ATTRIBUTABILITY AND AGENCY: MORAL ATTRIBUTABILITY FOR MENTAL STATES AS POSSESSION OF CARE-CONSTITUTIVE DESIRES

by

Thomas Yamilkoski

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

in Philosophy

at

The University of Wisconsin-Milwaukee

May 2018

ABSTRACT

ATTRIBUTABILITY AND AGENCY: MORAL ATTRIBUTABILITY FOR MENTAL STATES AS POSSESSION OF CARE-CONSTITUTIVE DESIRES

by

Thomas Yamilkoski

The University of Wisconsin-Milwaukee, 2018
Under the Supervision of Professor Edward Hinchman

A prominent line of thought owed originally to the work of Harry Frankfurt is that it is our identifying, in a certain technical sense, with our mental states which makes these states and the actions which emerge from them our own in a way distinctive of agents. Separately, moral attributability, a sort of responsibility located first by T. M. Scanlon, has recently attracted the attention of many philosophers. In this paper I will argue that we ought to aim to adopt theories of identification and moral attributability such that our capacity for the sort of agency involved in identification is a precondition for our capacity to have mental states attributable to us. Motivated by this point and the prominent line of thought that we identify through our cares, I will develop an account of moral attributability through caring, for which I argue it is plausible such an explanatory relationship can exist between our capacities for identification and attributability.

For both the family with me today and the family who have passed on before me.

TABLE OF CONTENTS

1.1: Introduction 1

2.1: Historical Motivation for Modern Reductive Accounts of Freedom of Action 4

2.2: Modern Accounts of Freedom through Identification 7

3.1: Moral Attributability 17

3.2: Attributability: Smith's Rational Relations View 21

3.3: Motivating a Care-based Approach to Attributability Further: The Prototype Account 24

3.4: Attributability as Sensitivity to Norms of Care-Constitutive Desires 28

3.5: Objections 33

4.1: Autonomy, Attributability, and Accountability 36

References 40

1.1: Introduction

Over the course of the last forty-or-so years, a certain literature has developed around characterizing the difference between mental states and actions which belong to us as persons, or we have or perform freely, and those which do not really express our characters. This literature holds that it is our *identifying*, in a certain technical sense, with our mental states which makes these states and the actions which they produce truly our own.[1] While various philosophers have tried different conceptions of identification toward characterizing these phenomena, in recent work on the subject Frankfurt, Agniezka Jaworska, and David Shoemaker have argued for accounts of identification which hold, through one mechanism or another, that it is our cares which allow us to identify with our mental states in this way which allows us to act freely, as agents.[2]

Separately, T. M. Scanlon in his "*What We Owe Each Other*" has located a certain sort of responsibility, responsibility-as-attributability, around which much literature has developed.[3] On Scanlon's account of this sort of responsibility, to say that something is attributable to an individual "is only to say that it is appropriate to take it as a basis of moral appraisal of that person."[4] Since Scanlon put his finger on this notion, much literature has emerged around characterizing this sort of responsibility and delineating it from others.[5] Two pertinent accounts,

---

[1] See Frankfurt's *Free Will and The Concept of a Person*. (1971)
[2] See Frankfurt's *The Importance of What We Care About*, Jaworska's *Caring and Internality*, and Shoemaker's *Caring, Identification, and Agency*. (1982, 2007, and 2003 resp.)
[3] Scanlon, 2000.
[4] Scanlon, 248.
[5] See, for instance, David Shoemaker's *Attributability, Answerability and Accountability: Toward a Wider Theory of Moral Responsibility*. (2011)

Scanlon's original proposal and a later account proposed by Angela Smith, attempt to explain the conditions of attributability, broadly speaking, in terms of norms of judgment.[6]

These approaches might be seen to represent two different methods of locating instances in which an agent is responsible for something: identification concerns whether what belongs to a person emerges in some correct way from their character, while attributability is interested in those conditions in which we, in practice, can appraise others as responsible for their actions. As can be easily demonstrated, however, we simply do, in practice, often hold others responsible in entirely mundane ways for both mental states and actions which do not emerge from their character in any way that is required for autonomy. The conditions of identification, in other words, differ from those of attribution.

One might view this as some sort of disagreement between our intuitions of what it is to be responsible, but I don't approach the issue this way in this paper. Instead, I attempt to answer another question, namely: what is the relationship between the things which belong to us as agents and those that are morally attributable to us? Beyond a vague feeling that there must be some such relationship, here is an argument that we ought to think there does exist some such connection: Only of persons can the sort of *aretaic* appraisal involved in morally attributing something to someone, one of them as a moral agent, be made. This suggests that the mental states which qualify somebody as an agent, at least *capable* of having mental states and acting freely, should have a significant part in explaining how our mental states and actions *can be* attributed to us. That is, given accounts of free agency and responsibility-as-attributability, it would be a significant theoretical boon for each (and both, taken together) if it were the case that

[6] Smith, 2005.

our *capacity* for this sort of agential freedom would be able to somehow explain our *capacity* for having mental states attributable to us.

Smith and Scanlon's own accounts, again, appeal to judgments and judgment-sensitive attitudes to explain moral attributability. But while Gary Watson proposed early in the course of the relevant literature that our identification with motives is a matter of our judging certain courses of action best, this account failed, and a significant amount of recent literature embracing the idea that it is instead our cares which make actions and mental states belong to us in this way crucial to our freedom has emerged. Speaking very broadly, if the attributability of a state to us is a matter of it having some relationship to our judgements, as Smith and Scanlon hold, it is at best unclear how we should be able to provide an account of this link between our capacities for agency and responsibility the above argument makes seem desirable.

In this paper, I motivate and subsequently provide an alternate account of moral attributability which is complemented in this way by the existing literature on identification-as-caring. What I argue is that the attributability of certain mental states and deficits thereof to us is a matter of our having care-constitutive desires either for or against taking these mental states. The truth of any deterministic thesis aside, on this account our responsibility for our actions lies not in the fact that we are radically free in any libertarian sense, but in the fact that we are emotionally tethered to other entities[7] in the world.

I do not argue that my operative conception of care is capable of characterizing also our identification with mental states and actions, as this could only be attempted as part of what would have to be a larger-scale project. It is desirable, as the above argument shows, that we should have conceptions of identification and attributability which allow us to explain our

[7] In the most general sense.

capacity for attributability in terms of our capacity for identification. What I hope to show in this vein is that it is *plausible* that we should be able to do this, at least for the case of mental states, through an account of moral attributability as susceptibility to norms of care.

In section 2.1, I provide a summary of the history of accounts of freedom of action which motivates appealing to the notion of identification under discussion. In section 2.2, I give a brief survey of modern accounts of identification, including three separate care-based accounts of identification developed by Frankfurt, Jaworska, and Shoemaker in their respective works.[8] In section 3.1, I discuss moral attributability and motivate an account of attributability through care for which our capacity for attributability can be explained through our capacity for agency. In section 3.2, I discuss and raise objections against a competing account of moral attributability for mental states fielded by Angela Smith.[9] In section 3.3, I propose and reject a prototypical account of attributability for mental states which suggests it is our higher-order desires which make certain mental states attributable to us. In section 3.4, I argue that the flaws of the higher-order desire-based account of attributability can be addressed through appeal to instead care-constitutive higher-order desires and provide my operative account of care. In section 3.5, I discuss two potential objections, and in section 4.1, I return to my claim that our capacity for this kind of attributability can be accounted for by our capacity for agency and give closing remarks.

2.1: Historical Motivation for Modern Reductive Accounts of Freedom of Action

In *Freedom and Action*, Roderick Chisholm considers an argument for a compatibilistic theory of responsibility used by "many philosophers in [the twentieth century], most notably, G. E.

