

# Neoptolemus and Huck Finn Reconsidered. Alleged Inverse *akrasia* and the Case for Moral Incapacity

Matilde Liberti<sup>1</sup>

Accepted: 20 March 2024 © The Author(s) 2024

Cases of *akratic* behavior are generally seen as paradigmatic depictions of the knowledge-action gap (Darnell et al 2019): we know what we should do, we judge that we should do it, yet we often fail to act according to our knowledge. In recent decades attention has been given to a particular instance of *akratic* behavior, which is that of "inverse *akrasia*", where the agent possesses faulty moral knowledge but fails to act accordingly, thus ending up doing the right thing. In particular, two literary examples are considered as exemplifying this kind of *akratic* situation: Huckleberry Finn (Arpaly & Schroeder 1999, Arpaly 2000, Hursthouse 1999, Kleist 2009, Holton manuscript) and Neoptolemus as understood by Aristotle (NE; Arpaly & Schroeder 1999). In this paper I will argue that those of Neoptolemus and Huck Finn are not cases of inverse *akrasia* (Holton manuscript) but are much better explained as instances of what Williams (1993) called "moral incapacity". In particular, the reason why they fail to act according to their original judgments is due to a lack of motivation to act accordingly, which is grounded in their moral self-identities (Blasi 1984; Vigani 2016).

The paper will unfold as follows: I will, first, argue that neither Neoptolemus nor Huck Finn show *akratic* behavior; thus, they cannot be legitimately labeled as "inverse *akratics*" (par 1.); then, I will argue that they act the way they do notwithstanding their faulty judgments because they are effectively motivated to do so. Such motivation originates in their moral self-identities and is experienced through the threat of self-betrayal (par 2); finally, I will argue that when an agent is motivated to act in a way that is integral with her moral self-identity, acting otherwise is experienced as something one *cannot* ultimately do; that is, as a moral incapacity (par 3). This "cannot" is neither a metaphor for an "I shouldn't", nor an instance of what has been recently labeled a "moral impossibility" (Caprioglio Panizza 2020, 2021), since it does not arise from the normative force of deontic judgments, and it is neither physically nor psychologically impossible for the agent to act otherwise.

Published online: 10 May 2024

Northwestern Italian Philosophy Consortium, University of Genoa, Genoa, Italy



Matilde Liberti libertimatilde\_unige@outlook.com

#### 1 Inverse Akrasia

The most influential account of inverse akrasia is that of Arpaly & Schroeder (1999), who define it as an instance of akratic action that "reverse[s] our usual expectations from akratic action" (162). While from akratic action we expect the wrong action. from inverse akratic behavior the outcome is, surprisingly, good. What is crucial in Arpaly & Schroeder's (1999) depiction is the fact that the good outcome is surprising because the agent is ignorant of the good – or, to put it differently, she does not know that her actions are ultimately good (e.g. Huck Finn who thinks that he is doing something wrong helping Jim). Kleist (2009) focuses on the element of irrationality that is peculiar of akratic behavior and states that "an inverse akratic act is one in which someone believes X, all things considered, is the correct act, and yet performs ~X, where ~X is the correct act" (257). Thus, he stresses the element of irrationality that is shared between cases of akrasia and inverse akrasia. The only difference between akrasia and inverse akrasia seems to be that in the former case. acting incoherently leads to bad outcomes, while in the latter acting incoherently leads to good outcomes; one may even think that the inverse akratic does the right thing somewhat accidentally<sup>1</sup>, since she does not know that what she is doing is right.

In light of this, let us draw a general definition of inverse *akrasia*: it is an instance of incoherence between the agent's moral knowledge (broadly – beliefs, judgments, deliberative premises), and her ultimate actions. In particular, her moral knowledge is faulty, while her actions are good. This is why that of Huck Finn is generally seen as a good example of inverse *akratic* action: Huck somewhat *knows* that slavery is a non-controversial, legit social system – that's what he had been taught, and he never questions it. Yet, apparently, he acts incoherently: he helps Jim eluding the slave hunters, and hides him multiple times throughout the book; crucially, he does so while maintaining that he is doing the wrong thing (Hursthouse 1999). How can we make sense of this?

