teorema
Vol. XLII/3, 2023, pp. 107-129
ISNN 0210-1602
[BIBLID 0210-1602 (2023) 42:3; pp. 107-129

Self-Manipulation and Moral Responsibility

Benjamin Matheson

ABSTRACT

In this paper, I first argue that sometimes freely and knowingly manipulating oneself does not fully preserve moral responsibility — namely, in cases of practically distinct self-manipulation. However, I argue that practically distinct self-manipulation preserves moral responsibility to some extent because such a self-manipulated person is *more* morally responsibility than an other-manipulated person. This is an important result: manipulating oneself doesn't always fully preserve one's moral responsibility for one's actions. But in what sense is the self-manipulated person more morally responsible? I argue the self-manipulated person is not a fitting target of the reactive attitudes but continues to have wrongdoing-incurred reparative obligations. This explains the intuitive judgement about the self-manipulated person, provides a better explanation of "tracing" cases, and reveals important requirements for a plausible theory of moral responsibility.

KEY WORDS: *Self-Manipulation, Moral Responsibility, Temporally Extended action, Tracing, Reparative Duties, Reactive Attitudes*.

RESUMEN

En este artículo, argumento en primer lugar que, algunas veces, la manipulación de uno mismo no preserva completamente la responsabilidad moral, a saber: en casos de auto-manipulación prácticamente distinta. Sin embargo, argumento que la auto-manipulación prácticamente distinta preserva hasta cierto punto la responsabilidad moral puesto que una persona que se auto-manipula es *más* moralmente responsable que aquella que es manipulada por otros. Este es un resultado importante: la manipulación de uno mismo no siempre preserva la propia responsabilidad por las acciones de uno mismo. ¿Pero en qué sentido es más moralmente responsable la persona que se auto-manipula? Argumento que la persona que se auto-manipula no es un objetivo adecuado de las actitudes reactivas, sino que continúa teniendo obligaciones reparadoras por haber incurrido en malas acciones. Esto explica el juicio intuitivo sobre la persona que auto-manipula, proporciona una mejor explicación de los casos de "rastreo" y revela importantes requisitos de una teoría plausible de la responsabilidad moral.

PALABRAS CLAVE: *auto-manipulación, responsabilidad moral, acción temporalmente extendida, rastreo, obligaciones de reparación, actitudes reactivas*.

What happens when a person manipulates herself? Suppose Jill takes a pill that she knows will cause her to enter a zombie-like state during which she will flail her arms around about a minute after ingesting it [King (2014), pp. 470-471]. Jill then enters a china shop and once the pill takes effect, she destroys all the precious china. Because she took the pill moments before, she lacked direct control over destroying the china: there was nothing she could at that time to stop herself from destroying it. Nevertheless, it seems that Jill is morally responsible for destroying the china.

One way of explaining this judgement involves appealing to *tracing*: because Jill freely and knowingly set herself up at t_1 (by taking the pill) such that she would lack control at t_2 (when the pill took effect), she is morally responsible for destroying the china. In effect, her moral responsibility is *traced back* to an earlier morally responsible action whose performance she could reasonably foresee would lead to her breaking the china, or at least to her causing damage of some kind. As such, this explanation says that she is indirectly or derivatively morally responsible for breaking the china [e.g., Timpe (2011); Fischer and Tognazzini (2009), (2011); Miller (2017)].

Another way of explaining this judgement involves appealing to temporally extended actions. Matt King (2014) holds that Jill's taking of the pill and Jill's destroying the china are part of the same temporally extended action. To see the idea, suppose a commando plants a bomb on a bridge with the intention destroying it. If the bomb is planted at t_1 but doesn't explode until t_2, on King's view the commando's action doesn't end until t_2. The idea is that actions can have parts, and just as the commando's action of *bombing the bridge* has a beginning and an end, so does Jill's action of *destroying the china*. It begun when she took the pill that would lead her to destroying it and ended when she actually destroyed it.

Whether one understands responsibility for self-manipulation in terms of tracing or temporally extended action, there is implicit agreement that freely and knowingly manipulating oneself necessarily fully *preserves* moral responsibility. A person cannot escape moral responsibility for an action she performs by deliberately making it such that she lacks direct control at the time of action. Call this the standard account of self-manipulation.

In this paper, I first argue that sometimes freely and knowingly manipulating oneself does not fully preserve moral responsibility — namely, in cases of *practically distinct* self-manipulation. However, practically distinct self-manipulation preserves moral responsibility to some extent be-

cause such a self-manipulated person is *more* morally responsibility than an other-manipulated person. This is an important result: manipulating oneself doesn't always fully preserve one's moral responsibility. This prompts a more important question: in what sense is the self-manipulated person more morally responsible? I then argue the self-manipulated person is not a fitting target of the reactive attitudes but continues to have wrongdoing-incurred reparative obligations. This explains the intuitive judgement that the self-manipulated person is not fully, but is to some extent, morally responsible, provides a better explanation of "tracing" cases, and reveals important requirements for a plausible theory of moral responsibility.

I. TRACY

Consider the following case:

> Tracy is nasty person who likes to hurt those she perceives to have wronged her. She believes that George has wronged her by getting a job that she thought she was the inside candidate for, and she is now planning her revenge. But rather than just kill George immediately, she decides to bide her time. She wants to kill George at the point at which it would be most harmful to do so – that is, when he has a good job, a family, and the like. Worried that she might change her mind, Tracy – who is a skilled, though nefarious, neuroscientist – creates a machine that will manipulate *her* to kill George, by implanting her with a genuinely irresistible and subconscious desire to kill him, *if* she has not already done so in the next fifty years. Tracy then erases her memory of creating the machine to safeguard against her changing her mind. It, however, has the unexpected side effect of also erasing the memories and beliefs gained in the previous few days, including those relating to George and the wrong she believed him to have done. Tracy, who is still a nasty person at this time, gets on with her business and attributes her loss of memory to drinking heavily, which was not uncommon for her to do at this time. Following a series of encounters with nice people, Tracy becomes less nasty. Over the next ten years, Tracy becomes nicer and nicer. Eventually all her previous nasty character traits and attitudes have gone. One day, she bumps into George and they reminisce about the job they had both applied for. Tracy remembers being up-

> set about not getting the job, but she remarks to George that she's glad she didn't get it because it would have only encouraged her bad character traits. Tracy and George fall in love, eventually marry, and have children. Marriage to George solidifies Tracy's good character traits and attitudes. Indeed, the kind of revenge she had planned for George is now unthinkable to her — that is, she has volitional necessities or moral incapacities that preclude her from normally being able to act in this kind of way [Frankfurt (1988); Williams (1993)]. Tracy regrets all her previous nasty behaviour — none of which was as bad as her long-forgotten plan to murder George. Forty years pass. Tracy, now 75 years old, has children and grandchildren. She is still such that her previous nasty behaviour is not psychologically open to her. The machine activates: it implants an irresistible desire in Tracy to kill George. Tracy kills George.