[8] See Frankfurt's *The Importance of What We Care About*, 1982, Jaworska's *Caring and Internality*, and Shoemaker's *Caring, Identification, and Agency*. (1982, 2007, and 2003 resp.)
[9] See her *Responsibility for Attitudes: Activity and Passivity in Mental Life*. (Smith, 2005)

Moore."[10] Suppose, as Chisholm does, that one man shot another. If some thesis of determinism is true, then it will seem to be the case that this individual could not have done otherwise than to shoot his victim, and the incompatibilist will on the basis of this claim he is not responsible for the act. What the classical compatibilist argues according to Chisholm is that the expression 'he could have done otherwise' in this context means only that 'if he had chosen to do otherwise, then he would have done otherwise.' Supposedly, then, the truth of this second expression is compatible with determinism, and so the truth first must be as well, their meanings being the same. Thus the shooter, the classical compatibilist argues, in fact could have done otherwise, and the incompatibilist's concerns are thus in some way confused. As Chisholm says, what the incompatibilist will deny of this argument is that is that 'he could have chosen otherwise' is true in any sense which licenses the inference from the first sentence to the second. If determinism is true, then our choices themselves will still have their roots in causal chains it is impossible to track back to any entity that is in a literal, nonreductive way 'us'. Given the truth of a thesis of determinism, then, it still seems unclear how an individual should ever be able get behind their actions in any way that they could properly be called responsible for them.

Suppose instead that determinism is untrue, and every event doesn't have a cause. "If the act... was not caused at all, if it was fortuitous or capricious, happening so to speak "out of the blue," then presumably no one, and nothing was responsible for the act."[11] To summarize, if determinism is true, then it is unclear how what seem to us to be our actions are ever more than the product of causal forces flowing 'through' us. On the other hand, if determinism is untrue and our actions instead arise without causes, then it is again unclear how we could ever be

[10] Chisholm, 1996, 15.
[11] Chisholm, 16.

responsible for our acts. So whether our actions emerge as the products of deterministic chains or they occur without causes, it seems as though we cannot get behind our actions in any way that could make us responsible for them.[12]

The radical solution Chisholm proposes at this impasse to save our responsibility is that "at least one of the events that are involved in [our acts] is caused, not by any other events, but by something else instead. And this something else can only be the agent—the man."[13] In other words, there must exist agents with the capacity for what he calls agent causation, or the ability to produce events in what we can only take to be a radically free way. As Harry Frankfurt describes Chisholm's theory, perhaps slightly unfairly, "Whenever a person performs a free action, according to Chisholm, it's a miracle."[14]

To be slightly fairer to Chisholm, he does make what I take to be deeply interesting arguments in defending his appraoch. But many philosophers do not take Chisholm on this matter seriously,[15] and I myself am skeptical of his notion of agent causation. I will however not be debating Chisholm's point. What is important to understand is that if we are unwilling to countenance the sort of agent causation which Chisholm proposes then there is a price to pay. We must accept that the actions we take are merely some combination of the product of causal forces which in some sense seem to only 'flow through us', or else emerge 'capriciously', without any order corresponding to that we tend to naïvely think we exercise when we act. The burden

[12] The above is adapted from Chisholm, 15-6.
[13] Chisholm, 17.
[14] Frankfurt, 1971, 18.
[15] As has been remarked by David Velleman in his *What Happens When Someone Acts?* (Velleman, 1992, 469)

which falls on us is then to produce a *reductive* account of our sense that we are agents who are capable of bringing about changes in the world, consistent with these possibilities

2.2: Modern Accounts of Freedom through Identification

The business of such a reductive account of Chisholm's phenomenon is, then, while accepting that our actions may have roots in causal chains or else be 'capricious happenings', to delineate instances in which we might naively regard as *persons* acting, unimpeded by *mere c*ausal forces or spurious events which might happen to 'flow through' them. Given the nature of this dichotomy between acts with their roots in mere 'outside' forces and in persons, we might couch this task in more mundane terms as that of describing when we *act freely*.

Much literature in the business of explicating this distinction has emerged since Chisholm's paper, the most seminal of which is likely Frankfurt's *Free Will and the Concept of a Person*.[16] Frankfurt's language in this paper might initially suggest to us he is interested only in freedom of the will, as he discusses what he calls 'freedom of action' rather less. In this work, however, by 'freedom of action' Frankfurt means only "the freedom to do what one wants to do."[17] This is closest to a classical conception of free action, utilized in the argument Chisholm attributes to classical compatibilists above, and not the more recently developed sense concerning a fuller notion of autonomy which has come to have an important place in this literature.

[16] Frankfurt, 1971.
[17] Frankfurt, 1971, 15.

What Frankfurt terms 'freedom of the will' in this work is in fact more closely associated with this fuller conception of autonomous action than what he terms freedom of action.[18] An important trope in this literature is that of the unwilling addict:

> It makes a difference to the unwilling addict, who is a person, which of his conflicting first-order desires [the desire to shoot up or the desire to not shoot up] win out. Both desires are his, to be sure; and whether he finally takes the drug or finally succeeds in refraining from taking it, he acts to satisfy what is in a literal sense his own desire In either case he does something he himself wants to do, and he does it not because of some external influence whose aim happens to coincide with his own but because of his desire to do it. The unwilling addict *identifies* himself, however, through the second-order volition, with one rather than the other of his conflicting first-order desires. It is in virtue of this identification and withdrawal, accomplished through the formation of a second order volition, that the unwilling addict may meaningfully make the analytically puzzling statements that the force moving him to take the drug is a force other than his own, and that it is not of his own free will but rather against his will that this force moves him to take it.[19]

We can see Frankfurt's theory as presented in Frankfurt, 1971 as the first which embraces a technical notion of *identification* with desires in order to provide accounts of free will and action.[20] Cases in which an agent is identified with a desire and resulting action are most easily contrasted with that of the unwilling addict, in which we might say a person's 'acts' are produced by what we would regard as *merely* by causal forces, which are in some way intruding on how the *agent themself* wants to act. Besides the cases such as the above in which we are alienated from our desires, we also other times simply fail to identify nor be alienated from our desires or actions, such as when we desire in an idle moment to flick away a speck of dust. In contrast, in the case that an agent's actions do emerge from desires with which he is identified then we can

[18] Although I think this should be uncontroversial, Gary Watson in his *Free Will and Free Action* would seem to generally be agreement with me on this matter. (Watson, 1987)
[19] Frankfurt, 1971, 13; italics mine.
[20] As I hope is clear, I mean free action in the non-classical sense.

according to Frankfurt take them to be both willing and acting freely in the fuller but reductive senses this debate is interested in characterizing.

We might take this notion of identification to apply to mental states more broadly, as Jaworska does,[21] taking it also to locate others sorts of mental states we in a similar sense have freely, and do not merely 'flow through us'. While I discuss this possibility more later in this work, much of the literature on identification, including Frankfurt, 1971, is interested only in identification with desires, being more interested in the questions of when we act and will freely. However, if we already appeal to a notion of identification with desires, of their 'really belonging' to an agent, to make sense of the cases of free action and will, it is difficult to see why one would deny that we also identify with other sorts of affective mental states as well.

Returning to discussion of Frankfurt, the criterion Frankfurt famously first proposed for an agent to be identified with a desire that this desire must be the object of one of their "second order volitions".[22] This means just that an agent must have a second order desire not only to have this desire, but also to have this desire as their active desire, or, by Frankfurt's usage, their will, in order to identify with it in the pertinent way. It is our second-order desires to take other desires as our will, according to Frankfurt, which distinguishes us as persons and allow for us to act as such.

As is well known, Gary Watson in his *Free Action and Free Will* quickly took issue with this account of identification.[23] His most fundamental criticism is that "we can see that the notion of 'higher-order volition' is not really the fundamental notion for these purposes, by raising the question, "[can't] one be a wanton, so to speak, with respect to one's second order desires and

[21] See Jaworska's *Caring and Internality*. (Jaworska, 2007)
[22] Frankfurt, 1971, 10.
[23] Watson, 1975.

volitions?"[24] Or, alternatively, as Irving Thalberg expresses the worry, "Why not go on to third-story or higher volitions?"[25] It seems entirely arbitrary as to why our second-level volitions should represent us, and not, say, our third or fourth.