If we consider Aristotle's structure of *akratic* action, we can see that inverse *akrasia* does seem to share the same psychological blueprint of *akrasia*: (i) it is action against one's *prohairesis*; (ii) it is voluntary; (iii) it is done out of some sort of ignorance; (iv) it is marked by conflict. In what follows I will consider Neoptolemus' and Huck Finn's cases and challenge (i).

#### 1.1 Neoptolemus

As just mentioned, according to Aristotle, *akratic* action is action against one's correct *prohairesis* (NE 3.2, 1111b13-15; NE 7.8, 1151a5-7; NE 7.8, 1151a29-33). The meaning of this claim is fairly obscure; some scholars interpret it as action against one's decision to perform particular actions (Irwin 1986; Wiggins 1978; Davidson 1980), while others, most notably Cagnoli Fiecconi (2018), interpret it as action against "one's general commitment to act on one's conception of one's ends overall"

<sup>&</sup>lt;sup>1</sup> As it may be the case with Arpaly's *Emily* (2000).



("broad" conception of *prohairesis*; 2018: 2). Practically speaking, according to the former interpretation, the *akratic* agent is the one who decides she will not eat the second slice of cake, yet she eats it; according to the latter, the *akratic* agent is the one who is committed to being healthy, yet she leads an unhealthy life made of particular choices against her general commitment to health, e.g. eating the second slice of cake. The difference is in the scope of *prohairesis*: a "narrow" conception sees it as particular choices, a "broad" conception sees it as commitments to the good.

I am not here disputing the correct interpretation of Aristotle; for the purpose of this paper it suffices to note, with Cagnoli Fiecconi (2018), that there are very good reasons for believing that in his discussion of akratic behavior, Aristotle employs a broad conception of prohairesis: to name one, when distinguishing the weak from the impetuous akratic (ἀσθένεια and προπέτεια, NE 7.7, 1150b18-28), Aristotle claims that the impetuous does not form a decision to perform an action (NE 7.8, 1150b19-23); yet, he acts akratically. Which means that he manifests irrationality, but not between behavior and particular choices. Furthermore, when he discusses stubborn opinionated agents (ἰδιογνωμονέω), Aristotle claims that they may seem enkratics, since they stick to their choices, but the stubborn agent actually has more to share with akratic agents, since they too act against their prohairesis (NE 7.9, 1151b5-16). Finally, as we will see shortly, when introducing the tragic hero Neoptolemus (NE 7.9, 1151b18-22), Aristotle does not classify him as inverse akratic<sup>2</sup>, since even though he acted against a particular choice, he ended up behaving in accordance with his prohairesis. Only if we concede that, at least in these cases, Aristotle has a broad conception of prohairesis in mind that does not limit it to particular choices, can we make sense of his claims on akratic behavior. In what follows I shall, thus, consider *prohairesis* as interpreted by Cagnoli Fiecconi (2018).

Aristotle describes *prohairesis* as a deliberative desire for things that are up to us (NE1113a10-11). We wish for some ends, we presuppose that those ends are good (NE 3.4, 1113a23-b2), then we deliberate about pursuing them (NE 3.3, 1113a1-15). *Prohairesis* is, thus, not deliberation on some ends, because ends are already presupposed to be good; but it is, rather, a rational desire to pursue those ends. In other words, "we wish to be healthy, but we do not form a *prohairesis* to be healthy. Rather, we form *prohairesis* to do things that make us healthy [...] (NE 3.2, 1111b29-30)" (3). It is, thus, "always something for the sake of something else" (4; EE1227b36-7). If I form a *prohairesis* to help my friend move house for the sake of helping her, I reveal wishes, goals and dispositions that are different from those that would be revealed by, say, me forming a *prohairesis* to help my friend for the sake of appearing good in the eyes of someone else. My goals, in particular, would be different in an important way; in the former case, the goal would be that of being a helpful friend, in the latter case, the goal would be that of appearing good to others<sup>3</sup>.