Is Tracy morally responsible for killing George?

Consider first what the tracing theorist would say: Tracy is not *directly* morally responsible because she fails to meet the relevant conditions on being directly morally responsible. For example, she is not at all reasons-responsive when she acts [e.g., Fischer and Ravizza (1998); Fischer (2012); McKenna (2012); Sartorio (2016)], and nor does she have an adequate opportunity to avoid acting as she does [e.g., Nelkin (2011); Brink and Nelkin (2013)]. However, because Tracy's action traces back to an earlier action of hers (creating the manipulation machine) for which she is directly morally responsible *and* she could reasonably foresee (indeed, she consciously intended) what the outcome of her action would be (killing George), Tracy is *indirectly* or *derivatively* morally responsible for killing George.

Now consider what the temporally extended action theorist would say: Tracy is morally responsible for killing George because this is the execution of a plan that she put into motion earlier in life. Her action started when she was twenty-five years old, and despite the lengthy gap, ended when she was seventy-five years old. This is no different than if the commando's bomb had been rigged to go off 50 years after he had planted it. Not only is Tracy morally responsible for killing George, but she also exercised control over killing him, her dearly beloved husband.

It seems implausible to me that Tracy *is* fully morally responsible for killing George and their family. It is true that Tracy at age twenty-five (Tracy-25) *is* fully morally responsible for killing George. It is true that she, at that time, put a plan into motion that aimed for and eventually resulted

in his death. And it is true that she lacked an excuse for doing so. However, I don't think Tracy *at age seventy-five* (Tracy-75) is fully morally responsible for killing George. She is victim of an extreme form of manipulation. The harrowing twist is that she has been manipulated by her earlier self — that is, she is a victim of *self*-manipulation. But it seems important that her earlier self is very different from her now: her earlier self has a completely different set of attitudes and so is practically distinct to her present self. So, I think that Tracy-75 has at least a partial excuse.

To further support the claim that Tracy is not fully morally responsible, first compare her to Jill. Between the time Jill takes the pill and destroys the china, she doesn't change at all. It makes sense, then, that Jill is fully morally responsible for destroying the china. However, between the time she initiates the plan and when the plan is finalised, Tracy changes dramatically. Such a dramatic, albeit incremental, change seems to make some kind of difference. While it makes sense to say that Jill is fully morally responsible, it doesn't seem to make sense to say that Tracy-75 is fully morally responsible.

This is compatible with holding that Tracy-75 is morally responsible *to some extent*. There does seem to be an intuitive difference between Tracy-75 and an other-manipulated agent. For example, suppose that Tony has the same kind of character and attitudes that Tracy-75 does. Suppose that Tony is also implanted with an irresistible desire to murder someone, and then proceeds to murder that person. Unlike Tracy-75, Tony was manipulated by other agents (e.g., nefarious neuroscientists). It seems intuitive to say that Tony is *less* morally responsible for the murder he commits than Tracy-75 is for the murder she commits (because it seems intuitive to say that Tony is not at all morally responsible for what he has done). So, self-manipulation seems to preserve moral responsibility to some extent. However, it is enough for my point that self-manipulation doesn't necessarily fully preserve moral responsibility, as this still suffices to undermine the standard account of self-manipulation.

In the next two sections, I will consider two ways one might try to debunk the intuitive judgement that Tracy-75 is not fully morally responsible but still morally responsible to some extent.

II. Debunking Strategy #1: Tracy-75 is a Different Person

I suspect that the most immediate response to this case will be to try to accommodate this intuition without accepting the implication that self-manipulation is sometimes not fully responsibility-preserving. And

there does seem to be an easy way for both the tracing theorist and temporally extended action theorist to do so. They can claim that Tracy-75 is a *different person* to Tracy-25. If this claim is true, it offers a way to explain why it seems intuitive that Tracy-75 is not morally responsible. She is not really a victim of self-manipulation, but rather just plain old manipulation: Tracy-75 was manipulated by another person. The fact Tracy-25 is morally responsible therefore has no impact on whether Tracy-75 is morally responsible.

But to speak of someone being a "different person" can mean different things [e.g., Shoemaker (1999), p. 397; Strawson (2015)]. One sense relates to metaphysical or numerical identity. This sense concerns a person's persistence conditions — that is, what it takes for a person, understood as an entity of some sort, to move through time and be correctly identified as the same persisting entity at different times.[1] The two leading approaches to metaphysical personal identity are the psychological approach and the biological approach. According to the biological approach, persons are essentially human animals; on this view, the persistence conditions of persons are just the persistence conditions of human animals [e.g., Olson (1997)]. According to the leading account of the psychological approach, personal identity is matter of unique psychological continuity [e.g., Parfit 1984].[2]

If we understand the "person" in "different person" to refer to a metaphysical sense of person, then Tracy-75 is the same person as Tracy-25. This is obvious on the biological approach because they are biologically continuous with one another — that is, they are one and the same human animal. This approach is compatible with no similarity in psychology whatsoever between the same person at two times. And according to the psychological approach, they are the same person because Tracy-75 is uniquely psychologically continuous with Tracy-25. On this approach, *X* is psychologically continuous with *Y* if and only if there are overlapping chains of *strong* psychological connectedness between *X* and *Y*; strong psychological connectedness holds when "the number of direct connections, over any day, is *at least* half the number that hold, over every day, in the lives of every actual person" [Parfit (1984), p. 206]. Thus, it is possible for two individuals at different times to have entirely distinct psychologies (e.g., different beliefs, desires, values, cares) but be the same person according to the psychological approach. Indeed, this is true according to any adequate account of metaphysical personal identity because a criterion of metaphysical personal identity must be transitive, and similarity relations are not transitive. This means that a person's

character can entirely change over a long enough period of time without the person going out of existence if that change is incremental rather than sudden. This means that tracing theory and the extended action theory cannot accommodate the intuition that Tracy-75 is not morally responsible by claiming that she is literally a differently person. Moreover, these theories wouldn't be able to capture the claim that Tracy-75 is *to some extent* morally responsible. If Tracy-75 is a literally different person, there does not seem to be a basis to explain this. This debunking strategy therefore fails.

