017) calls this the "death of desire principle". He describes it as follows:

> *Imagine that Sam desires to see Niagara Falls. Mary offers to take him there. There they are, enjoying the breathtaking panorama. At some point, Sam says, "I want to see Niagara Falls." "Sam, you are seeing Niagara Falls," replies a quite surprised Mary. We understand Mary's astonishment. It is strange to express a desire to see something while in the midst of seeing it. Sam might express a desire to continue seeing the false, but this is a different desire than a desire simply to see the Falls. How could he desire simply to see*

> *the Falls while he is seeing them and is aware of his doing so? It appears that desire is incompatible with the representation that its content obtains.*

Dispositionalists seek to account for this "death of desire" principle by the fact that, if p is true, the agent cannot be disposed to act so as to realize p. While Sam is looking at the waterfall, Sam cannot be disposed to act to make it the case that he is looking at the waterfall. This, according to Lauria, is why it seems odd for him to say, while looking at the Falls, "I want to look at the Falls." It is because he cannot want to (desire to) look at the Falls if he is looking at them. On the other hand, he can want to *continue* to look at the falls. This is because he can be disposed to continue to look at the falls. However, this desire to continue to look at the Falls is a different and distinct desire.

I disagree.

To show what I think is wrong with this, I want to look at a different desire – a desire to have lunch. Agent is hungry. He has a desire to have lunch. To fulfill this desire, he goes to the store, buys food for a salad, takes it home, chops up the vegetables, adds salad dressing, and lifts the first fork full of salad to his mouth.

It would be odd to say that the desire that motivated him to prepare lunch suddenly vanished the instant he started to eat lunch, at which point a second and completely different desire to continue to eat lunch arose in its place. Rather, it was one and the same hunger that motivated him to prepare lunch, and then to continue to eat lunch after it was prepared.

Imagine, if you will, if this separate desire to continue to eat lunch never emerged. The desire to eat lunch vanished the instant "I am eating lunch" became true. At that point, the agent puts down the fork, puts the salad in the refrigerator, and then headed over to the couch to watch some television.

This description of how the desire to eat works not only seems intuitively wrong, but it has a poor fit with what we know about the biology of hunger.

One of the factors relevant to the desire to eat is the level of ghrelin in the blood. High levels act on the brain to promote hunger. As one eats, the stomach expands, and ghrelin levels drop.

The hunger that motivates the person to fix a meal to eat, motivates the agent to continue eating until full. It is, then, one and the same hunger that motivates the agent to prepare lunch, and then to eat lunch once it is prepared. There is no desire to eat lunch that lasts up to the point that lunch is prepared, to be replaced by a separate desire to continue to eat lunch.

This, then, provides an alternative account of what happens is happening in the case of Sam's desire to see the false. There is a single desire – a desire that can be expressed (from Sam's point of view) as a "desire that I see Niagara Falls". This desire motivates the agent to act so as to realize the desired state – to make true the proposition "I am seeing Niagara Falls." Then, once this proposition is made true, the same desire motivates the agent to keep this proposition true, at least for a while. As with hunger, as Sam gets his fill of the Falls (gets bored with it), his desire "that I see Niagara Falls" fades to the point where, eventually, he leaves the Falls to fulfill some other desire, such as a desire to have lunch.

This will explain the oddness of Sam, stepping up to see the Falls and, the very instant his desire to see the Falls is fulfilled, loses all interest in seeing the Falls and leaves for lunch. It will explain the oddness of Agent going through the effort to prepare lunch and the instant he lifts up the first fork full of salad, leaves his lunch to watch television. Desires do not die when the state that is desired is realized. They persist, and motivate the agent to preserve or continue that state.

So, how do we account for this oddness that Lauria mentions whereby it is odd for Sam to say, while he is looking at the Falls, that he has a desire to see the Falls? Lauria explained this by saying that he no longer had that desire – it died the instant he started looking at the Falls.

I suggest that it is like the oddness of talking about "the car I am going to buy" up until one buys the car, at which time referring to it as "the car I am going to buy" no longer makes sense. Once Agent purchases the car, he needs to refer to the car as "the car I bought". This does not imply that "the care I am going to buy" ceases to exist and an entirely different car – the "car I bought" suddenly pops into existence. There is only one car, but we refer to it in two different ways.

In the case of Sam having a desire "that I see Niagara Falls", we have a way of talking about this desire when the proposition, "I am seeing Niagara Falls" is false, and a different way of referring to that same desire when "I am seeing Niagara Falls" is true. When one is not φ-ing, we use the phrase, "I want to φ." When one is φ-ing, one uses the phrase, "I want to continue φ-ing." Yet, both statements refer to one and the same act of φ-ing – the act that is yet to occur (and might never occur) in the first instance, and the act that is occurring in the second.

As it turns out, the dispositional theorist can accept all of this without abandoning a dispositional theory of desire. She can say, "I was wrong. There is a "desire that p" can survive the realization of p because there are still dispositions to act. Before p is realized, the agent is disposed to make p true. After p is realized, the agent is disposed to preserve p. Combined, these make up a disposition to realize p. It is just, like all dispositions, realized in different ways under different circumstances."