<sup>&</sup>lt;sup>3</sup> In both cases the goal would be presupposed as good, and *prohairesis* would take the shape of a deliberative desire to perform actions to achieve such goals. It is in this sense that *prohairesis* is character revealing (Rhet. 1366a14-16; Fiecconi 2018).



<sup>&</sup>lt;sup>2</sup> Arplay & Schroeder (1999) employ Neoptolemus' case as an exemplification of inverse *akratic* behavior, which is indicative of the fact that what they have in mind is a "narrow" conception of *prohairesis*. Their account does not consider the details of Aristotle's discussion of Neoptolemus' case.

If I form a *prohairesis* to help my friend for the sake of helping her, I reveal my commitments to being a helpful friend; if I do not help her in the end, then I show *akratic* behavior. Let us now see how this may work for cases of inverse *akrasia*; according to this picture, the inverse *akratic* agent is the one who forms an incorrect *prohairesis*, but then acts against it (or fails to act accordingly), thus ending up doing the right thing. In particular, according to a broad conception of *prohairesis*, the inverse *akratic* is the one who performs actions against her commitments to act on what she incorrectly deems morally good. As mentioned above, when discussing possible instances of *akratic* behavior, Aristotle considers the case of Neoptolemus as depicted in Sophocle's *Philoctetes* and claims that, although it may seem like a case of (inverse) *akratic* behavior, it is actually not so.

In the play, Odysseus and Neoptolemus sail off to the isle of Lemnos with a specific mission: they need Philoctetes, along with his bows and arrows, in order to finally conquer Troy. However, Philoctetes would not be easily convinced to join the mission to sack Troy; he was abandoned on Lemnos, wounded and in pain, nine years earlier, by orders of Odysseus himself. He deeply resents him and would never be willing to give him his precious bows and arrows. Odysseus knows all that and convinces Neoptolemus to trick Philoctetes into believing that Neoptolemus too hates him, so that he could build the necessary trust to get both Philocettes and his weapons. Odysseus recognizes that Neoptolemus' nature is honest, not inclined to such mischief, but he insists that victory is such a high price that he can bend his fair nature just for once [100-110]. Neoptolemus, after some hesitation, firmly sets to find Philoctetes and manages to trick him. However, when it comes to sailing off with Odysseus, Neoptolemus does something unexpected: he faces Odysseus, telling him that he is going to give the weapons back to Philoctetes and convince him to join the mission on his free will. It is important to note that Neoptolemus, as the son of Achilles, is the perfect Greek hero; he indeed feels pity and compassion for Philoctetes, but his main concern is for the status of his honor. Tricking Philoctetes is a shameful action, as opposed to, say, engaging him in a deathly fight.

Aristotle claims that Neoptolemus does not actually show *akratic* behavior: among those who do not firmly stick to their choices, not all who fail to stick to their choices do so because they are incontinent (ἀκρατής): "[ό Νεοπτόλεμος]. καίτοι δι ἡδονὴν οὖκ ἐνέμεινεν, ἀλλὰ καλήν: τὸ γὰρ ἀληθεύειν αὖτῷ καλὸν ἦν, ἐπείσθη δ' ὑπὸ τοῦ Ὀδυσσέως ψεύδεσθαι" (NE 7.9, 1151b18-22). In this passage Aristotle explains that it was indeed due to pleasure that Neoptolemus could not stick to his original choice, but to a kind of pleasure that was "morally beautiful"; this is crucial, since the incontinent is the one who, on the other hand, responds to passions that are "ugly" – that is, either morally dubious or pure appetites. Now, since Neoptolemus was initially convinced to lie, Aristotle cannot classify him as fully "virtuous", but he goes on to claim that those who follow passions are neither intemperate, nor vicious, nor, as we have seen, incontinent; only those who follow *shameful* passions belong to such categories<sup>4</sup>. If we add that Neoptolemus' *prohairesis* broadly understood takes the form of a deliberative desire to perform actions for the sake of being