III. Debunking Strategy #2: The Agent's Perspective

Another strategy for explaining away the intuition involves claiming that the intuition stems from not properly fleshing out the details of the case — in particular, Tracy-75's *perspective*. How would she feel? How would she respond? It seems clear that a person like Tracy-75 would be absolutely devastated for what she had done. Beyond her grief, it also seems plausible to me that she would feel guilt.

Does the fact she would feel guilt mean she is morally responsible for killing her husband and family? Not necessarily. While Tracy-75 might feel negative emotions that seem like guilt or remorse, it could also be that she is feeling fitting *agent-regret* – that is, the kind of emotion that a person feels over events she has caused but is not culpable for causing [e.g., Williams (1976); Baron (1988); Matheson (2017); Wojtowicz (2018), (2022)]. The classic case involves a lorry driver who, through no fault of her own, hits and kills a person. Daniel Jacobson (2013), however, argues that there is no such thing as agent-regret and the lorry driver instead might feel admirable but unfitting guilt. In other words, it is sometimes admirable for people to feel certain emotions when they realise they have caused harm, even if they also realise they are not culpable for causing that harm.

However we understand the emotion at play here, we can make sense of Tracy-75 experiencing negative emotions (beyond grief) upon killing her husband that does not imply she is (or finds herself to be) blameworthy for doing so. Consider also that it seems likely that a person who had been manipulated by another person to kill someone would also feel a raft of negative emotions – including agent-regret or admirable but unfitting guilt, as well as grief, horror, disgust, dismay, and shame – upon being manipulated in the manner Tracy-75 has been to kill another person.[3] So,

the fact that Tracy-75 might feel such negative emotions does not give us clear reason to think she must be fully morally responsible for killing her husband.

This explanation for Tracy-75's negative emotions seems plausible when we assume she does not know she herself devised the plan earlier in her life. Does it still seem plausible if we add to the story that Tracy-75 comes to discover that she is metaphysically or numerically identical to the person who manipulated her? I think so. To see this, imagine that you are manipulated in a similar manner to do something equally as harrowing as Tracy-75 and you then find out that a person that you happen to be metaphysically identical to but with whom you share no character traits or distinctive attitudes. Would this make you feel less or more responsible? Would you feel weaker or stronger guilt or regret? I suspect that, while it would be shocking that you are causally connected in more ways to the crime than you initially realised, you would still feel the same about the actions you have been manipulated to perform. You would share the mix of guilt, grief, horror, disgust, dismay, and shame that other-manipulated agents feel merely on the basis that one is a *cause* of heinous events. But I do not think you would feel a greater or lesser sense of ownership over your action because your earlier, metaphysically identical, self put this heinous plan into action. With respect to how to you relate to the action, I contend that you would feel no different than an other-manipulated agent would. Such disconnected past selves do not feel like us, even if we accept that they are, strictly speaking, us.

The crucial factor is that your earlier self will *seem* like a different person, even if you have just been told a true metaphysical story that says that they are the identical to you — that is, literally same entity as you but earlier in time. In other words, the radical change of character and attitudes – and the overall psychological disconnection – between you now and you back then will make your earlier self-seem like a different person. This is also true, I contend, for Tracy-75. We might therefore say that while her earlier self and her later self are metaphysically identical, her earlier self and later self are *practically distinct*.[4] So, it is *as if* another individual has manipulated her. The unfortunate twist is that it is in fact her earlier self. It is this practical sense of "different person" that I think our intuitive judgement that Tracy-75 is not fully morally responsible latches onto. While this conception of "person" might not be adequate to explain our essential nature as persons (or animals, or whatever) and how we persist through time, it does seem more suited to at least certain practical or normative questions. In particular, it seems to explain why

practically distinct self-manipulation does not preserve fully moral responsibility — namely, because the (earlier) self in question is practically distinct from the (later) self that is being manipulated.[5]

So, the standard view of self-manipulation is false. Freely and knowingly manipulating oneself does not necessarily fully preserve moral responsibility. Tracing theory and temporally extended action theory must be amended. One possible way to amend these views is to place a practical identity restriction on tracing and actions. I leave that task to tracing and extended action theorists. I have introduced the case of practically distinct self-manipulation so that we can investigate more deeply the nature of moral responsibility. In the next section, I argue that extant theories of moral responsibility struggle to explain why Tracy-75 is less than fully morally responsible but still morally responsible to some extent for killing George.

IV. Moral Responsibility

What does it mean to say that Tracy-75 is *more* morally responsible than Tony? While Tracy-75 is a victim of practically distant self-manipulation, Tony is a victim of good old fashioned other-manipulation. In order to ascertain the sense in which Tracy-75 is more morally responsible than Tony, I will first consider whether leading theories of moral responsibility can explain this intuitive judgement about Tracy-75. Through showing why these theories fail, I will motivate my alternative proposal.

IV. 1. *The Gateway Conception*

According to John Martin Fischer (2007), p. 185, moral responsibility is a gateway concept. This means that when a person is morally responsible, they are *in the running* for further responsibility practices such as praise and blame, but it is not yet the case that they are either praiseworthy or blameworthy. By comparison, on what Robin Repko Waller (2014) calls a thick conception of moral responsibility, a person who is morally responsible for an action is either praiseworthy or blameworthy, depending on the moral valence of the action.

If the gateway conception is correct, then it is not clear exactly what follows from a person being morally responsible for an action. It seems that we can say that the action belongs them in a particular kind of way, but it is not clear what exactly this means. Presumably, it means something more than just being a cause, given that causal responsibility is

weaker than moral responsibility. Indeed, it is also the case that being causally responsible is a kind of gateway to further responsibility practices, such as praise and blame. So, the "gateway" metaphor doesn't seem particularly illuminating.