To some extent, it seems Frankfurt anticipates this crucial objection even in his original work, and he does later make fuller arguments with the aim of skirting this objection, to be discussed shortly. Besides merely raising this objection however, Gary Watson in Watson, 1975 also argued for his own, radically different account of this sort of identification, on which we identify instead through our evaluative judgment. As Watson helpfully puts it in discussing an example case, "it is because [a kleptomaniac's] desires express themselves independently of his evaluative judgments that we tend to think of his actions as unfree."[26]

One reason why Watson's account is so notable, besides the fact that it occupies such an early place in the literature, is that it is of a decidedly non-Humean bent. At the end of *Free Agency*, Watson quotes Nietzsche as writing "Man's belly is the reason why man does not easily take himself for a God."[27] The meaning Watson intends to convey is clear: rather than in some way constituting our identification with our psychic states and thereby our actions, as Frankfurt claims it does, he means to suggest our affect only *obstructs* our free action instead.

If we appropriate this meaning to Nietzsche's aphorism, then we might take the ensuing literature on identification as good evidence that he was incorrect. As Watson now acknowledges, there exist what he has come to call "perverse cases", or cases in which one acts contrary to their evaluative judgments of what it is best to do. "I might fully 'embrace' [and

---

[24] Watson, 1975, 349.
[25] See Thalberg's *Hierarchical Analyses of Unfree Action*. (Thalberg, 1978, 219)
[26] Watson, 1975, 351. Such an (implicitly reluctant) kleptomaniac of course fills the same role in Watson's argument as the unwilling addict did in Frankfurt's.
[27] Watson, 1975, 251.

identify with] a course of action I do not judge best; it may not be thought best, but is fun, or thrilling; one loves doing it, and it's too bad it's not also the best thing to do, but one goes for it without compunction."[28] Even beyond identifying with desires and courses of action we do not judge best, however, we can also even be alienated from our values *themselves*. As Velleman observes, it is not difficult to imagine an individual recoiling from his own materialistic values.[29]

While there have been no attempts to save Watson's 'rationalist' approach to free action as far as I am aware, Frankfurt alone has on the other hand published a mass of literature on the importance of higher order desires and also later care to our free agency, and others have done so after him. Returning to Frankfurt's account, then, if we interpret Frankfurt in his 1971 work as I have suggested, in the face of Watson's criticism the account he presents seems clearly lacking. Frankfurt to a certain extent does, however, anticipate the arbitrariness objection originally raised by Watson. What Frankfurt suggests near the end of this work to really be important to the issue of identification with desires is that one must instead somehow actively desire a first order desire such that "this commitment "resounds" throughout the potentially endless array of higher orders."[30] It is difficult to determine exactly what Frankfurt means here, and he himself in a later work admits that the language in the few paragraphs in which he describes this solution is "terribly obscure".[31]

In his more recent works *Identification and Externality* and *Identification and Wholeheartedness* Frankfurt builds upon this earlier, rather oblique suggestion.[32] His comments in the former of these two works, which is shorter and the earlier of the two, are more cautious,

---

[28] Watson, 1987, 150.
[29] See Velleman's *What Happens When Someone Acts?* (Velleman, 1992, 472)
[30] Frankfurt, 1971, 16.
[31] See Frankfurt's *Identification and Wholeheartedness,* page 167. (Frankfurt, 1987)
[32] Frankfurt, 1976 and Frankfurt 1987.

so I here discuss only his account as he presents it in *Identification and Wholeheartedness*. In this later work, Frankfurt draws an analogy between the way in which we make decisive identifications with desires and the way in which a mathematician may settle on a solution to a mathematical problem:

> Suppose [the mathematician] is confident that he knows the correct answer. He then expects to get that answer each time he accurately performs a suitable calculation. In this respect, the future is transparent to him, and his decision that a certain answer is correct resounds endlessly in just this sense: It enables him to anticipate outcomes of an indefinite number of possible further calculations.[33]

Similarly to how a mathematician can become confident in the correctness of his calculations after a point and decide that it would be pointless to check them further, persons can be led to have a similar confidence in the fact that a first-order desire will meet with no challenge further up the hierarchy. Although we cannot as 'finite beings' ever in our lifetimes consider all of our nth-order desires as they may extend up the hierarchy, we can after a point 'cut off' reflection and decide confidently enough that there is no conflict waiting for us further upwards. Such a reflection-culminating decision is not arbitrary, according to Frankfurt, because at some point in a series of reflections because the only reason one might continue to reflect would be to deal with what they take to be a potential conflict; "[given] that the person does not have this reason to continue, it is hardly arbitrary for him to stop."[34]

This account is vulnerable to a line of objection raised by John Christman is that it is not difficult to imagine an individual being led to make a reflection-culminating decision which doesn't seem to have any significance at all to their character. Take, as Christman does, the example of an individual hypnotized to desire strawberries. If also hypnotized to ignore the

[33] Franktfurt, 1987, 168.
[34] Frankfurt, 1987, 169.

hypnosis in their reasoning, then we can imagine such an individual decisively identifying with these desires, yet we don't think that such a desire has any significance to his character, or that his actions emerging from it are free in the sense under discussion.[35]

While Frankfurt went on to propose further modulations of his original hierarchical account,[36] much of his later work on identification instead appeals to various notions of care, and this line of approach has attracted significant interest. In what follows I discuss three prominent accounts put forward by Harry Frankfurt, David Shoemaker, and Agenzia Jaworska in their respective works. Structurally speaking, these accounts are very similar; they all hold that that being identified with mental states is a matter of them being in some way constitutive of our cares. Accordingly, when actions arise from desires with which we are identified, then we are also identified with the action.

This shift in the literature was again precipitated by a work of Frankfurt's, specifically, his *The Importance of What We Care About.*[37] Although Frankfurt has since publishing this work written extensively discussing different conceptions of care and its species love concerning their roles in agency, I limit discussion to his earliest work on the subject here to avoid spending time delineating the accounts Frankfurt has since developed.

Care, as Frankfurt characterizes it in Frankfurt, 1982, "is constituted by a complex set of cognitive, affective, and volitional dispositions and states."[38] He claims that what a person cares about "coincides in part with the notion of something with reference to which the person guides

[35] The above is adapted from Christman's *Autonomy and Personal History*. (Christman, 1991, 8-9)
[36] See for instance his *The Faintest Passion.* (Frankfurt, 1992)
[37] Frankfurt, 1982.
[38] Frankfurt, 1982, 85.

himself in what he does with his life and in his conduct."[39] Although animals can certainly guide themselves in their conduct in various ways, "[caring], insofar as it consists in guiding oneself along a distinctive course or in a particular manner, presupposes both agency and self consciousness."[40] Additionally, on Frankfurt's conception of care "[a person who cares about something] identifies himself with what he cares about in the sense that he makes himself vulnerable to losses and susceptible to benefits depending on whether what he cares about is diminished or enhanced."[41] We might summarize that on Frankfurt's 1982 account, a care is a collection of dispositions and mental states focused around the fortunes, positive and negative, of certain entities, produced at least in part through some reflective process. On this picture of care, it seems reasonable to take an individual to be identified with those desires which constitute their cares.