Unfortunately, this does not work in a case where that which is done cannot be undone. We can use Graham Oddie's (2017) example of Clinton desiring to be first female President of the United States. If she had won the election, and made it through the inauguration, she would have been the first female President. The dispositional theorist would have to say that her desire that she be the first female President dies at that instant. She could not desire to continue to be the first female President because there was nothing she could do to prevent it from being true. Thus, it would not dispose Clinton to any actions.

However, there appears to be no reason to treat her desire "that I be the first female US President" any different from Sam's desire "that I see the falls" or my desire "that I have lunch". The fact that the former cannot be undone does not change the fact that she continues to value being the first female US president in the same way that Sam values continuing to look at the falls and I value eating the lunch that I have prepared simply because Sam can stop looking at the falls, I can stop eating lunch, but Clinton would have been unable to stop having been the first female President of the United States.

At this point, the dispositional theorist can assert that she is going to distinguish between valuing something (certainly Clinton could have values being the first female President of the

United States) and desiring it. Desiring would be distinguished from valuing precisely on the grounds that a desire requires at least the possibility of acting. This is why it makes no sense to say that Clinton either desires to or desires to continue to be the first female President of the United States – precisely because there is no possibility of acting.

However, establishing that desire requires the possibility of action is not the same as establishing that desire is a disposition to act. At this point, we have "valuing" doing all the work. It provides the motivation where action is possible and the other symptoms of valuing where action is not possible. It is "valuing" that makes the disposition an end or a goal, and the absence of valuing that makes other dispositions into habits. Everything that is important in a theory of desire now needs to be asked of a theory of valuing. Desiring only plays a subordinate role of "valuing when the possibility of action exists".

Above, I wrote about valuing "provides the motivation where action is possible and the other symptoms of valuing where action is not." That was my first mention of the "other symptoms of valuing". These other symptoms make it possible to have a different dispositional theory of desire – one that looks not just at dispositions to act, but dispositions to have these other symptoms as well.

Maria Alvarez (2017) provides a broad definition of dispositions that goes beyond dispositions to act. It includes dispositions to express and to feel. She wants to argue for a dispositional theory of desire in this broad sense, suggesting that desires, in fact, are "manifestation-dependent". They must show themselves in some way, or they do not exist.

Broad Dispositional Theory

I have discussed problems with the idea that a desire is a disposition to act. Habits are dispositions to act, but are not desires. Furthermore, we can value things we have no ability to change – that we have no ability to understand in terms of dispositions to act.

However, understanding desire in terms of a disposition to act may be too narrow. Maria Alvarez (2017) discusses desire in terms of a larger set of dispositions. It includes not only dispositions to act, but dispositions to express, and dispositions to feel. Nomy Arpaly and

Timothy Schroeder (2014) also discuss the wide range of influences that desires have, not only on action, but on other mental states such as feelings, thoughts, and attention. Note that Arpaly and Schroeder do not advocate a dispositional or action-based theory of desire. They argue against such theories. However, they do agree that the effects of desires is not limited to causing behavior.

This opens up the possibility of a dispositional theory that also is not limited to dispositions to act.

Alvarez introduced the idea that desires can manifest themselves in more than just action in a defense of the thesis that desires are, what she called, manifestation-dependent dispositions. There are no desires without at least one manifestation.

> *[T]here seem to be dispositions that are what might be called "manifestation-dependent": the absence of the manifestation over the lifetime of the object implies the absence of a disposition. . . My claim is that it is part of the concept of desire that someone has a desire at time t, only if the desire has been manifested in any of the various ways I described above at some point up to and including time t.*

She asserts that attributing to somebody the disposition of being a smoker if they never smoke, or felt any inclination to smoke, or felt frustration over the fact that they do not smoke, makes little or no sense. Similarly, calling a person generous who has never given anything of value for the sake of others or felt some sort of genuine regret that they could not have done something to help would be a mistake.

I think we can dismiss this version of her thesis.

The first reason to reject this thesis is that it seems quite possible for a person to have a desire or aversion that never manifests itself. I know, I said that I do no like exotic thought experiments, but this experiment is not as exotic as it might seem. Imagine, if you will, a species of intelligent life that has a deep aversion to spiders. Spiders on their home world are extremely deadly, and a genetic, inheritable aversion to spiders was such an aid to survival that it spread

throughout the species. They acquired an advanced level of technology, and eventually reached a stage where they effectively killed all of the spiders. This produced a massive environmental catastrophe of such a scale that this race entered into a dark age lasting thousands of years. When they emerged on the other end, they had lost all knowledge of spiders.

This story does not imply that these people lost their aversion to spiders. Indeed, an alien interstellar race who had been watching the people on this planet through the millennia can demonstrate that their aversion to spiders remains. All they need to do is drop a spider into a room where several members of this species had gathered for a birthday party and observe their reaction. We can test our hypothesis by bringing a spider to this world and dropping it into the middle of a birthday party and seeing how the natives react. Their innate, previously unmanifested aversion to spiders should be able to produce some predictable observable results.