Consider again self-manipulation cases. Is it normatively important if Tracy-75 is morally responsible in the gateway sense? It is not clear because it is not clear what exactly is at stake when one is morally responsible in this sense, other than the question of whether Tracy-75 instantiates a particular property. If it implies nothing else about her – that is, about whether she is blameworthy, owes an apology, has a duty to reform and to make amends, and the like – then it isn't clear there is any problem saying that she is fully morally responsible in this sense. This is just as it does not seem incorrect to say she is causally responsible for killing George and his family. And it is undeniable that she is causally responsible: she is the same entity that brought about this horrific event. Again, while moral responsibility in the gateway sense presumably means something more than merely being a cause, it does not seem normatively significant to say that a person is morally responsible in this sense.

Merely being morally responsible in this thin sense means that a person is not a fitting target of praise or blame or any reactive attitudes, they do not owe an apology, they do not owe compensation, and so on. The person must satisfy *further conditions* in order to become subject to any of these responsibility practices [Fischer (2007), p. 186]. So, even if Tracy-75 is (indirectly) morally responsible in this thin sense (she cannot be directly morally responsible in this way because she lacks control at the time she kills George), this does not help us explain the intuitive difference in responsibility between Tracy-75 and Tony because it seems that Tracy-75 is open to some, but not all, responsibility practices.

IV. 2 *The Tripartite Account*

According to Shoemaker (2015), there are three faces of responsibility: attributability, answerability, and accountability. If it turns out that Tracy-75 is responsible in one sense but not one or both of the others, this could explain why Tracy-75 is not fully morally responsible but still more morally responsible than Tony. However, I will now show that Tracy-75 is not responsible in any of these senses. So, we will have to look elsewhere to explain this intuitive judgement.

On Shoemaker's understanding, when a person is responsible in the attributability sense, we can appropriately attribute certain ethical predicates and fittingly feel certain emotions. A person who is responsible in

this sense might be vicious, kind, hard-hearted, callous, thoughtful, thoughtless, and so on. According to Shoemaker, this face of responsibility is grounded by *quality of character*. That is, it is grounded in an individual's attitudes where those, when appropriately clustered, constitute the individual's traits. For Shoemaker, the relevant attitudes are cares and values. Most minimally, to care about X is to be invested in X such that one will experience emotional highs when X flourishes and one will experience emotional lows when X flounders. Values, on other hand, are evaluative judgements about which courses of action, all-things-considered, an individual would prefer. To value X, then, is to judge that X is better to do, all-things-considered. On the basis of these attitudes constituting particular character traits, certain aretaic appraisals are justified. For example, suppose that George cares about and values his family. His cares and values will constitute certain traits, such as him being kind, generous, thoughtful, and so on.

Now consider Tracy-75. Is she responsible in the attributability sense for killing George and his (and her) family? At the time of action, she has no attitudes that would make her responsible in this sense. Her action is genuinely out of character. Because attributability is grounded in presently possessing particular cares and values, it cannot be coupled with either tracing theory or temporally extended action theory.

What about accountability? According to Shoemaker (2015), accountability involves reactive attitudes that are communicative in that they issue demands – in particular, a demand not to be disregarded. A person must be able to appreciate and internalise that demand to be an appropriate the target of it. In other words, the target must be able to take up the demandee's perspective, which involves a certain sort of empathy.

For Shoemaker, accountability is grounded by *quality of regard*. When an individual is accountable for A, where A is a "slight", it is fitting to respond with anger because A manifests an insufficient quality of regard (toward a particular individual). Anger is *communicative*: it communicates that inadequate regard has been shown. To show the adequate level of regard towards another – that is, not to slight them – is to *take them seriously*, where "my taking you seriously is a matter of the extent to which I take your specific normative perspective to bear a weight in my own deliberative perspective in the generally valenced way it does for you" [Shoemaker (2015), p. 97]. Those who are unable to take someone's specific normative perspective to bear on their own deliberation are, therefore, unable to be accountable for their actions. This is because they cannot take up the message that anger communicates. Those who have

certain *empathic* deficiencies (such as philosophical psychopaths) are unable to appreciates others' normative perspectives, and so are not accountable for their action.

At the time of action, Tracy-75 is certainly not accountable: she is not able to take up the perspectives of others at that time. But perhaps she is *derivatively* accountable. Perhaps because Tracy-25 is *directly* accountable for creating the manipulation machine (because she could take up the perspectives of others at that time) and she could reasonably foresee creating the machine would lead to her later self killing George and his family, Tracy-75 is derivatively accountable. However, this depends on Tracy-75 being a fitting target of accountability attitudes, such as anger, indignation, and resentment. However, because of her significant change of attitudes over time, while she was a fitting target of such attitudes at age 25, she cannot be a fitting target of such attitudes at age 75.

To see this, first consider blaming attitudes that are fitting on the basis of character traits, such as contempt. For contempt to be a fitting response to a person – that is, to accurately evaluate that person – the person must currently possess contempt-worthy traits. For example, a person might feel contempt for their slobbish roommate [Mason (2003), p. 249]. If a person lacks those traits, then there is nothing that makes contempt fitting. It would be a mistake to continue to feel contempt for a person who no longer has contempt-worthy traits, just as it would be a mistake to continue to feel fear about a bear that lost its teeth, claws, arms, legs, and anything else that made it fearsome. So, the fittingness of contempt depends on certain things presently being true of a person.

Of course, these are trait-focused attitudes. What about *action*-focused attitudes, including accountability attitudes? While such attitudes take actions as *part* of their focus, they are not exclusively focused on a person's actions.

Consider the distinction between an emotion's *particular* and its *formal* objects.[6] The former is the object towards which the emotion is felt. For example, the particular object of contempt is the person. Likewise, the particular object of indignation is also the person. We feel contempt and indignation about persons. The formal object is what makes the emotion fitting. For example, the formal object of contempt is a person's contempt-worthy traits. The slobbish roommate is worthy of contempt because they have contempt-worthy traits. Whereas the formal object of indignation seems to be person's action. However, it cannot just be the person's action. A person might perform an action and yet not be more than causally responsible for it. Indignation is not a fitting response to such ac-

tions. Further things must be true of an action for indignation to be a fitting response. It may seem to need to be a wrongdoing, but some, including Shoemaker, disagree that indignation is only fitting for wrongdoings. As noted, he holds that slights – which need not but can be wrongs – are the formal object of anger and its associated attitudes, such as indignation.