David Shoemaker in his *Caring, Identification, and Agency*, similarly claims that "identification is, for the most part, a passive process, garnering its authority for self-determination from one's nexus of cares."[42] Again in line with Frankfurt's conception, Shoemaker holds what is most distinctive of caring is that "in caring for X, [one is] rendered vulnerable to gains and losses—to emotional ups and downs—corresponding to the up-and-down fortunes of X."[43] While caring entails such emotional vulnerabilities, Shoemaker holds that the connection in the other directions is just as strong; in his words, "[an] emotional reaction, in

[39] Frankfurt, 1982, 82.
[40] Frankfurt, 1982, 83.
[41] Frankfurt, 1982, 83.
[42] Shoemaker, 2003, 90.
[43] Shoemaker, 2003, 91.

other words, must be a reaction to events affecting something [we care about]; otherwise, it just cannot be an emotional reaction."[44]

In association with these "emotional vulnerabilities", Shoemaker holds that "certain dependent desires crop up, desires to act on behalf of the cared-for object, for example, to protect it or contribute positively to its maintenance or development."[45] These desires are those we identify with, and those which allow us to express our agency in action. To make this discussion slightly more concrete, Shoemaker gives a helpful passage:

> Consider first the relation between caring and emotional dispositions. Suppose I care a great deal for my dog. Such caring will dispose me to a variety of affective states. For example, when I come home from work and see his tail wagging and his mouth open in what I take to be a smile, my affection for him will come rushing to the forefront (producing dependent desires as a result: I will want to pet and play with him and give him a treat for not tearing up the house while I was away). And when I see my dog limping or listless, I will be distressed (wanting, as a result, to get him to the vet and fix whatever the problem is). Indeed, caring requires such emotional vulnerability; I would, after all, reject any attempt to block the feelings of distress attached to seeing my dog in pain.[46]

On Shoemaker's picture we might summarize that cares are best understood as relations individuals bear to various sorts of things around which emotional reactions and dependent desires emerge tied to the affairs of these things emerge. What is more, it is only due to our cares that mature adults have emotional reactions at all.

Yet more recently, Agniezka Jaworska in her *Caring and Internality* has developed an account of care motivated by fact that individuals in what she terms 'marginal cases' don't participate in reflective processes but still seem to be able to fully express their agency in their

[44] Shoemaker, 2003, 93.
[45] Shoemaker, 2003, 92.
[46] Shoemaker, 2003, 92.

actions and emotions.[47] Young children and Alzheimer's patients, for instance, still seem able to care about things and act in accordance with these cares, even though they seem not to assess their stances towards the things they care about or the emotions involved with doing this consciously.[48] In order to accommodate these cases, Jaworska takes the relevant species of caring to be, in psychological terms, a secondary emotion, or an emotion which involves "conscious, deliberate considerations you entertain about a person or situation."[49] This is in contrast to primary emotions, which include reactive responses such as fear and disgust and arise in us without any sort of conscious processing.

Additionally, Jaworksa expands upon the notion of identification used in earlier literature, claiming "[identification] is meant to track attitudes that "speak for the agent," that are fully the agent's own and internal."[50] This in a certain way synchronizes with her picture of cares as "a structured compound of various less complex emotions, emotional predispositions, and also desires, unfolding reliably over time in response to relevant circumstances."[51] Not only can we identify with desires, according to Jaworksa, but also with a wide variety of other, care-constitutive-emotions, such as "grief, regret, hope, and so forth."[52] As I have expressed, this

---

[47] Jaworksa, 2007.
[48] If Jaworska is correct in this claim, which I take it she probably is, we might note two things. First, many if not all of Frankfurt's earlier accounts which make some amount of reflective activity necessary for identification face a strong objection to contend with. Second, the fact that individuals with so little rational activity are capable of expressing themselves fully in these ways provides strong evidence that something like caring, as opposed to, say, our judgmental activity, is what allows us to identify with our actions and mental states. might be noted that these cases provide strong evidence that something along the lines of caring is at the heart of the issue of identification. I don't discuss either of these points any further in this paper.
[49] Jaworska, 555. Jaworska takes this definition from a work by Antonio Damasio. See Antonio Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*, 136.
[50] Jaworska, 537.
[51] Jaworska, 2007, 560. This last clause is meant to capture that cares involve secondary emotions.
[52] Jaworska, 562.

strikes me as correct; while the literature's initial focus was on free will and action, and desires are the mental states most pertinent to these discussions, there is no reason why desires should be the only mental states that we are capable of having as 'really our own.' Just as we can ask if some individual desires freely, we can coherently ask, for instance, if they have one or another attitude freely, or if it is merely 'passing by' or else is intruding.

3.1: Moral Attributability

Frankfurt's original goal in characterizing his technical notion of identification was to give the conditions under which we act and will freely, or when our actions and desires are part of our characters it in a way which distinguishes us as persons. As I have suggested, we ought to follow Jaworska holding that this identification has a wider scope, and that we are capable of identifying with other affective mental states[53] than just desires. Conceived in this way, an additional attractive point of an account of what an individual identifies with, though not the only one, is that we can see it as providing an account of the mental states and actions an individual is responsible for.

There are however many ways of attempting to get at what an agent is responsible for,[54] and one sort of responsibility I would now like to consider is moral attributability, first identified by T. M. Scanlon in his "*What We Owe to Each Other*".[55] To say that someone is responsible for something in this sense is to say they are a viable target for a certain variety of *aretaic* appraisal because of it; alternatively, as Scanlon himself puts it, it "is only to say that it is appropriate to

---

[53] Both Jaworska and I do not make any serious attempt to delineate exactly which species of mental states we can identify with. It is part of my case in this paper that we can identify with desires, attitudes, and patterns of awareness. What exactly a 'pattern of awareness' is will become clearer as the reader continues.

[54] See, for instance, Shoemaker's *Attributability, Answerability and Accountability: Toward a Wider Theory of Moral Responsibility*. (Shoemaker, 2011)

[55] Scanlon, 2000.

take it as a basis of moral appraisal of that person."[56] Following Scanlon, other writers have taken an interest in the role of this sort of responsibility. Angela Smith, for instance in her *Responsibility for Attitudes: Activity and Passivity in Mental Life* discusses attributability as

> ... the conditions under which something can be attributed to a person in the way that is required in order for it to be a basis for moral appraisal of that person. To say that a person is responsible for something, in this sense, is only to say that she is open to moral appraisal on account of it (where nothing is implied about what that appraisal, if any, should be).[57]

I take it that Smith and Scanlon's conceptions of this notion are more or less the same. What moral attributability concerns itself with is not whether some mental state or act 'belongs to someone' as an agent in something like Frankfurt's sense, but instead whether under everyday moral norms appraisals could be made of these individuals on the basis of these things, that is, whether they would on the basis of these things be a proper target of such an appraisal, putting aside what such an appraisal might be, or any practical consequences expressing it to them may have.[58]

For the sake of contrasting these two sorts of responsibility, those implied by identification and attributability, consider the case of an individual with deep-seated racist tendencies which we might suppose they are not even aware of. This individual might, for instance, desire not to hire individuals of a race other than their own, and this desire might manifest itself in their behavior. They might at the same time insist earnestly if asked that they are indifferent to the race of the individual they intend to hire. Certainly we feel such prejudiced tendencies do render an individual an appropriate target for moral assessment. But it is at best

---

[56] I take Scanlon to be using 'person' in a loose sense. (Scanlon, 2005, 248)
[57] Smith, 2005, 238.
[58] In this paper I do not differentiate between attributability and moral attributability and use the corresponding expressions interchangeably.

unclear, on most of the accounts of identification we have considered,[59] whether this individual is identified with such a desire.

Here is a more decisive case: consider an individual with the same desire against hiring an individual of a different race, but one who is aware of this desire, alienated from it, and actively attempts to avoid allowing it from influencing their actions. The higher-order desires they have against having the prejudiced desire aside, such an individual is at any rate made a proper target for moral assessment by having the prejudiced desire (their second order desire, we might even plausibly assume, was born of *their own* moral assessment *of themself*). In this case, not only is the prejudiced individual not identified with the prejudiced desire, but they in fact seem *alienated* from it. This desire, however, is no less attributable to them despite this.