 $<sup>^4</sup>$  οὐ γὰρ πᾶς ὁ δι' ἡδονήν τι πράττων οὕτ'ἀκόλαστος οὕτε φαῦλος οὕτ' ἀκρατής, ἀλλ' ὁ δι' αἰσχρά (NE 7.9, 1151b21).



honorable, and if we add the fact that, as we noted before, honor was, for an ancient Greek, something morally good, then we can see that Neoptolemus' action is performed *in accordance with* his *prohairesis*, and not against it. It is, thus, not *akratic* behavior (actions against one's good *prohairesis*), nor, crucially, inverse *akratic* behavior (actions against one's bad *prohairesis*).

#### 1.2 Huckleberry Finn

In his unpublished manuscript Richard Holton argues that the case of Huckleberry Finn is one of weakness of the will, but not one of akrasia. The difference can be summarized as follows: akrasia involves action against one's best judgment, while weakness of the will involves "an over-ready revision of one's resolution" (3). Huck Finn, as Holton reads him, is indeed quick in giving up his resolution to give Jim in ("judgment shift") due to his call of conscience. By the time he acts he has already shifted his judgment from giving Jim in to helping Jim, thus, he does not seem to be acting against his previous judgment. In other words, he does not show akratic behavior - specifically, since the previous judgment would have been a faulty one, he does not show inverse akratic behavior. Holton supports his reading of Huck's weakness of will through cognitive dissonance theory: agents seem to be quick to reinterpret previous states of minds in order not to feel dissonance between judgments and inclinations (5). We want to have a coherent, integral picture of ourselves, and when we feel inclined to act in a way that does not cohere with our self-image, we tell ourselves narratives in order to adhere once again to an integral picture of who we are (Karniol & Miller 1983; Harmon-Jones & Harmon-Jones 2007). Imagine the following scenario: I judge that at tonight's party I prefer to refrain from eating too much. I have already been eating a lot lately, I find it hard to fall asleep due to digestion, and, generally, I feel fatigued by all the food consumption I have undergone. Then, I arrive at the party and I immediately feel tempted by all the food I see in front of me. This temptation does not cohere well with my original judgment; due to its pressure I begin to tell myself a story that would make sense were I to eat more food than previously decided - "This is just a one-time party, my friends have put so much effort into preparing this food. I actually have reasons to dig in". This is what Holton calls "judgment shifts": by the time I act on my temptation, my original judgment has already changed in order to avoid cognitive dissonance (7). Crucially, the temptation does not coincide with the action of over-eating but, rather, with the reconsideration of my original judgment that I should not eat too much (9). Thus, I do not show akratic behavior; rather, a form of weakness of will. Truly akratic behavior would be acting against my original judgment just out of pure desire, without having reconsidered it, and this, according to Holton, is a much rarer phenomenon than what philosophers would allow. Back to Huck Finn failing to turn Jim in, he writes:

Huck thinks he has shown weakness of will, and I think we should agree with him; admirable though his inability is, it is hard to see the revision as stemming from any new information or insight of the kind that should rationally have him



reconsider. [...] But if Huck Finn is weak-willed, is he *akratic*? That depends on whether, at the time he has acted, he believed that turning Jim in was the right thing to do. And I think that we have plenty of evidence to think that the answer to that is far from clear (16).