One way to understand slights is that they are actions that have a *morally bad meaning* [Wolf (2011)]. Such meaning might be a message of disrespect, an insult, or a threat towards others [Radzik (2009); Hieronomyi (2004); Hampton and Murphy (1988)]. Actions that are fitting targets of indignation might be wrongs, given that wrongs have morally bad meanings, but they might also be morally permissible acts that manifest a bad meaning. What, then, gives an action a morally bad meaning?

Actions can have meaning in different ways [Archer and Matheson (2019)], but the way that concerns responsibility depends on a person's attitudes, such as her intentions, her cares, and her values [McKenna (2012)]. So, while indignation might take a person's action as part of its formal object, it also takes *part of the person* as its formal object — namely, the part of the person that *confers the morally bad meaning on the act*. A person's traits and attitudes are the best candidates for meaning-makers because there does not seem to be anything else about the person that could confer meaning on acts — at least not in a sense that grounds accountability emotions.

A consequence of this is that changes in the character traits and attitudes of a person change what action-focused attitudes are fitting. This does not collapse action-focused attitudes into trait-focused ones. It is still the case that action-focused attitudes take actions as their formal object and trait-focused attitudes take traits as their formal object. My point is that action-focused attitudes take actions as their *primary* formal object and traits/attitudes as their *secondary* formal object — that is, such attitudes ascribe properties to a person's actions and to the person. Both action-focused and trait-focused attitudes, then, can cease to be fitting depending on changes in the person.

The upshot is that because Tracy-75 changes sufficiently – that is, she does not have the relevant cares and values – she cannot be directly or derivatively accountable for killing George and his (and her) family.

Similar considerations apply to answerability. Answerability involves how one regulates and judges one's attitudes, and it makes a different set of reactive attitudes fitting. Shoemaker (2015), p. 65, cites shame as a possible emotion rendered fitting by being answerable-responsible. But shame is trait-focused attitude like contempt. We have

already seen that it is uncontroversial that a person can be a fitting target of such an attitude at t_1 but cease to be a fitting target of such an attitude by t_2.

So, Tracy-75 does not remain responsible in any of the three senses that Shoemaker identifies. Therefore, his account does not provide a way to explain why Tracy-75 is not fully morally responsible but still morally responsible to some extent. We have to look elsewhere for an explanation.

IV. 3. *Duties and Responsibility*

The problem for Shoemaker is that his account is silent on an important aspect of our responsibility practices. He focuses entirely on what attitudes are rendered fitting. Because he finds different conditions under which different clusters of attitudes are fitting, he concludes that there are three distinct senses of responsibility. Indeed, his view seems to be a reductionist one: the conditions on responsibility just are the conditions on the fittingness of different reactive attitudes [see also Shoemaker (2017)]. It doesn't help him to claim that a person can cease being morally responsible for a past action they were once morally responsible for if they change traits and attitudes sufficiently. On this view [Shoemaker (2012); Khoury (2013), (2022); Matheson (2014), (2019)], Tracy-75 would turn out to be not at all morally responsible for killing George or for creating the manipulation machine. This is because, on this view, moral responsibility is ultimately (or purely) a matter of being a fitting target of the relevant reactive attitudes. Because Tracy-75 seems to be morally responsible to some extent and yet cannot continue to be a fitting target of negative reactive attitudes, we have to explain the way in which she is morally responsible by appealing to something other than reactive attitudes.

Responsibility is also associated with duties and entitlements. A person who acts wrongly, or otherwise substandardly, typically incurs reparative duties, such as the duty to apologise, the duty to reform, the duty to compensate, and the duty to remember. A person who does something exemplarily will often be entitled to things like awards and commendations. In other words, a consequence of responsible action is that sometimes we owe things to others, and at other times people owe things to us. What we therefore need is an account of responsibility that makes explicit both kinds responsibility practice — namely, reactive attitudes and reparative duties and entitlements.

An account of moral responsibility that gets closer to capturing this important aspect of our responsibility practices is Scanlon's (2015). He distinguishes between *moral response* and *substantive* responsibility. The

former includes responses to responsible actions — such as the reactive attitudes. The latter revolves around obligations a person has and obligations that others have to that person. While Scanlon is more explicit about his talk of obligations, he doesn't mention reparative ones. So, his account still isn't the right one to explain the intuitive judgement about Tracy-75.

In order to find the right kind of account, we can start by asking whether Tracy-75 has any reparative duties. Let's start by considering whether Tracy-75 owes an apology for killing George and his family.

First notice she is in a similar position to a person who unwittingly or unintentionally harms another. For example, the lorry driver, through no fault of their own, who hits and kills a pedestrian. As some have argued [e.g., Williams (1976); Capes (2019); Piovarchy (2020)], it is plausible that the driver owes an apology for killing the pedestrian. Not only does it make sense for the driver to feel agent-regret (or admirable but unfitting guilt), the driver also ought to communicate this feeling through its natural expression: an apology.

Second, notice that even though Tracy-75 is in a different position to the lorry driver because she – more precisely, her earlier self – created the circumstances in which she was now forced to kill George, and even though she has forgotten the plan that her earlier self devised and put into motion, it is still the case that she incurred an obligation for performing the action at the age of 25. Here I draw on Scanlon's (2015), p. 107; my emphasis) insight that, "substantive responsibility is in an important respect a *morally residual notion*". While Scanlon isn't referring to reparative obligations by "substantive responsibility", there is something important about the claim that duties are "morally residual". Unlike the fittingness of reactive attitudes – which are fitting because they are grounded in aspects of a person's attitudes – duties are perhaps such that they can remain despite many kinds of change in the person.

There are some duties that Tracy-75 has discharged without meeting. For example, her duty to reform. A person can have this duty discharged if their character and attitudes were reformed without their input, such as by being brainwashed by neuroscientists to have a better character. There would now no longer be any basis for this duty, and so it is discharged. But the person has not met the duty because they have not done anything to meet it. Even so, there is no longer any basis for requiring her to reform her character if it is already reformed.