The two cases above both draw evidence that the conditions of identification are different from those of moral attributability from the fact that we can and do in practice hold others responsible in a very commonplace way for states and actions with which they are not identified. I take it that they demonstrate this definitively. I doubt this point is novel, and I do not take myself to be exposing any substantial problem, *per se*, for accounts of freedom which rely on the notion of identification as the ones I have so far discussed do. I think we have an intuition, though, that there is at least some sort of connection between those states we have as agents and those actions and states which are morally attributable to us. Here is an argument to help make this feeling more tangible: only of persons can the sort of *aretaic* appraisal involved in morally attributing something to someone, one of them as a moral agent, be made. This suggests that the *preconditions* of someone being an agent relative to their actions and mental states, should

[59] The exception being Frankfurt's account as elaborated in *Identification and Wholeheartedness*. Given that the individual we are discussing is unaware of this desire, it seems fair enough to conclude that they could not have decisively committed themself to it.

explain, if at least only partially, when they *can* have things attributable to themself in this way. Put in other words, given accounts of free will and responsibility as attributability, it would be a significant theoretical boon for each (and both, taken together) if it were the case that our *capacity* for free action would be able to somehow explain our *capacity* for attributability.

With this said, the goal of this paper is to provide an account of attributability in terms which are amenable to theories of free will, free action, and the free possession of mental states which make appeal to a concept of care, for which we can explain our intuition that at least our capacity for agency can explain our capacity for attributability. What I argue in the latter half of this paper is that it is our desires constitutive of our cares (for a certain conception of care) which make us responsible for our mental states or our lacks thereof. More precisely, I argue that having a mental state or lack thereof attributable to oneself is a matter of having a care-constitutive desire for or against taking it.[60] While it would be desirable to also provide an account of attributability for actions, their case is ostensibly more complex and space is limited so I do not discuss the topic further here. In order to facilitate discussion of my own account and to make some of the discussion contained herein about attributability more concrete, before presenting my view I discuss and provide criticism of a recent competing account of moral

[60] One might wonder: why not take a different approach, and say a mental state is attributable to someone when they have care-constitutive desires *toward taking* it, and the dearth of such a state is attributable only when there are care-constitutive desires *towards lacking* this state? This approach gives the wrong result in cases where one might have objectionable first-order attitudes without having care-constitutive desires towards having these attitudes. In the case someone has, say, a racist attitude which they don't have care-constitutive desires directed towards taking, we still feel like we could make an aretaic appraisal of them for taking the attitude. If we add to the first of the examples discussed above, for instance, that the individual has no desire to preserve their deep-seated racist tendencies, this does nothing to the fact that these desires still seem attributable to them; they might even make aretaic appraisals of themselves, experiencing it as a failure on their part that they have these desires.

attributability for mental states formulated by Angela Smith which locates the attributability of our mental states to us in our evaluative judgments.

3.2: Attributability: Smith's Rational Relations View

In her paper *Responsibility for Attitudes: Activity and Passivity in Mental Life*, Angela Smith makes two interesting claims: (1) that we are often responsible for what she calls the 'patterns of awareness' we display and fail to display and the attitudes we both take and fail to take, and (2) that it is the rational connections these things bear to our evaluative judgments in such cases which makes these things attributable to us.[61]

Smith draws upon compelling cases to support her claim that these sorts of mental states or the lacks thereof are often attributable to us, and I agree with her on this matter. What Smith calls our 'patterns of awareness' are, very broadly, "what we notice and neglect, what does and does not occur to us, and what we see as relevant in our practical deliberations what either occurs or fails to occur to us."[62] Her leading example is that of an individual who forgets a friend's birthday.[63] In such a case, it seems clear that one becomes open to a certain variety of criticism. To give a few more of Smith's example cases: "If I do not notice when my music is too loud, when my advice is unwelcome, or when my assistance might be helpful to others, these again can be described as involuntary failings; nevertheless these failings are commonly taken to be an appropriate basis for moral criticism."[64] More commonplace than her claim about patterns of awareness are those involving the attitudes we take in our everyday lives. Although Smith discusses these rather less, it is not hard to produce examples: when one is envious of the

[61] Smith, 2005.
[62] Smith, 270.
[63] Smith, 235.
[64] Smith, 244.

successes of a friend, takes racist attitudes, or is subject to certain vicious desires, for instance, we feel as though that someone takes one of these attitudes is often equally legitimate grounds to confront them as responsible for behavior.

On the rational relations view which Smith argues for, the attributability of the attitudes and patterns of awareness we display is accounted for by two separate sorts of normative, rational connections that these mental states have to our judgments. The attributability of patterns of awareness to us is accounted for by rational connections they bear to the judgments we make. The relevant norms in the case of patterns of awareness are I take it essentially those of practical reason. If I judge someone a good friend, for instance, I should, rationally, think about them enough to not forget to call them on their birthday. Separately, in order to account for the attributability of our attitudes, Smith relies on the judgmentalist theory which holds that many of our intentional mental states have judgments partially constitutive of them. Given such a theory, the judgments partially constitutive of our attitudes will have the potential to conflict with those we make more explicitly, and in such cases requests for justification can be made on the basis of these conflicts. Smith thus models the attributability of attitudes on what she takes to be rational conflicts such as the following:

> If I sincerely judge that there is nothing dangerous or threatening about spiders, I should not be fearful of them. The emotion of fear is conceptually linked to the judgment that the thing feared is in some way dangerous or threatening; therefore, my judgment that spiders are not in any way dangerous or threatening rationally entails that I should not be fearful of them.[65]

What I take to be the largest problem with Smith's proposal is that it relies on the controversial judgmentalist theory that judgments are constitutive of many of our attitudes. Consider the case of an arachnophobe who is fully aware that their fear of common house spiders

[65] Smith, 253.

is entirely unwarranted. This individual might recoil from a house spider while at the same time acknowledging it to be utterly harmless. But in such a case, it is clear what this individual judges of the spider: it is completely harmless. If this is the case, then analyzing this and similar cases in terms of conflicting judgments simply will not work.

To get a better grasp of the problem this presents for analyzing moral attributability, we can again imagine the case of someone with racist tendencies, who can't help but see individuals of races other than their own as, say, untrustworthy. They might be ready to earnestly assert that their attitude has no basis in the character traits of these individuals and detest themselves for being host to them, but nevertheless possess the racist attitude. While this attitude is morally attributable to them, this circumstance simply does not seem properly frameable as one in which judgments conflict, but something else altogether.

Smith admits that her account faces problems posed by "recalcitrant attitudes", and it is conflicts such as the above, I take it, to which she is referring.[66] What exactly constitutes judgments is a difficult question with much literature surrounding it, and I do not explore it in any depth in this paper. While I think we plausibly do more often than not have associated judgments alongside our attitudes, the above cases give strong evidence that this is not always true. It is unclear to me how Smith should go about changing her account to remedy this problem. The fact that judgments simply don't seem to be what is at the heart of these cases suggests that this problem runs deep in her approach.

The problem more pertinent to the earlier discussion is, I take it, somewhat smaller. This is that the literature on free will/action/mental state possession gives strong indication that it is not our ability to make judgments, *per se*, which makes our acts and mental states belong to us as

[66] Smith, 271.

agents, but instead our cares. It may be the case that there is a strong argument, perhaps proceeding from claims involving conceptual connections, that there is some link between caring about things, which makes us persons with respect to our mental states and actions, and the judgments which may make us responsible for our mental states. Or, it might turn out that Watson was right all along, and it is judgment sensitivity which accounts for our agency with respect to these things. But I see no strong reason to think that either of these states of affairs is the case, and the literature on the issue to the contrary gives us good reason to think that it is our cares, on some conception, though some dynamic, which allow for us to be agents relative to our actions and mental states. In short, it is at best unclear on Smith's theory how we can explain our capacity for attributability in terms of our capacity for agency.

I do not think this second criticism is in any way definitive by itself. But when we weigh accounts of phenomena against one another, unintuitive points in this vein do add up and make the case for one view or the other stronger or weaker. In any case, if an account of moral attributability were able to provide a clear account of this connection, we should think this an advantage of it. This motivates an account of moral attributability as care sensitivity.