Holton ends the manuscript suggesting that something important has indeed played a role in Huck's seemingly unreflective change in attitude; "it matters that he is acting, as he puts it, from something like a visceral sense of equality" (16); this would also explain why he does not merely lie to the slave hunters, but he also goes on to tell them a story that would send them away for good. There are two forces that make *akrasia* rare: the first is that of adhering to desires. Huck Finn desires to protect his friend Jim, thus he shifts his judgment in order not to be dissonant with such desire. The second Holton does not fully explain, but it sounds just like the plain force of being motivated to doing good. This motivation is what drives us to tell ourselves stories that put us in a good light, but also to making sense of what it is that is ultimately right independently of what others have taught us (e.g. Huck Finn rejecting Miss Watson's morality; Holton manuscript: 17), or what our conscience itself tells us we should do (e.g. Huck Finn's conscience telling him to turn Jim in; Holton manuscript: 14).

In what follows, my aim is that to take it where Holton left it and offer a novel interpretation of alleged cases of inverse *akrasia* that makes sense, particularly, of the latter force – that which drives us towards doing good notwithstanding opposing, often entrenched conditions. I agree with Holton that inverse *akrasia* is not a common phenomenon, but more needs to be said as per why ultimately agents like Huck Finn and Neoptolemus act well. What is the origin of their motivation to act the way they do? If their actions are in contrast with their moral knowledge and if they are not *akratic* – that is, they are not purely irrational – how can we make sense of them? In the next section I will argue that the origin of their motivation to act well is to be found in their moral self-identities.

# 2 Moral Self-identity

Moral self-identity has been explored both from a philosophical and a psychological perspective. Broadly, moral self-identity is the form of our will (Frankfurt 1971; Taylor 1989); it is actively shaped by what we care about most and, at the same time, it shapes the kind of moral agents we are (Blasi 1984). For this reason, it is worth exploring if we wish to find the origin of the kind of motivation that guides Huck Finn and Neoptolemus. I will start from the philosophers and then introduce Blasi's psychological account.

#### 2.1 Why Do They Fail to Act? Motivation and the Centrality of Commitments

According to Frankfurt (1971) and Taylor (1989), what moves all the way to actions are, for the former, "second-order volitions" and, for the latter, commitments. For



Frankfurt, one can have first-order desires ("A wants to X") and second-order desires ("A wants to X", where X is a first-order desire), but when it comes to choices, one needs to have second-order volitions in order to choose willfully, where second-order volitions are second-order desires in which the "want" is motivationally charged. Thus, if I have a second-order desire, it does not imply that that very same desire is also a second-order volition: in perfect akratic style, I may want to want to go for a run, yet not go for a run because, in the end, I am much more effectively moved by my desire to watch YouTube videos. However, if the second-order desire is effective - that is, it actually moves me all the way to action - then I am exercising my will. Taylor (1989) goes in a similar direction by claiming that, in order to choose willfully, one needs to have a kind of "moral map", which is offered by one's deep-seated commitments. According to Taylor, if I did not have any such commitments, I would not possess any map to navigate my moral possibilities and, thus, I would not be able to choose willfully between options. Note that the suggestion is not that when one has commitments, then the choice is easy; quite the contrary, it is very likely to be extremely hard. The claim here is that, if one does not have deepseated commitments or second-order volitions, it is *impossible* to choose willfully, not that it is hard.

The idea that second-order volitions and commitments ground one's moral selfidentity was taken seriously by psychologist Augusto Blasi (1984). Inspired by both Frankfurt and Taylor, Blasi developed his seminal theory of moral self-identity, arguing that when one's moral commitments are central to who we are as agents, then our moral self-identity is constructed on such commitments. To give an example, I may have a first-order desire of the kind "I want to be compassionate" and also a second-order desire of the kind "I want to want to be compassionate"; but in order for this second-order desire to be a second order volition – that is, to move me all the way to action – I need to actively endorse it, to place it at the foundations of my agency. It is around such actively endorsed volitions that my moral self-identity takes shape – and, of course, can be re-shaped<sup>5</sup>. As noted by Lapsley (2008), not everyone constructs one's self-identity around moral concerns; that is, people may have moral concerns, but such concerns need not be central to the construction of their identities: "[S]ome have only a glancing acquaintance with morality but choose to define the self by reference to other priorities; or else incorporate morality into their personality in different degrees; or emphasize some moral considerations ("justice") but not others ("caring")" (2008: 35). From this we can conclude that there is a variety of self-identities; those that, along with possibly other concerns, have moral considerations at their core, may be categorized as peculiarly "moral" self-identities:

moral identity is a dimension of individual differences, which is to say, it is a way of talking about personality. One has a moral identity to the extent that moral