While Tracy-75 has arguably met her duty to reform, she has not met her duty to apologise. Apologies are primarily communicative: they

communicate emotions about and one's stance towards an action, and often as well one's commitments to change and improve in light of the action being apologised for [see Smith (2008); Radzik (2009); Matheson (2017); Capes (2019)]. Tracy-75 has not done anything to meet this duty for her earlier act. While she might have changed and so cannot make commitments to change, she has not communicated her emotions about and stance towards her wrongful plan to kill George and his family. Her change of character and attitudes (that is, her reform) does not discharge of her duty to apologise, for this duty is a primarily communicative one. We therefore have a plausible candidate for a duty that Tracy-75 continues to have. This might, then, give us a basis for holding that Tracy-75 is morally responsible *to some extent* but not fully morally responsible.

One might respond at this stage that some blaming attitudes must be fitting for Tracy-75 to owe an apology. Apologies, after all, are thought to express emotions such as guilt and remorse. When a person owes an apology, she effectively, in part, owes these an expression of these emotions. It would be odd, so the objection goes, for a person to owe these emotions but for these emotions to be unfitting. But this is what the above view implies: Tracy-75 is not a fitting target of blaming attitudes, including self-blaming attitudes, but she allegedly owes an apology, nevertheless.

We've already seen the seeds of a reply to this worry. Earlier I noted that we might feel an unfitting emotion for admirable reasons. In short, we might feel an emotion because it is a valuable thing to feel and not because it correctly evaluates the world. Tracy-75 might then provide an apology that expresses emotions that are unfitting but given for admirable reasons — for example, to acknowledge her causal role in bringing about a harmful result.

There is one view of moral responsibility – in particular, of blameworthiness – that can make sense of this [e.g., Nelkin (2013); Tierney (2022); cf. Carlsson (2022)]. On this view, the root of blameworthiness is having reparative obligations. According to Tierney, a person remains blameworthy – and thus morally responsible (assuming a thick conception) – as long as she has unmet or undischarged reparative obligations.

A problem for this view is that it makes the fittingness of reactive attitudes *dependent* on having reparative obligations. I agree that meeting or discharging one's reparative obligations undermines the fittingness of these attitudes. Once a person meets her all her reparative duties – in particular, her duty to reform – there is now no longer any continued basis for the fittingness of these attitudes. But, as I've argued, it is possible for a person to cease being a fitting target of these attitudes while still

having reparative duties — such as the duty to apologise. While Tierney might point to the fact that we sometimes express the blaming attitudes towards those who continue to have reparative duties, this doesn't mean that these attitudes are fitting. They can instead just be appropriate because they help to enforce a person's reparative duties.

I propose that we understand the fittingness of the reactive attitudes and reparative duties as two distinct responsibility practices. The reason they might not seem to be distinct practices is that they often overlap — that is, a person is often both a fitting target of the reactive attitudes and has reparative duties because of an action she has performed. Consider the lingering pull towards holding that Tracy-75 is morally responsible to some extent. I propose that this is best explained by holding that Tracy-75 continues to have reparative duties of some kind for her earlier creation of the manipulation machine and its devastating consequences. Even though Tracy-75 has changed significantly since she was 25 years old, she continues to have wrongdoing-incurred reparative duties for her earlier actions because duties are morally residual.

V. NO NEED FOR TRACING

In this section, I will argue that the account of the nature of moral responsibility that I have proposed in this paper provides a better account of tracing cases in general. I'll focus on one kind of case: drunk driving.[7]

Suppose Bernie gets blackout drunk and then gets in his car and drives over a person, injuring them. Is Bernie morally responsible for injuring the person? Both the tracing theorist and the extended action theorist say yes. However, both fail to explain how Bernie can be a fitting target of negative reactive attitudes for more than just getting blackout drunk. This was certainly a reckless, irresponsible act. But it is significantly different act from running over a person.

Consider first the problem for the tracing theorist. It is not clear why being a fitting target of negative reactive attitudes *for an act* means that one is also a fitting target of reactive attitudes *for its consequences*. Those consequences, after all, do not change the nature of the reckless act: it remains just as reckless whether or not any harmful consequences follow from it. The tracing theorist therefore has an explanatory problem: why do actual consequences affect what attitudes a person is a fitting target of?

Consider now the problem for the extended action theorist. On this view, Bernie's running over of the person is part of a temporally extended action that started when Bernie got blackout drunk in the first place. Bernie can therefore be a fitting target of negative reactive attitudes because this temporally extended action is grounded in his own attitudes. But it's not clear why his drunken actions are part of one big temporally extended action, and why they are not just new actions undertaken by his drunken self. The extended action theorist therefore has its own explanatory problem: what unites actions under the banner of one temporally extended action?

One alternative to tracing theory and temporally extended action theory is to deny that there is resultant moral luck [e.g., Khoury (2012); (2018)]. On this view, we are only morally responsible for our actions (or willings) and not for the consequences of our actions (or willings). Consequences might, at best, be evidence of our moral responsibility for our actions (or willings). But this solution is controversial because there is a lingering worry that consequences make some difference to moral responsibility.

Another alternative is to hold that we can be fully and directly morally responsible for our drunken (and otherwise out of control) behaviour [e.g., Reis-Dennis (2018)]. On this view, we can be fully morally responsible for our drunken actions even if they do not disclose (or express) any aspect of our self. On this view, moral responsibility (in particular, blameworthiness) depends on its social effects. If we offend someone whilst drunk, we ought to feel bad and apologise for doing so, and others can fittingly blame us. This view requires holding that actions are the *exclusive* focus or object of blame. However, as discussed earlier, this involves an implausible account of the reactive attitudes, as it is undeniable that a person is, in some way, part of intentional object of blame.

The view I have defended avoids the problems with each of these proposals. It implies that the fittingness of reactive attitudes is not affected by the consequences of our actions, such consequences simply raise the salience of our actions and make it harder for us to deny that these actions make us the fitting targets of (positive or negative) reactive attitudes [cf. Khoury (2012)]. But it also implies that whether we have reparative duties can be affected by the consequences of our actions. The idea is, I think, simple enough: what we owe because of our wrongs (or what we are entitled to because our right actions) depends on what actually happens. But what it is fitting to feel about us doesn't depend on the consequences of what we do, but just on what we do (and who we are).