3.3: Further Motivating a Care-based Approach to Attributability: The Prototype Account

While my goal, speaking broadly, is to argue that our cares which make things attributable to us, considering first a similar, simpler but incorrect Humean view also helps motivate appealing to care. Constructing and criticizing such a view is the goal of this section.

Attributability is at least plausibly a precondition of accountability. Being held to account over something, or being 'held responsible' for it, seems to involve, among other things, an aretaic appraisal, and being the proper subject of such an appraisal over something is just to have

it attributable to oneself. As a commonplace, being held to account involves a certain sort of sting—it is unpleasant. As another commonplace, when our desires are frustrated, proportional to the strength of the desire we experience a sort of affective penalty. These things together suggest, even if only very obliquely, that when our attitudes or patterns of awareness are attributable to us, it is at least often because we have a desire which our own conduct frustrates that can be appealed to.

That we do take desires both for and against the mental states that we both have and could have strikes me as another commonplace. While Smith gives many examples to demonstrate that the attributability of mental states to us is closely tied to our judgments, it is not difficult to repurpose her cases in a certain way to demonstrate the relevance of our higher-order desires[67] to these practices, often times more compellingly. I give discussion to four cases of the attributability of attitudes and patterns of awareness or lacks of these things, suggesting that in each case we can explain the attributability of some mental state or lack thereof to us through the presence of higher-order desires.

Consider the case of two good friends who are employees at a company, Charles and Eric. Whether through good fortune or hard work, suppose Charles gets a promotion. If they are good enough friends, it is reasonable to expect that Eric will want to be happy for Charles in light of his good fortune, to in a certain way affectively share in his success. Were he to fail to be happy for him, we expect he would be receptive to criticism on the basis of this, and we can explain this receptivity through a higher-order desire to feel good for Charles.

[67] From herein I use 'higher-order desires' to refer to desires directed at, not just other desires, but other mental states more generally.

Keeping with the same characters, Eric would also plausibly desire not to take certain attitudes towards Charles on the basis of his success. While he would rather be happy for Eric in the face of his successes, he would simultaneous be all the more troubled were he in such a circumstance envious of him. This is just to say we expect that Eric would additionally not want to be envious of Charles. Again, we expect he would be receptive of criticism on the basis of having these attitudes and can explain this through the presence of desires against taking them.

We move on to the cases of patterns of awareness. To straightforwardly adapt one of Smith's cases to this purpose, suppose Jenny has a good friend, Bill, whose birthday she's forgotten.[68] Upon realizing this, she would likely feel a certain degree of guilt for forgetting the date, and experience it as a failure on her part that she didn't give Bill what she takes to be due consideration. This experience of deficiency is explainable, at least in part, by the frustration of a desire to at least be aware about considerations involving Bill in her reasoning.

Again using one of Smith's cases regarding objectionable attitudes, were it to occur to Jerry that he might solve some problem in his life through committing or orchestrating some act of violence, it would likely disturb him that such a thing would even cross his mind.[69] Even if some experience of this sort of guilt is not immediate, upon reflection or being confronted over the fact that he would consider such an action, such sentiments could likely be brought out. The root of these sentiments, it seems most plausible, are in the fact that he to some degree does not want such options to cross through his mind.

In each of these cases some individual experiences/is vulnerable to experiencing what they take to be a failure on their part in having or failing to have certain mental state(s), and in

[68] Smith, 236.
[69] Smith, 246.

each case we seem to be able to explain this experience of failure by the frustration of their higher-order desires. Additionally, in each these cases the protagonist is intuitively responsible for lacking or having the relevant mental states. This lends more plausibility to the idea that it is our higher order desires for and against taking attitudes and patterns of awareness that make these mental states (or lacks of mental states) attributable to us. According to what I will call the prototype account, then, when we have a desire for or against having some attitude or pattern of awareness, the mental state or lack of it, respectively, thereby becomes attributable to us.

At least two weaknesses present themselves for this view. The more substantial problem is that it seems conceivable that one could cultivate within oneself a higher order desire that has nothing to do with being responsible for anything. Suppose someone in an idle moment sees a crumb in front of them. Unremarkably enough, they might come to have some weak desire to flick it away. It also seems at any rate conceivable that they might, if they were so inclined, be able to convince themselves to not only desire to flick away this crumb, but also to desire to desire to do so. This certainly seems like an odd case, but I see no compelling reason to deny that it is possible. If the prototype account is correct, then such an individual would have in their everyday musing made themselves responsible for desiring to flick away the crumb. This of course seems implausible.

The second worry is that even supposing the prototype account perfectly picks out every instance in which one has some relevant sort of mental state attributable to them, it seems as though it might fail to provide a complete enough explanation of why these mental states are attributable to them. Suppose, then, that Fred notices Bill is angry at their mutual friend Eliza in a way which strikes Fred as unfair to Eliza. Perhaps Eliza missed a gathering due to being called in to work, and this has angered Bill. Motivated by this perceived injustice to Eliza, Fred might

chastise Bill, "you really shouldn't be so angry with Eliza." Plausibly, Fred is trying to hold Bill to account, or we might say ‘confront him as responsible’, in such a case. Suppose Bill admits, trying to calm himself, you're right; I shouldn't." While it is plausible that this holding-to-account might have been successful, this case is actually underdescribed. If the only reason Bill seems receptive of this criticism is that his anger is getting in the way of him enjoying himself or else doing his work, then he would plausibly have missed the message communicated to him by Fred entirely. His attitudes toward Eliza are attributable to him according to the prototype view, as he has a desire to be rid of them, but at the same time something has failed in this instance of account. Plausibly, a more complete account of why similar states are attributable to us will be able to explain at least more fully what happens in such a case.

3.4: Attributability as Sensitivity to Norms of Care-Constitutive Desires

The largest difficulty with the prototype account is that it maintains we have more mental states attributable to us than actually are. If we wanted to amend instead of throwing out the prototype account, the approach this suggests is finding a narrower collection of second-order desires which make our mental states attributable to us. I propose that it is our care-constitutive desires, for a certain account of care, which suited to fill this role. It may sound *prima facie* implausible that to be responsible for a mental state we must have a care 'in the area' with an appropriate desire constitutive of it. The conception of care that I claim can do this philosophical work I do not propose as an account of care *simpliciter*, and I do not claim conforms to our every use of the term. Judged as an account of care *simpliciter*, I expect it would fail. It promises to do good philosophical work in explaining attributability, and this is what I take to justify exploring it.

Consider, then, some comments from Frankfurt's presentation *Taking Ourselves Seriously*:

> When we do care about something, we go beyond wanting it. We want to go on wanting it, at least until the goal has been reached. Thus, we feel it as a lapse on our part if we neglect the desire, and we are disposed to take steps to refresh the desire if it should tend to fade. The caring entails, in other words a commitment to the desire.[70]

Frankfurt suggests here that caring about something involves desiring it in a certain way. I take caring about something to involve a more complex collection of higher order desires than Frankfurt might be taken to suggest here. Intuitively, when we care about something, not only do we take desires toward other desires, but also toward an array of other mental states, including attitudes and patterns of awareness. When we come to care about someone as we do when we become good friends with them, to look back to multiple of the prior examples, not only do we want to have desires to do certain things with or for them, but we desire not to be envious of them, to be happy for them in their successes, and to involve their affairs in our reasoning more than we would otherwise.

What is most important for the conception of care that I propose is that caring involves taking a complex collection of desirative attitudes of a certain nature towards and against having certain mental states. Besides these higher-order desires, the passage I initially took from Frankfurt suggests that caring about something in his sense involves taking certain first order desires centered around what we care about as well, and the other accounts of care I reviewed earlier in this paper also suggest this. It is certainly the case that when we care about something, we very often do have the lower-order attitudes/patterns of awareness, including first-order desires, we take as the objects of these higher-order desires. Regardless, I do not treat them as strictly necessary to our caring about something.