In response to this type of counter-example, Alvarez writes:

> *First, if the object of desire is something the agent was not at all familiar with, then it is wholly implausible to suggest that the agent's reaction is a manifestation of a desire that was there all along.*

She says little more about this type of case. The hypothetical example of the spider seems to suggest that this is not the case. The object of the aversion is something that, ex hypothesi, the people at the birthday party have been entirely unfamiliar with. Yet, their reaction is a manifestation of an aversion that was there all along. Alvarez would have to argue that they could not continue to be averse to spiders unless they continued to be familiar with spiders. Nobody born ignorant of spiders and who remained ignorant could have such an aversion.

Against this, we can ask whether a person who is born without experiencing pain and who has not yet experienced pain can have an aversion to pain. If she has no aversion to pain until she has had her first experience of pain, then we have to ask: What caused her first reaction?

This leads to a second reason to reject her thesis: How would a desire manifest itself the first time if there is no desire before the first manifestation? If the child mentioned above did not

have an aversion to pain until its first manifestation, then what was it that caused her reaction to her first experience of pain?

The standard explanation is that humans are wired for an aversion to pain well before its first manifestation, and that it is the presence of this wiring that explains and predicts our first reaction to pain. It has to exist before the first manifestation, or there will not be a first manifestation.

We can ask other questions about this account. For example, is a single manifestation enough? Do we say that an agent has a desire that p if, for one fleeting moment, he had a manifestation and then it ceased? If a desire can go extinct, when does it go extinct? Does it go extinct with its last manifestation, or when it can on longer be triggered? Indeed, does a desire exist between manifestations? Or does it go extinct and then re-emerge with each new manifestation?

For my purposes here, I do not need to address these questions. It is sufficient that a desire is a disposition for certain types of behavior, expression, or psychological state to result from certain circumstances. The individuals of the alien species described above have an aversion to spiders just in case they are disposed to have an aversive reaction in the presence of spiders, even though they never see a spider. A child has an aversion to pain just in case if something caused the child to experience pain the child will not like it.

This goes along with what I said at the start of this section about what dispositional theorists got right. An agent has a desire that p just in case if she were to think that there was something she could do to realize p then she would be motivated to try it. She probably would not be sufficiently motivated to outweigh reasons not to try it, However, in the absence of such considerations, “Why not?” (Literally.) She can desire that p and think, “There is nothing I can do about p.” However, she cannot desire that p and think, “If there was something I could do about p I would be entirely indifferent about doing it.”

So, does a counterfactual dispositionalism – where to have a desire that p is to have a disposition to have some sort of manifestation (e.g., a reaction to the discovery that p is the case, an inclination to act so as to realize or preserve p if one were available) in the appropriate circumstances – allow us to save the dispositional theory of desire?

The problem with this move is that it is so broad that it is an issue not only for desires, but for everything. If something has no manifestations of any type, should we say that it does not exist, or that we do not have any reason to postulate its existence, or that we should accept claims about its existence anyway based on some special way of knowing? We have, in short, an issue that takes us far outside the issue of the best theory of desire.

The types of desires that we are interested in must at least counterfactually manifest themselves somehow. The question is: how? We have been looking at theories that say that desires are simply dispositions to act in particular ways. Those accounts of desire have been problematic. On this issue, I want to consider one last objection.

Multiple Desires

How do dispositional theories of desire handle multiple desires?

The versions of dispositional theories that we have looked at so far describe a single desire. An agent has a desire that p in virtue of having a disposition to act in a particular way under particular circumstances. However, the "particular circumstances" that determine whether a particular desire will manifest itself in behavior (and how) includes the other desires that the agent has at the time and the relationships between those desires, actions, and possible outcome.

It can get complicated very quickly.

What I want to do here is take seriously the "force" metaphor that is commonly used when talking about desire. We talk explicitly about motivational force and the weighing of different considerations in making a choice. Desires have a particular strength – where a weak desire can be easily resisted while a strong desire can be overpowering.

A force has two major components – a direction and a magnitude. In the realm of physics, we typically depict force as a vector – an arrow. The direction that the arrow points indicates the direction of the force, while the strength of the force is depicted using the arrow's length.

We can compute the sum of a set of vectors by putting them nose to tail – the nose or point of vector 1 being the starting point for vector 2, and its end point being the start for vector 3. We

can then compute the effects of all of these forces by calculating a vector sum – an arrow going from the starting point of vector 1 to the end point of the last vector.

If force 1 goes to the right with a strength of 5 units, and force 2 goes in exactly the opposite direction with a strength of 6 units, the sum is a vector going in the opposite direction of vector 1 with a strength of 1 unit.

If we apply this metaphor to desire, we can easily see that desires have a direction. A desire that p provides a force for realizing a state of affairs in which 'p' is true. Desires also have a strength – a measure of how important it is to the agent that 'p' be made or kept true. This will determine whether the desire will be outweighed by or outweigh any competing considerations that might come along.

If we look at dispositional theories of value, they count for the direction of desire – a desire that p is a disposition to realize p. However, there is nothing in their formulae that account for the strength of the desire. There is nothing that tells us how to weigh a desire against other considerations.

Imagine that you do have a balance and a set of weights. Try to develop a counterfactual account that will allow you to predict which way the scales will tilt as different weights are placed on the scale.