Even so, what happens can make others more aware of what it is fitting to feel about us.

VI. Conclusion

By distinguishing between two general responsibility practices – one which grounds the fittingness of the reactive attitudes and one which grounds reparative duties (and commendatory entitlements) – we can explain both why Tracy-75 is less morally responsible than Jill but still morally responsible to some extent *and* specifically more morally responsible than an other-manipulated person, such as Tony. Tony seems to lack moral responsibility completely, whereas Tracy-75 seems to be morally responsible to some extent. We can also provide a better explanation of tracing cases than is currently available: drunk drivers, for example, are not fitting targets of reactive attitudes for their drunken actions, but can gain reparative duties because of them. Finally, I have emphasised that accounts of moral responsibility must accommodate our two main responsibility practices. While I have argued that these practices are independent from one another, there is conceptual space for holding that attitudes are prior to duties, duties are prior to attitudes, or that they are mutually entailing. What cannot be denied, though, is that a theory of moral responsibility must explain both of these crucial facets of our responsibility practices.

Institut of Philosophie
University of Bern
Länggassstrasse 49a
3012 Bern, Switzerland
E-mail: matheson.philosophy@gmail.com

Acknowledgments

An earlier version of main case in this paper was presented at summer school on free will and moral responsibility organized by the Moscow Centre for Consciousness Studies in 2014. Parts of this paper were also presented at the The Joint Session of the Mind Association and the Aristotelian Society 2016, and Rocky Mountain Ethics Congress 2017. Thanks to participants at each event, including John Martin Fischer, Matthew Talbert, Daniel Miller, Sofia Jeppsson, Yishai Cohen, Artem Besedin, Anton Kuznetsov, for helpful feedback. I would also like to thank participants at the workshop in honour of Carlos

Moya held at the University of Valencia in 2022 for their feedback. Thanks also to Robert Hartman, Andrew Khoury, Daphne Brandenburg for comments on this or an earlier version of this paper. Thanks also to the many referees who took the time to comment on this and earlier versions of this paper. Work on the final version of this paper begun during my Maria Zambrano fellowship at the University of Valencia and was completed with the support of a SERI-funded ERC Starting Grant at the University of Bern under contract number M822.00083. ("SERI-funded" means it was directly funded by the Swiss State Secretariat for Education, Research and Innovation.)

NOTES

[1] If my wording seems to presuppose a particular account of persistence (e.g., endurantism), then readers should feel free to replace this for wording that fits another account (e.g., perdurantism). Nothing of substance hangs on this.

[2] There are other accounts of personal identity, such as bodily continuity theories [e.g., Williams (1970)], dualist accounts [e.g., Swinburne (1984)], constitution accounts [e.g., Baker (2000)], and embodied mind accounts [e.g., McMahan (2002); Parfit (2012); Ostaku (2017)]. As each account must posit a transitive criterion (otherwise these are not adequate accounts of *metaphysical* personal identity) which implies the possibility of complete psychological change [Khoury and Matheson (2018)], it is fine to use these two better known accounts as representatives of these two main approaches.

[3] For a fleshed out fictional example of this, see season 1 of either Marvel's *Jessica Jones* or *Twin Peaks*.

[4] If Shoemaker (2007) is right, then there may be a different "identity" relation for each practical concern. In my preferred language, there may be a different sense of practical identity for each practical concern. See also Matheson (2017).

[5] Does this talk of earlier and later selves sneak in some controversial views about persistence or personal identity? No. All views of persistence and personal identity must accommodate change – any view that cannot accommodate change is a non-starter – and so we need a convenient way to talk about a person at different times. We might then happily talk about earlier and later selves, person-stages, or a person-at-a-time without worrying we are sneaking in controversial metaphysical claims by the back door.

[6] See, for example, Kenny (1963), Teroni (2007), Scarantino and de Sousa (2018), and Kauppinen (2019).

[7] Philosophers use stylised versions of drunk driving cases because they typically assume that the drunk driver is completely out of control when they drive. However, it is arguably that many actual drunk drivers retain some level of responsibility relevant control, and so their blameworthiness for drunk driving can be explained without appealing to tracing or temporally extended action theory.

REFERENCES

ARCHER, A. & MATHESON, B. (2019), 'When Artists Fall: Honoring and Admiring the Immoral'; *Journal of the American Philosophical Association* 5 (2), pp. 246-265.

BAKER, L. R. (2000), *Persons and Bodies: A Constitution View*; Cambridge University Press: Cambridge.

BARON, M. (1988), 'Remorse and Agent-Regret'; *Midwest Studies in Philosophy* 13 (1), pp. 259-281.

BRINK, D. O. & NELKIN, D. K. (2013), 'Fairness and the Architecture of Responsibility'; *Oxford Studies in Agency and Responsibility* 1, pp. 284-313.

CAPES, J. A. (2019), "Strict Moral Liability" *Social Philosophy and Policy* 36 (1), pp. 52-71.

CARLSSON, A. B. (2022), 'Deserved Guilt and Blameworthiness Over Time'; in A. B. Carlsson (Ed.), *Self-Blame and Moral Responsibility*; Cambridge University Press, pp. 175–197.

FISCHER, J. M. (2012), *Deep Control: Essays on Free Will and Value*. Oxford University Press: Oxford.

FISCHER, J., KANE, R., PEREBOOM D. & VARGAS, M (eds) (2007); *Four Views on Free Will*; Wiley-Blackwell.

FISCHER, J. M. & RAVIZZA, M. (1998), *Responsibility and Control: A Theory of Moral Responsibility*; Cambridge University Press: Cambridge.

FISCHER, J. M. & TOGNAZZINI, N. A. (2009), 'The Truth about Tracing', *Noûs* 43, 3, pp. 531-556.

— (2011) 'The Triumph of Tracing'; in Fischer, J.M. *Deep Control*, Oxford University Press: Oxford.

FRANKFURT, H. G. (1988), *The Importance of What We Care About: Philosophical Essays*; Cambridge University Press.

FRITZ, K. G. (2014), 'Responsibility for Wrongdoing Without Blameworthiness: How it Makes Sense and How it Doesn't'; *Philosophical Quarterly* 64 (257), pp. 569-589.