[70] Frankfurt, 2004, 18-9.

Caring as these three authors characterize it also involves what we might describe as an emotional investment in the affairs of the object cared about. Analogously to the discussion in the above paragraph, while vulnerability to such sentiments often accompanies caring about something, for caring as I characterize it the only emotional involvements required are higher-order ones, for and against taking certain mental states.

In principle, anything can be cared about. We can care about other people, fictional characters, our relationships with other people, social justice, Nazism, that we drive safely, and that the walls in some room be painted green. If Jerry cares about social justice, for instance, he will plausibly desire to desire to see it implemented, desire to desire to help realize it, desire to do away with any latent discriminatory tendencies he or others may have, and desire to consider it when concerns related to it may be pertinent. Such a list could be extended.

The reader might demand a more detailed description of the complex of higher-order desires involved in caring about something than has been given. That the objects of cares are so diverse precludes giving too detailed of an account of precisely the variety of higher-order desires involved in caring about something. While the authors herein discussed take emotional investment in what happens to what is cared about as closely tied to caring about something, what I instead take as characteristic of care is emotional investment in respectively *manifesting and not manifesting the lower-order mental states, which gives rise to desires for and against taking them*. What is necessary to caring as I conceive it are not first-order attitudes, but those 'one degree higher.' Caring involves taking desires (and corresponding emotional investments) for and against taking lower-order mental states. While there may be some sort of expectability relation which holds between caring about something and taking the relevant first-order mental states, these lower-order states are not what is characteristic of care itself. I take it that this aspect

of care makes the specialized use I am giving to the word differ rather markedly from that which we commonly give it, but I again have no intent of capturing the notion *simpliciter*.[71]

The account of care I am presenting here I take to be essentially a modification of that Shoemaker proposed in *Caring, Identification, and Agency*. In this work Shoemaker again takes cares to be 'emotional tethers' to entities due to which all of our emotional reactions emerge, and in accordance with which desires are produced. The two modifications I am making are mandating that (1) care involves emotional investment in and the taking of higher-order desires towards a wide array of affective mental states and (2) all that is necessary for care is the presence of the emotional investments and higher-order desires mentioned in (1). I take past examples to provide strong evidence that we do take higher-order desires towards and against taking the sorts of mental states here under discussion, so the emotional investments and higher order desires mentioned in (1) I take to be required by care on Shoemaker's account, though he does not explicitly discuss them. The purpose of this second modification will be made clear in my discussion of the second potential objection I consider in the following section.

As a concise statement one may hold me to, the view which I am here defending holds that to have some mental state (or lack of such a state) attributable to oneself is to have one's higher-order care-constitutive desires weigh in for or against manifesting the corresponding mental state. Call this view the care account.

The care account skirts both of the issues initially raised for the prototype account relatively straightforwardly. Firstly, even if one is capable of coming to have 'stray' higher-order

[71] The comments in this paragraph and the few above it are liable to look curious to the reader. My reasoning in singling out this conception will be clear by the time I discuss the second objection in the coming section, if not before then.

desires, as one might come to have towards flicking a crumb, this will not be enough to make one responsible for having the first order attitudes/patterns of awareness. Unless the higher order attitudes we take emerge from our nexus of cares, they have no significance to the attributability of any mental state. Related to this topic, Smith claims that "when certain slogans or song lyrics keep running through one's consciousness", we are not responsible for such occurrences because they do not depend on our judgmental activity.[72] The explanation I offer instead is that the lack of attributability for our stray desires to flick crumbs or The Spice Girls being stuck in our head arises from the fact that on the level of what we care about we are indifferent to these occurrences. When our cares fail to weigh down for or against some mental state we are with respect to these states little different from most animals; we might have desires and attitudes, but we will lack any substantial conscience which makes us them attributable to us.

The second issue with the prototype account concerned its ability to provide a complete enough explanation of what occurs when one individual confronts another as responsible. Returning to the case of Fred chastising Bill for his anger at Eliza, the care account can explain what went wrong. Fred most plausibly confronted Bill with an appeal to those attitudes constituent of his caring about Eliza, rather than any belonging to a care he had for his own well-being or enjoyment. Bill, however, misinterpreted this appeal as one of the latter sort. This suggests a partial account of what goes on in holding others to account: when we hold another to account over their mental states, we do so on the basis of some *certain* care they have, not merely the fact that they have higher order desires in general, without regard as to where these desires emerge from.

[72] Smith, 260.

3.5: Objections

With the above, I have laid out the essential points of my account of attributability. Before returning to some of my earlier claims concerning attributability and agency, I discuss two objections to the account of responsibility-as-attributability I have provided in the care account.

Objection 1: As the existence conversationalist theories of responsibility evidence, when we hold others responsible there is often the potential for a rich dialogue between the those holding and held to account. If having something attributable to oneself is simply a matter of caring about something, we should expect we would often not have the ability to engage in this sort of exchange. Thus we ought be skeptical of the care account.

Response: Although it may not be obvious, our cares provide us with a rich basis for both moral conversation. This is accommodated by the two facts. The first is that we are often opaque to ourselves with regard to what we care about. The second is that we seem to care about many things only instrumentally.

Suppose Todd accepts that he cares about everyone being treated fairly, but at the same time professes to be indifferent to, say, issues of social justice. If confronted over his indifference to some issue of social justice, he may thus not be immediately receptive of criticism. But we might attempt to convince him that he should acknowledge his fault in such a case by either attempting to convince him of something like a conceptual connection between social justice and fairness, or else that what attitudes he takes in some instance aren't appropriate for one who cares about fairness. There is plenty of room for moral discussion through such exchanges. What is more, participating in such exchanges seems to me what we naturally tend to do when we meet with difficulties in confronting someone as responsible.

Objection 2: On the care account, we won't be able to hold many individuals responsible for their attitudes who we believe we can. A devout racist, for instance, clearly does not at all care about social justice or racial equality, yet we clearly can hold such an individual responsible for their objectionable attitudes in the way which is relevant to attributability.

Response: This objection is liable to arise due to a particularly large tension between our use of the word 'care' in everyday language and my specialized use. The accounts care we have discussed in Frankfurt's, Shoemaker's, and Jaworska's works are undoubtedly closer to our common usage in that they require us to be directly emotionally invested (in one way or another) in the fate of what we care about. When we say someone cares about something in everyday speech, what seems most centrally supposed is that this person has first-order desires of a certain strength towards promoting whatever they see to be the good of what they care about. Consistently with this, we often say that animals can care for their young or other members of their social grouping.

In caring in the sense I have developed, strong ‘first-order’ desires or other attitudes need not be present. Instead, a care might have little constitutive of it but relatively weak latent second order desires which hardly manifest themselves at all in our lower-order mental states or actions. This is so because it simply is the case that certain mental states or deficits thereof are often attributable to people even in the absence of these strong first-order attitudes. While we can imagine some petulant racist who takes highly objectionable attitudes without being at all reflective on them, my impression is such individuals are rarely beyond caring. It is difficult to support such generalizations, e.g., that most racists care about racial equality, but similar generalizations have had a certain amount of support in the past. The most notable example will

be Hume's claim that "[w]hatever conduct gains my approbation, by touching my humanity procures also the applause of all mankind, by affecting the same principle in them."[73]

Besides historical support, there is an analogous case which makes this claim seem highly plausible. Suppose you are eating a hamburger, and in doing so attract the attention of a devout, exceptionally pushy vegetarian. She might rail against you, claiming that your ability to order and eat food the production of which involves the inhumane treatment and slaughter of other sentient creatures without a second thought is disturbing. Plausibly contained in such a tongue-lashing is an appeal to your cares. If I had to wager, the reader has had at multiple points in their life feelings very similar to those the devout vegetarian is attempting to inspire, and have felt the associated pangs of guilt, yet, statistically speaking, the reader is no vegetarian. Essentially all of us care about other sentient life and are unsettled by such a rebuke. At the same time, most of us manage to get by without experiencing such feelings when we order hamburgers.