<sup>&</sup>lt;sup>5</sup> Self-identity need not be construed accidentally, or unconsciously – in order to "save" the rational element of morality (Lapsley 2008), Blasi argued that at least a significant part of the construction of self-identity can be organized around the influence of moral reasons: "[F]undamentally, [...] the direction of influence would be from moral understanding to moral identity, rather than the other way around " (1984: 138). Moral commitments are affective, but also susceptible to reasons; thus, the possibility of rational moral choice is maintained (Lapsley 2008: 35).



notions, such as being good, being just, compassionate, or fair, is judged to be *central*, essential, and important to one's self-understanding. One has a moral identity when one strives to keep faith with identity-defining moral commitments; and when moral claims stake out the very terms of reference for the sort of person one claims to be (Lapsley 2008: 35, my italics).

If my second-order volition to be compassionate is foundational to my moral self-identity, then I will feel compelled to being compassionate, or at least to try as much. My integrity would depend on my being faithful to such commitment to compassion, lest I would experience my actions as ones of *betrayal* (Blasi 1984, 2004; Lapsley 2008). The urge not to avoid betraying ourselves is the source of motivation to act in accordance with our core second-order volitions and commitments (Vigani 2016).

Blasi's self-model bridges the gap between judgments and action through the following components: judgment of responsibility, moral identity and self-consistency (Darnell et al 2019; Vigani 2016; Blasi 1984). When faced with a choice, the agent begins by filtering a moral judgment through a judgment of responsibility towards herself; in other words, she asks herself whether this action is necessary in order not to betray herself. Whether she will feel that she needs to pursue it and to what extent she will feel motivated to do so, depends on the centrality that moral concerns have in her moral self-identity. Finally, if she sees it as a necessity, she will also experience the drive to be self-consistent (Darnell et al 2019: 6). Those who reliably act on their judgments, do so because "not to act according to one's judgment should be perceived as a substantial inconsistency, as a fracture within the very core of the self" (Blasi 1983: 201). Thus, when second-order volitions and commitments are central to one's moral self-identity, they move all the way to action (Frankfurt 1971); they are effective in motivating.

However, it is still not clear how the agent experiences the urge of self-consistency. Saying that one needs to be faithful to one's deep-seated commitments in order not to betray one's self may just be a romantic metaphor for the mere fact that one decides that such commitments are to be followed and then acts on the normative force that such decision implies. In this case, saying that the agent needs to act in order to be consistent would merely – and redundantly – mean that she feels that she *should* act in order to be consistent. When considering motivation to internal self-consistency, Blasi identifies a set of skills that are needed in order to be integral with one's self ("integrity skills", e.g. being truthful, transparent, avoiding self-deception, etc.; Lapsley 2008). It is of course true that an agent feels responsible to be integral, but if her commitments are *central* features of her moral self-identity (that is, they are its foundations) and not peripheral, then she will have second-order volitions to live up to them; that is, she will *desire*<sup>6</sup> to live up to them, and it is this desire that Blasi identifies as generating not just the responsibility to be integral, but also and

<sup>&</sup>lt;sup>6</sup> See Vigani (2016) for a thorough discussion on how Blasi's identity-based motivation can account for our mixed intuitions on motivational externalism and internalism. Not all moral judgments are motivating, but those that are central to one's moral self-identity are indeed so. Similarly, see Kristjannson (2013) on a hybrid account of Aristotelian motivation that is compatible with Blasi's model, where externalism explains the continent, while internalism explains the virtuous.



primarily as the *necessity* to be so (Blasi 2005; Lapsley 2008: 36-37). As succinctly explained by Lapsley (2008): "Integrity is felt as identity when we imbue the construction of self-meaning with moral desires" (37).