If there are only two weights, this is easy. Weight 1 will tilt the scale in its direction unless countered by weight 2, in which case the scale will tilt away.

If there are three weights, the issue becomes more complicated. Weight 1 will tilt the scale in its direction unless countered by weight 2, but not weight, though a combination of weight 2 and weight 3 will tilt the balance away from weight 1. However, if weight 1 and weight 3 are combined, then they will tilt the balance in their direction even if countered by weight 2.

With 4 weights, this becomes much more complex, and the complexity increases exponentially as the number of desires increases.

There are two more complications to consider. The first is that desires are not always strictly additive. If a person desires to spend some time with his friend and that he wants to see a movie, this does not imply that he wants to see a movie with his friend. After all, his friend might be an obnoxious companion at movies, constantly talking and interrupting.

Furthermore, some desires are variable in strength. A person's desire to eat something is much weaker shortly after finishing a meal than it can be expected to be four hours later. So, whether desire 1 (a desire to eat) outweighs desire 2 (a desire to play tennis) might depend on how long it has been since the last time the agent has eaten.

How are we going to use dispositions to explain and predict intentional behavior in the face of these complexities?

The point here is not that the dispositional theory is impractical. After all, there is no law of nature that says that everything we may want to do must be easy. The point is that our actual ability to explain and predict the behavior of others seems to be a lot easier than the dispositional model would predict. We can make estimates of how other people will behave based not only on what they want, but on how important those things are compared to other things the agent wants. We can create a mental model into which we feed in our understanding not only of the directions of the other person's desires, but their strength as well, and use this to produce useful results.

Since dispositional theories leave out a consideration of the motivational force or the strength of the desire, it cannot account for the ways in which we model the behavior of others in order to explain and predict their behavior. To accomplish this, they will need to add an element for the strength, motivational force, or weight of different desires.

This missing step of adding in an element for motivational force describes exactly the area where evaluation-based theories differ from dispositional theories. Evaluation-based theories assign a value to the object of a desire – an amount of goodness that the object of evaluaton is either believed to have or is perceived to have. Adding this missing component takes us out of dispositional theories of desire and lands us into evaluation-baed theories of desire. I want to look at this case next.

Summary

In this section, I have looked at theories that try to say that a desire is a mere disposition to act. I am holding this theory up against a rival that says that a desire is a state that attaches a particular importance to a state – an importance that translates into motivational force when an act that may realize or prevent the realization of such a state seems possible.

Dispositional theories have trouble distinguishing that which is desired from that which is done out of habit. Precisely what distinguishes these is the measure of importance given to that which is desired, but that we don't find in actions performed out of habit.

Dispositional theories have trouble talking about things that are important to us in the sense that we would like to see them realized when, in fact, there is nothing we can do to realize them. We can attach a measure of importance to things that have happened in the past, or to things that are going on at this very instant. Both sets of states are beyond our capacity to change (where there is no action we can be disposed to perform) but can still have value.

Dispositional theories do not fit well with the vector metaphor for desires that we use when we try to explain and predict the behavior of multiple desires. A dispositional theory would require that we consult a table identifying how different desires produce actions when placed in a context that included other desires. On the other hand, by looking at the importance of certain states we can look at where the greater motivational force is pointing. This allows us to explain why past behavior when the direction it did, and to predict the direction that future behavior will go.

There is a set of theories that look at desire as being concerned with the value of things. These include theories that take desire to be beliefs or perceptions as things as good (or bad) or as ought to be (or ought not to be). I will look at them next.

4. Evaluation-Based Theories of Desire

In the previous section, I looked at the idea that a "desire that p" can be understood as a disposition to realize p. I should make clear that a desire, in this sense, is not something that

disposes a person to realize p, but the disposition itself. As such, where there is no disposition, there is no desire.

I mentioned a number of problems with this way of thinking about desires.

First, dispositional theory fails to distinguish between desires and habits. Habits are dispositions to act, but to act out of habit is not the same is acting out of a desire.

Second, dispositional theory suggests that one cannot desire that which (one believes) already obtains, since we cannot have a disposition to bring about that which already obtains.

Third, dispositional theory leaves out a factor for the force magnitude or strength of a desire, which we use to explain and predict the behavior of agents having more than a few desires.

This suggests that we should look to the leading rival to dispositional theory, which are evaluation-based theories.

Evaluation-based theories are theories that assert that a desire that p can be understood in terms of p having some sort of value. The paradigm example of an evaluation-baed theory says that a desire that p is to be understood as a belief that p is good.

Such a theory easily handles the problems identified with dispositional theories. We can distinguish desires from habits precisely on the grounds that any relationship that exists between habit and what we value is contingent. We can value things that already exist. And we can relate the quantity of goodness to the strength of the desire, giving a better account of how we handle explaining and predicting people's behavior.

However, the common forms of evaluation-based theories have their own problems.

The most serious problem with evaluation-based theories is that they seem to require the existence of some sort of intrinsic value. At the very least, they assert a belief that the object of desire has intrinsic value (even if the belief is false) or the perception that the object of evaluation has value (even if it is an illusion). Either way, there is some sort of error involved in a desire that p.