HIERONYMI, P. (2001), 'Articulating Uncompromising Forgiveness'; *Philosophy and Phenomenological Research* 62, pp. 529–555.

JACOBSON, D. (2013), 'Regret, Agency, and Error'; in David Shoemaker (ed.), *Oxford Studies in Agency and Responsibility: Volume 1*, Oxford: Oxford University Press), pp. 95-125.

KAUPPINEN, A. (2019), 'Ideals and Idols: On the Nature and Appropriateness of Agential Admiration'; in A. Archer and A. Grahlé (eds) *The Moral Psychology of Admiration*, Rowman and Littlefield.

KENNY, A. (1963), *Action, Emotion and Will*; London: Routledge and Kegan Paul.

KHOURY, A. (2012), 'Responsibility, Tracing, and Consequences'; *Canadian Journal of Philosophy*, 42, 3-4, pp. 187-207.

— (2013), 'Synchronic and Diachronic Responsibility'; *Philosophical Studies*, 165, 3, pp. 735-752.

— (2018), 'The Objects of Moral Responsibility'; *Philosophical Studies* 175 (6), pp. 1357-1381.
— (2022), 'Forgiveness, Repentance, and Diachronic Blameworthiness'; *Journal of the American Philosophical Association*, 8 (4), pp. 700-720.
KING, M. (2014), 'Traction Without Tracing: A Solution for Control-Based Accounts of Moral Responsibility'; *European Journal of Philosophy* 22, 3, pp. 463-482.
MASON, M. (2003), 'Contempt as a Moral Attitude"; *Ethics*, 113,2, pp. 234-272.
MATHESON, B. (2017), 'More Than A Feeling: The Communicative Function of Regret'; *International Journal of Philosophical Studies* 25 (5), pp. 664-681.
— (2014), 'Compatibilism and Personal Identity'; *Philosophical Studies*, 170, 2, pp. 317-334.
— (2019), Towards a Structural Ownership Condition on Moral Responsibility; *Canadian Journal of Philosophy* 49 (4), pp. 458-480.
MCKENNA, M. (2012), *Conversation & Responsibility*; Oxford University Press: Oxford.
MCMAHAN, J. (2002), *The Ethics of Killing: Problems at the Margins of Life*; New York, US, OUP USA.
MILLER, D. J. (2017), 'Reasonable Foreseeability and Blameless Ignorance'; *Philosophical Studies*. 174, 6, pp. 1561-1581.
NELKIN, D. K. (2011), *Making Sense of Freedom and Responsibility*; Oxford, GB: Oxford University Press.
OLSON, E. (1997), *The Human Animal: Personal Identity Without Psychology*; New York, Oxford University Press.
OTSUKA, M. (2017), 'Personal Identity, Substantial Change, and the Significance of Becoming'; *Erkenntnis* 83 (6), pp. 1229-1243.
PARFIT, D. (1984), *Reasons and Persons*; Oxford: Oxford University Press.
— (2012), 'We Are Not Human Beings'; *Philosophy*, 87, 1, pp. 5-28.
PIOVARCHY, A. (2020), 'Blame in the Aftermath of Excused Wrongdoing'; *Public Affairs Quarterly* 34 (2), pp.142-168.
RADZIK, L. (2009), *Making Amends*; Oxford University Press.
REIS-DENNIS, S. (2018), 'Responsibility and the Shallow Self'; *Philosophical Studies* 175 (2), pp. 483-501.
SARTORIO, C. (2016), *Causation and Free Will*. Oxford, United Kingdom: Oxford University Press UK.
SCANLON, T. (2008), *Moral Dimensions: Permissibility, Meaning, Blame*; Belknap Press of Harvard University Press.
SCARANTINO, A. & DE SOUSA, R. (2018), 'Emotion'; in E. N. Zalta (ed.) *The Stanford Encyclopaedia of Philosophy*. *<https://plato.stanford.edu/archives/win2018/entries/emotion/>*
SHOEMAKER, D. (1999); 'Selves and Moral Units'; *Pacific Philosophical Quarterly* 80 (4), pp. 391-419.
— (2012), 'Responsibility Without Identity'; *Harvard Review of Philosophy*, 18, 1, pp. 109-132.

— (2015), *Responsibility from the Margins*; Oxford University Press.
— (2017), 'Response-Dependent Responsibility; or, A Funny Thing Happened on the Way to Blame'; *Philosophical Review* 126 (4), pp. 481-527.
SMITH, N. (2008), *I Was Wrong: The Meanings of Apologies*; Cambridge University Press.
SWINBURNE, R. (1984), *Personal Identity: Great Debates in Philosophy*; (eds: Shoemaker, Sydney & Swinburne, Richard), Blackwell: Oxford.
TERONI, F. (2007), 'Emotions and Formal Objects'; *Dialectica*, 61(3), pp. 395–415.
TIERNEY, H. (2022), 'Don't Suffer in Silence: A Self-Help Guide to Self-Blame'; in A. Carlsson (ed), *Self-Blame and Moral Responsibility*, Cambridge: Cambridge University Press, pp 117-133.
TIMPE, K. (2011), 'Tracing and the Epistemic Condition on Moral Responsibility'; *Modern Schoolman* 88, 1/2, pp. 5-28.
WALLER, R. R. (2014), 'The Threat of Effective Intentions to Moral Responsibility in the Zygote Argument'; *Philosophia* 42 (1), pp. 209-222.
WILLIAMS, B. A. O. (1970), 'The Self and the Future'; *Philosophical Review* 79 (2), pp. 161-180.
— (1976), 'Moral Luck'; *Aristotelian Society Supplementary Volume* 50 (226), pp. 115-151.
— (1993), *Shame and Necessity*; University of California Press.
WOJTOWICZ, J. (2018), 'Bernard Williams on Regarding One's Own Action Purely Externally'; *Journal of the American Philosophical Association* 4 (1), pp. 49-66.
— (2022), 'Agent-Regret, Accidents, and Respect'; *The Journal of Ethics* 26 (3), pp. 501-516.
WOLF, S., (2011), "Blame, Italian Style", in Wallace, Kumar, and Freeman (eds.), *Reasons and Recognition: Essay on the Philosophy of T. M. Scanlon*, New York: Oxford University Press