The above is a powerful testament to the fact that human beings have a rather disturbing capacity to ignore their second-order desires, their conscience, when it is expedient for them. In light of this and similar cases, it seems very plausible to me that virtually all racists and individuals in like cases are to some degree susceptible to being appealed to on the basis of their cares, and this is all that the theory I propose requires. In the event that in some case someone is completely indifferent, on the level of their cares, to the objectionable mental states they as a matter of fact have, whatever they are, then they are beyond being responsible for them. But such individuals are so exceptional that we ought to think that this is the correct conclusion in such cases; with respect to these mental states, they are not unlike most animals.

[73] Hume, *"An Enquiry Concerning the Principles of Morals"*, 76.

4.1: Autonomy, Attributability, and Account

With these objections addressed, I now return to my original claim, that this account of attributability in terms of our cares will plausibly be able to give an account of the connection between our abilities to take actions, will, and have mental states as agents and have mental states attributable to us. If we consider care as I have suggested, then identifying with mental states might be most intuitively construed as a matter of having desires constitutive of care weigh in toward having a mental state. But this will not be enough. Care as I have construed it, in order to avoid the second objection posed above, need not involve actually having the mental states which are the objects of our care-constitutive higher-order desires. One cannot have a mental state freely if one does not, in the first place, have the relevant mental state. The simple solution which this suggests and which I take is that we identify with mental states as agents just when we (1) have the mental state and (2) have care-constitutive desire(s) toward having the relevant mental state.

Strictly speaking if this proposal is correct then it will be that an individual can have states attributable to them without identifying with any mental states at all, although such an occurrence, we should expect, would be a rarity. The most easily imaginable case would be an individual that only cares about one thing but doesn't have any of the first-order mental states they desire to take. Although one may not have the first-order mental states which their care-constitutive desires demand they do, and thus may not be identified with any of their mental states, the ability to have mental states attributed to oneself is explained by those states which give us the capacity to identify with these states.

The goal of this paper is to argue such an account of attributability which meshes with the literature on freedom in this way is plausible. As I prefaced this project, I do not intend to argue

that this account of identification is capable of doing the work of characterizing when we act and take mental states in the agent-specific way which the literature in this vein attempts to. Still, one reason that we should think this account of identification or perhaps one very similar should be able to fill this role is that I take it to be essentially an extension of Shoemaker's account of identification. The principle differences are that it (1) follows Jaworska in taking identification to be a wider-ranging phenomenon which can hold between persons and different types of mental states and (2) holds that second order desires and associated emotional investments are required for identification with these mental states. These qualifications do not strike me as at all objectionable. In fact, I see no reason why anyone should take identification to be a phenomenon restricted only to desires and actions and think that Shoemaker's own view probably entails that second order desires of this sort are required for identification, given the nature of the emotional connections that caring on his view requires. Shoemaker however doesn't speak on either of these points, so I am cautious about attributing this view fully to him.

I take it this parity between my account of attributability and having these mental states as agents as highly desirable. In closing I would like to remark on two interesting results which fall out of it. This view suggests a picture of the process by which we hold others responsible for mental states, or account, as a distinctively interpersonal manner of communication. When we hold others to account, we engage with them as a being with the capacity for conscience and agency—we make appeals to their care-constitutive desires, appeals to them as an agent to take the mental states they might identify with to be of a character[74] more consistent with their cares. When we fail to do this, or when we merely attempt to make deals with others or threaten to harm or sanction them, we fail to engage with them in this distinctively interpersonal way.

[74] I use the word in a loose, non-technical sense here.

This is not to say that when hold others to account we are in every case doing so with the express intent of having them realize their cares. Nor is it to say that when we do so threats or sanctions are never present. They generally are; when we are moved to take reactive attitudes, this is generally[75] because something or someone we are emotionally invested in has been harmed or stands to be harmed. Very often, our desires to protect or avenge what we care about are simply strong enough that we are willing to bargain with or impose various costs on others to do so, whether we can justify this by some practice-external stance or not. But if all there is in some instance is an attempt to bargain with a person or sanction or threaten them, then one fails to hold such an individual to account, to interact with them as a person, but instead only 'takes account' of them as an obstacle to their goals, as one might most animals.

This is by no means a complete picture of account and accountability, but it yields another compelling result. While there is often some threat of sanctions contained in instances in which we hold others to account, this is not true in all cases, and the care view accommodates this. Consider what Lawrence Stern says when he claims that:

> [King and Gandhi] publicly rebuked others for doing wrong. We have every reason to believe that the rebuke was the sincere expression of a sentiment. Therefore, in some important sense they blamed others for wrongdoing. Yet it is not clear that they wished those they blamed to suffer. Nor did they exclude them from the moral community. For their method of action was to appeal to the conscience of their adversaries.[76]

At least as regards our mental states, appealing to someone's care-constitutive desires is not so different from appealing to their conscience. While Stern uses the examples of these men to point towards a distinction between moral approbation and moral indignation, they also

[75] And we might note if Shoemaker is correct in claiming that adult emotional reactions are in all cases care-constitutive, plausibly *always*.
[76] See Lawrence Stern's *Freedom, Blame, and Moral Community*. (Stern, 1974)

provide strong evidence that holding someone to account need not involve much more than an appeal to care.

References

Chisholm, Roderick M. 1966. *Freedom and Action*. Printed in *Keith Lehrer's Freedom and Determinism*. Random House, 1966. 11-44.

Christman, J. 1991. Autonomy and Personal History. *Canadian Journal of Philosophy* 21: 1-24.

Damasio, A. 1994. Descartes' Error: Emotion, Reason, and the Human Brain. New York: Avon Books.

Hume, D. 1983. *An Enquiry Concerning the Principles of Morals*. Edited by J. B. Schneewind. Hackett: Hackett Publishing Company. (Original work published 1777)

Frankfurt, H. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy* 68: 5–20.

—. 1976. *Identification and Externality*. Reprinted in his *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press, 1988. 58-68.

—. 1982. *The Importance of What We Care About*. Reprinted in his *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press, 1988. 80-94.

—. 1987. *Identification and Wholeheartedness*. Reprinted in his *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press, 1988. 159-177.

—. 1982. *The Importance of What We Care About*. Reprinted in his *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press, 1988. 80-94.

—. 1992. *The Faintest Passion*. Reprinted in his *Necessity, Volition, and Love*. Cambridge: Cambridge University Press, 1999. 95-107.

—. 2004. *Taking Ourselves Seriously*. Printed in his *Taking Ourselves Seriously & Getting It Right*. Stanford: Stanford University Press, 2006. 1-26.

Jaworska, A. 2007. Caring and Internality. *Philosophy and Phenomenological Research*, vol. 74, no. 3. 529–568.

Scanlon, T. 2000. *What We Owe to Each Other*. Harvard: Belknap Press.

Shoemaker, D. 2003. Caring, Identification, and Agency. *Ethics* 114: 88-118.

—. 2011. Attributability, Answerability and Accountability: Toward a Wider Theory of Moral Responsibility. *Ethics* 121: 602-632.

Smith, A. 2005. Responsibility for Attitudes: Activity and Passivity in Mental Life. *Ethics* 115: 236-71.

Stern, L. 1974. Freedom, Blame, and Moral Community. *The Journal of Philosophy* 71: 72-84.

Thalberg, I. 1978. Hierarchical Analyses of Unfree Action. *Canadian Journal of Philosophy* 8: 211-26.

Velleman, D. 1992. What Happens When Someone Acts?. *Mind* 101: 461-81.

Watson, G. 1975. *Free Agency*. Reprinted in *Free Will*. Oxford: Oxford University Press, 2003. 335-351.

—. 1987. Free Action and Free Will. *Mind* 96: 145-172.