In what follows, I want to say a little more about the type of evaluativist theories I will be considering in this section. After that, I wish to examine this problem of intrinsic prescriptivity in more detail.

Two Types of Evaluation

We can define the common form of evaluation-based theories into four groups based on two pairs of distinctions. The first distinction separates evaluativist theories that take a desire that p has value from those that take a desire that p to be a case of perceiving value in p. The second distinction separates theories that take the value found in b to be goodness from theories that take the value found in b to be ought-to-be-ness.

All four types follow the model of evaluation-based theories that Doring and Eker (2017) provides:

> *(ME) Necessarily, for any agent a, any proposition p, any time t, if, at t, a desires that p, then, at t, a evaluates p positively (as good).*

The first distinction applies to the idea of "evaluates" in the sense that, in one case, the evaluation is a belief or judgment whereas, in the other, the evaluation is a perception or a seeming (as in, "it seems to be the case that").

A belief is a cognitive state, like believing that the Statue of Liberty is in New York City or that water is made of H2O. The main difference between an evaluative belief and a traditional belief is that an evaluative belief comes with motivational force. It provides the agent with a reason for acting. At the same time, just as with other beliefs, the belief that p is good is subject to evaluation. We can ask if the agent has enough of the right type of evidence for her belief. In other words, agents can be asked to and expected to justify their values.

Perceiving that p is much like believing that p in that we can ask whether we are perceiving the object of evaluation correctly. We can ask whether there are aspects of the environment that may be distorting our perception. In the same way that a white object held under a red light can appear to be pink, a good object examined in the context of anger or envy can be perceived as being bad. We are invited to look around and see what might be distorting appearances.

The second distinction applies to the term “positively,” though, in fact, it applies to what we evaluate negatively as well. It is a distinction between the belief/perception that p is good, compared to the belief/perception that p ought to be.

The “guise of the good thesis” – the thesis that nothing is desired except insofar as it presents itself to the agent as good – has a very long philosophical pedigree. Health, the esteem of friends, a loving relationship, the raising of healthy and happy children, all of the things that we want out of life are things we also classify as good. The two are intimately related.

The “guise of the ought to be” thesis came about much more recently. It states that what an agent desires presents itself as something that ought to be. This “ought to be ness” relates the object of desire to action. If it “ought to be”, then it is something that the agent should bring about (to the degree that the agent has the power to bring it about and there is no conflict with other things that ought to be).

This, then, leaves us with four major variations of evaluativist theories of what it is to have a desire that p; as a belief that p is good, as perceiving that p is good, as believing that p ought to be, or as perceiving that p ought to be.

We will not need to be going into these theories in detail, since they all share a common problem. This problem concerns the “direction of fit”.

Direction of Fit

A common way of distinguishing beliefs from desires is in terms of “direction of fit”.

If a person believes that p, and p is false – that is, if ‘p’ does not describe the world correctly – then the proper way to bring the belief and the world into alignment against is to change one’s belief. If a person believes that raising the minimum wage will have no effect on unemployment, when in fact it will cost 50,000 jobs, then the best way to bring the world and his beliefs into alignment is for him to change his beliefs. Beliefs aim at describing the world correctly. Of course, if a person believes that a room is well illuminated, when in fact it is dark, can bring the belief and the world into agreement by turning on the light. That is to say, he can

change the world. However, it makes no sense for him to turn on the light if he believes the room is already illuminated.

In contrast, when an agent desires something, and the world does not conform to one's desires, the correct response is to change the world. If one has a desire for some chocolate cake, and one does not have any chocolate cake, the attitude to take is not one that says that the desire is mistaken. The correct attitude to take is to go get some chocolate cake. Insofar as we desire health, we have a reason to change the world so as to realize a state in which we are healthy. Insofar as we have an aversion to exerting ourselves and a desire for chocolate cake, then we have reason to sit on the couch eating a chocolate cake.

The most common versions of evaluativist theories get this relationship backwards with respect to desires.

This is easiest to see in the face of belief versions of the thesis; e.g., to desire that p is to believe that p is good. A belief that p is good is a belief that 'p is good' is true – that it is an accurate description of the world. In this way, the belief that p is good is like the belief that raising the minimum wage will not impact unemployment. It suggests that if one has a belief that p is good, and p is not good, then one should change one's beliefs. If one has a belief that p ought to be, and p does not have ought-to-be-ness, then one should change one's beliefs.

This raises the question: How are we to determine if p is good, or if p has ought-to-be-ness? What are we looking for? It cannot simply be that we desire it, because that would create a vicious circle. We desire p because p is good, and p is good because we desire it. Being good, or having ought-to-be-ness, are things that have to be true of the world independent of our desire, and something that we can discover. If they are not, then our beliefs are flawed, and we should change our beliefs.

Perception versions of these theories face the problem. According to the perception versions, to desire that p is to have it appear to be the case, to have it seem to be, and to have it look as if p is good or p ought to be the case. However, if p is not good, or p ought not to be the case, then the correct attitude to adopt is that our perceptions are mistaken. What we are perceiving is an illusion or a mirage. That is not to say that everything must appear the same to everybody

from every perspective. A cylinder can appear as a circle to somebody looking down from the top and as a rectangle for somebody looking at it from the side. Similarly, a song can appear as a moving melody to one person and as an unpleasant noise to another.

However, a cylinder appears as a circle to anybody who looks at it from the top – seeing a cylinder as a circle is the correct way to see a cylinder from the top. The person who looks at the cylinder from the top and who does not see it as circular is, in some way, defective. Similarly, two agents approaching a glass of wine from the same perspective would both have to have the same attitude towards the wine, or one of them would have to be considered defective. Not only is it the case that they have conflicting reasons to change the world. One of them has a reason that is mistaken – that is a wrong attitude to take towards wine from this perspective.

The evaluativist has a ready response to these types of concerns. "Of course desires can be mistaken. Pedophilia. Addiction. Obsessively counting blades of grass. These are cases in which a desire that p is mistaken precisely because p is not good, or ought not to be done, or lacks those qualities that make it a fit or appropriate object of desire. It is so easy to come up with examples of mistaken desires that you will be needing to provide some sort of argument to show that desires cannot be mistaken."

I will discuss one such argument in the next section. Briefly it says that the story of evolution leaves little room for the idea that desires can be correct or incorrect. All that matters, so far as evolution is concern, is that the desires generally work to bring about viable offspring. However, even though desires cannot be correct or incorrect, they can be good or bad. Or, on other words, people can have reasons to promote or inhibit certain desires.

Desires and Evolution

The story of evolution makes it unreasonable to believe that that our desires can be correct or incorrect in the same way that our beliefs or our perceptions can be. This is precisely because of the different roles that each play in behavior.

The evolutionary story of belief and perception provides a reason to believe that there is a mind-to-world direction of fit for beliefs. If a lion is going to catch an antelope for dinner, it would be useful if the lion can accurately see the antelope – see where it is, and that it is an antelope (or, at least, that it is something to kill and eat for dinner). At the same time, it is useful for the antelope to be able to tell when there is a lion in the area, and in which direction to run if it is to avoid becoming dinner. This does not eliminate the possibility of errors – even useful errors. However, it does suggest that, insofar as we believe correctly, our beliefs provide us with an accurate description of the world.

On the other hand, we can expect that the evolution of desires to have worked somewhat differently from this. We can imagine evolution randomly assigning desires to creatures – an aversion to crowds over here, a preference to killing one's step children over there, a preference to eat a food that happens to be nutritious versus a fondness for the taste of a food that leaves the creature sterile. What makes a particular desire "correct" or "incorrect" is not that it accurately identifies some value quality in the world. What matters is that the creature with the desire (and that acts on it in its environment) produces viable offspring that, in turn, produce viable offspring.

Evolution explains a part of our acquisition of ends, but it does not provide the ends. We cannot look to evolution to determine if a desire is correct or incorrect. Evolution has sculpted the aversion to pain so that it leads to the avoidance of bodily damage that could interfere with survival and procreation. However, the end that the aversion of pain picks out is the avoidance of pain, not the prevention of boldly damage that could hinder survival and procreation. The end that the desire for sex picks out is not procreation, but to have sex, though evolution favored those who had this desire because they did tend to procreate. The objective of hunger – the desire to eat – is to eat. The objective of nurturing one's children is to have healthy, well-functioning children. The value of being a respected member of the tribe is found in being a respected member of the tribe, not in the evolutionary advantages that spring from membership.

All of this suggests that our desires could have been different. The desires we acquire are the results of random genetic mutation interacting with a contingent environment. After all, evolution has not only given us our aversion to pain and concern for our offspring. Every predator and parasite has an evolutionary history. It has made tigers into effective killers, unlikely to see any aversion to inflicting pain on the antelope. It has created brood parasites – birds that lay their eggs in the nests of other birds, who are then disposed to raise (and invest resources in) these chicks.

It would be a mistake to say that all of our desires are strongly genetically determined – that is not true. We have desires that form through an interaction between genes and environment. A particular interaction between genes and environment forms our sexual orientation. Genetics influences the amount of testosterone in the fetus' environment, and the amount of testosterone present influences the development of the brain to as to dispose the resulting person to being attracted to males or females. This process is separate from the process that influences development of the body, allowing for the existence of individuals with a male body attracted to males, and with a female body attracted to other females. This happens when other factors influence the amount of testosterone or effects of testosterone on the body or the brain.

A third and extremely important source of desires is through the functioning of the reward system of the brain. Roughly speaking, the reward center takes input in terms of experiences with one's environment and produces rules of behavior. An individual that is stung by a bee not only learns a means-end relationship between avoiding bees and avoiding pain, it acquires an aversion to avoiding bees. On other words, the individual will come to avoid bees even when the individual is fully aware of the fact that he is not at risk of suffering pain. He "just doesn't like bees." Means-ends reasoning takes a fair amount of cognitive power. On the other hand, an acquired desire or aversion accomplishes nearly the same result without the need for sophisticated cognitive capacities. The person with an aversion to bees can avoid the consequences that bees may produce without the sophisticated understanding of the relationship between bees and pain.

The acquisition of desires from experience means that different individuals can have different desires and aversions without it being the case that one views the object correctly and another incorrectly. Two people can be otherwise quite similar, and yet one of them can enjoy Mozart while the other finds it insufferably boring. Furthermore, the same individual could have acquired different desires and aversions based, not on differences in the qualities of the things desired but based on her experiences. A parent who praised and encouraged her drawings could have produced somebody who liked to draw whereas a parent who denigrated her interest in art as a waste of time may have produced somebody with a different set of interests. The version with an interest in art does not see a goodness in the production of drawings that the other fails to see, but simply has a brain that is wired to give a different value to producing art.

In fact, there are situations where people have reason to prefer a diversity of desires – different people having different degrees of interest in the same thing, rather than all of them seeing the same thing as good or bad, or ought to be or ought not to be. We see this in the set of examples that go into the choice of profession. It is a good thing that some people find value in being teachers, others find value in being physicians, still others like engineering, and others prefer growing food or manufacturing tools or driving a truck. If we all had the same preferences, we would have to suffer in a world where a few people got the few jobs that were at the top of everybody's list and everybody else had to endure a job further down the list. Instead, with a diversity of interests, we can have a situation where different people get a job nearer the top of their list and still obtain the benefits of different people doing different things.

These facts fit uncomfortably with the idea that desires contain some sort of "mind to world" direction of fit – with the idea that a desire that p can be correct or incorrect. With any true proposition p, there is only one correct belief-attitude that one can have towards p, and this is that p is true. However, there is nothing in any proposition p that dictates a correct desire-attitude to have towards p. Whether it becomes the object of a desire or aversion depends on the evolutionary history of the creature that has come to desire it. This depends, in part, on random chance – the random genetic mutations that came about. It depends on the contingent

facts regarding the environment in which those dispositions emerged. It depends on luck. If it came about through the activation of some type of learning system, it depends on the random and contingent facts that tuned that learning system, and on the contingent facts that defines the creature's experiences. These determine the ways in which the creature seeks to change the world. There is no fact of the matter that makes any particular desire correct or incorrect.

If the creature has a belief that p is good or that p ought to be then either it is a belief that p is something the agent wants to realize, or the belief is false. If the creature perceives goodness or oughtness in p, which is something than perceiving that he wants it to be the case that p, then his perception is an illusion.

The direction of fit goes from the desire that p to the goodness or oughtness of p, and not from the latter to the former.

Argument from Neuroscience

Arply and Schroeder (2014) provide another argument against evaluation-based theories of desire based on neuroscience. They criticize the idea that desires are cognitive states (beliefs or perceptions that something is the case) by arguing that a study of people with brain damage is inconsistent with this hypothesis. Specifically, damage to the cognitive portions of the brain does not impair the brain's capacity to produce the types of symptoms characteristic of desire, such as motivation to act or a tendency to feel particular emotions. On the other hand, damage to the reward system does have these defects.

Parkinson's disease destroys cells in the reward system – the parts of the brain that uses dopamine transmitters to process rewards and punishments. This disease leaves the cognitive portions of the brain intact. Parkinson's disease does not impair a person's ability to form judgments or beliefs. However, it does impair a person's ability to act on those judgments or beliefs. This is a problem for a theory that identifies desires with a belief-dependent or a judgment-dependent action and gives the advantage to a theory that locates desires somehow in the functioning of the reward system. The reward system picks out what to do, not the belief or judgment system.

> *An important fact about the paralysis that can result from Parkinson disease is that it is not relieved by beliefs about what is right or good or reasonable. No such belief suffices to restore the capacity to act on its own, independently of what is constituted as a reward (i.e., intrinsically desired).*

Against this argument, Gideon Yaffe (2013) suggests that the reward centers of the brain are, themselves, seats for certain types of judgments.

Schroeder (2017) responds that if this were the case, then we would have to interpret Parkinson's Disease as a disease that destroys people's capacity to make these types of judgments – as people who cannot judge that they have a reason to perform some action.

More importantly, Schroeder argues that the reward centers of the brain – unlike the cognitive centers – are poorly connected to the parts of the brain having to do with memory and consciousness. Consequently, we would have difficulty being conscious of making a particular judgment or of recalling a judgment from memory. If we tend not to have these types of difficulties, then this suggests that these beliefs or judgments are formed in the same parts of the brain as other beliefs or judgments. Yet, if this were the case, we return to the problem that beliefs or judgments formed in these parts of the brain seem incapable of producing action – that is to say, people do not act because they have the belief or the judgment. Either we cannot be conscious of or remember these judgments, or they cannot cause action.

Against Reward Theory

I have made extensive use of Arpely and Schroeder's empirical data in this paper. However, I have to object to their proposed solution to these problems.

Arpely and Schroeder defend what they call a "Reward Theory of Desire".

According to this theory:

> *Reward Theory of Desire: to have an intrinsic desire regarding it being the case that P is to constitute P as a reward or a punishment. To have an intrinsic appetitive desire that P is to constitute P as a reward. To have an intrinsic aversion to P is to constitute P as a punishment.*

This theory seems to suggest that it would not be possible for an agent to have a desire or an aversion that lacks the ability to influence the formation of malleable desires by serving as a reward or punishment. By definition, researchers cannot discover the existence of a hard-wired desire, that provides an agent with a reason to perform or to forbear from some action – even though it is not the case that obtaining more (or less) of that which is desired than expected also generates a reward signal.

In contrast, it seems quite possible that nature provided creatures with an aversion to pain, a desire to eat and drink, and a desire to seek a comfortable environmental temperature, long before we had a reward system. The problem with these hard-wired desires is that the creature's suitability to a particular environment would have been fixed. Either those desires and aversions kept it alive in a given environment, or it did not. A mutation at this point that altered the strengths or the objects of these desires and aversions based on experience could provide an advantage in allowing the creature to acquire behavioral dispositions suitable for a wider variety of conditions, or to adapt to changing conditions. The malleability of desires may well have come after there were desires to be made malleable.

I am not claiming that this represents our actual evolutionary history. I am arguing that it is an option worthy of investigation. Yet, Arpely and Schroeder's reward theory of desires do not allow for this possibility. On their account, desires could not have existed unless and until there was a reward system with the capacity to alter desires that did not yet exist. In order for p to constitute a reward or punishment, there must be a reward or a punishment system. If there is no reward or punishment system, then nothing could constitute a reward or punishment. This implies, on Arpely and Schroeder's theory, it would not have been possible to have an intrinsic desire.

I also want to stress that the existence of a problem with Arpely and Schroeder's theory of desire does not imply that the evaluation-based theories they criticize are out of the woods. This is no truer than it would be to say that the problems with evaluation-based theories implies that the reward theory is out of the woods.

5. Assignment Theory of Desire

6. References

Alvarez, Maria (2017), "Desires, Dispositions, and the Explanation of Action". In Deonna J. & Lauria F. (eds). *The Nature of Desire*. Oxford: Oxford University Press.

Anscombe, G.E.M. (1963), *Intention (Second Edition)*, Basil Blackwell, Oxford: Oxford University Press.

Arpaly, N. and Schroeder, T. (2014). *In Praise of Desire*, New York: Oxford University Press.

Davidson, Donald, 1963, "Actions, Reasons, and Causes", *The Journal of Philosophy*, Vol. 60, No. 23, American Philosophical Association, Eastern Division, Sixtieth Annual Meeting (Nov. 7, 1963), pp. 685-700

Blackburn, Simon (1985), "Errors and the Phenomenology of Value" in Honderich, Ted (ed.), *Morality and Objectivity: A Tribute to J.L. Mackie,* Routledge and Kegan Paul.

Doring, Sabine A. and Eker, Bahadir (2017), "Desires without Guises". In Deonna J. & Lauria F. (eds). *The Nature of Desire*. Oxford: Oxford University Press.

David Hume, 1896, *A Treatise of Human Nature* by David Hume, reprinted from the Original Edition in three volumes and edited, with an analytical index, by L.A. Selby-Bigge, M.A. (Oxford: Clarendon Press, 1896). 7/8/2018. http://oll.libertyfund.org/titles/342.

Hursthouse, R. (1999). *On virtue ethics*. Oxford: Oxford University Press.

Lauria, Fredrico (2017), "The 'Guise of the Ought-to-Be': A Deontic View of the Intentionality of Desire". In Deonna J. & Lauria F. (eds). *The Nature of Desire*. Oxford: Oxford University Press.

Mackie, J.L. (1977), *Ethics: Inventing Right and Wrong*, Penguin Press

Nozick, Robert (1974). *Anarchy, State, and Utopia*. New York: Basic Books

Oddie, Graham (2017), "Desire and the Good: In Search of the Right Fit". In Deonna J. & Lauria F. (eds). *The Nature of Desire*. Oxford: Oxford University Press.

Quinn, Warren (1993), "Putting Rationality in its Place", in *Morality and Action.* Cambridge, UK: Cambridge University Press.

Schroeder, Timothy (2004). *Three Faces of Desire,* New York: Oxford University Press.

Schroeder, Timothy (2017), "Empirical Evidence against a Cognitivist Theory of Desire and Action". In Deonna J. & Lauria F. (eds). *The Nature of Desire*. Oxford: Oxford University Press.

Smith, Michael (1987), "The Humean Theory of Motivation", *Mind*, Vol. 96, No. 381, pp. 36-61.

Smith, Michael (2012). "Four Objections to the Standard Story of Action (and Four Replies)", *Philosophical Issues, Vol. 22, Action Theory*, pp. 387-401.

Stanford Encyclopedia of Philosophy, "Desire", First published Wed Dec 9, 2009; substantive revision Thu Apr 9, 2015, https://plato.stanford.edu/entries/desire/.

Stanford Encyclopedia of Philosophy, "Reasons for Action: Internal vs. External", First published Thu Sep 4, 2008; substantive revision Fri Aug 18, 2017, 7/8/2018, https://plato.stanford.edu/entries/reasons-internal-external/.

Street, Sharon, "A Darwinian Dilemma for Realist Theories of Value," Philosophical Studies 127(1):109-166

Yaffe, Gideon (2013), "Are Addicts Akratic?" in N. Levy (ed.) *Addiction and Self-Control*, New York, Oxford University Press, 190-213.