From this we can conclude that depending on the degree of centrality to the construction of my self-identity that commitments possess, there will be things that I will feel I *should* do and that I *should not* do – when commitments are imbued with normative force – but also things that I *cannot* do, when commitments are imbued with moral desires. The stress is both on "cannot" and "I": "cannot" involves an incapacity to pursue a certain course of action given my moral self-identity, and the "I" expresses the fact that it is *me* who possesses this incapacity because were I to act on it, I would not be myself anymore (maybe someone else could, Bauer et al 2017)<sup>7</sup>. As nicely put by Bergman (2002):

The best answer to the question, Why be moral?, may thus be, Because that is who I am, or, Because I can do no other and remain (or become) the person I am committed to being. Commentators on the Shoah often observe that both rescuers of Jews and those who refused to take such risks on behalf of desperate and hunted strangers have explained their behavior in similar words: "'But what else could I do?' "(Monroe 1994, p. 201). Everything depends on how the "I" understands itself and its responsibilities (Bergman 2002: 123).

The "could" in "what else could I do?" is not a metaphor: there was nothing else *they* could have done. People may contend that everyone should behave as they did, which is hard to deny, but also hard to live up to. Maybe sometimes the normative force of principles is not strong enough: if commitments are not central to one's moral self-identity, then they are not as effective in issuing action. It is when they are central that one feels not only that there is an underlying principle with normative force, but also that *they* cannot do otherwise.

#### 2.2 Neoptolemus' and Huck Finns' Motivations

To summarize, we have seen that when second-order volitions and commitments are central to one's moral self-identity, then they are the source of moral motivation. Let us now see if all this applies to our examples. We know that Neoptolemus is committed to being honorable, and this commitment is central to his moral self-identity (he would much rather engage in a fight than deceive<sup>8</sup>). We can see this in the fact that Odysseus manages to convince him precisely by stressing how honorable it



<sup>&</sup>lt;sup>7</sup> Darnell et al (2019) challenge appeals to moral self-identity when it comes to bridging the knowledge-action gap, since moral identities may just be confabulations, or self-deceptions on our ends (2019: 8). If the necessity is experienced exclusively as a "should", then the risk identified is real. On the other hand, if the necessity is experienced also negatively, as an "I cannot", then the risk of self-deception does not stand. In such cases, the agent herself may find out something authentic about her moral apparatus, independently of her own confabulations on herself. This will be addressed in the final section.

<sup>8 &</sup>quot;NEOPTOLEMUS: Son of Laertes, I hate to carry out an order which it hurts to listen to. It's not my nature to do anything based on deceit" [100-110].

would be to be the heroes who finally conquer Troy<sup>9</sup>. Neoptolemus is surely guided by his commitment to honor; he just does not seem to know that he is not capable of being dishonorable, not even in the face of the promise of a much greater honorable deed (at least, in the eyes of the Greek hero)<sup>10</sup>. The possibility that an agent may not be aware of what grounds her moral self-identity is touched upon by Blasi (1999) when talking about moral motivation and emotions: it is true that motivation arises from commitments, but such commitments have an emotional element that makes them liable to the spontaneity and uncontrollability of emotions. Thus, it is perfectly possible that an agent may feel the urge to live up to her commitments, without fully understanding either their content or the actual centrality of such commitments to her moral self-identity; that is, without being aware of the fact that they are much more effective in motivating than they might expect, as opposed to other (moral and non-moral) priorities. They would not be fully "mature" moral agents, in Blasi's terms, since forming a moral self-identity requires reason, effort and time (Vigani 2016: 223; Blasi 1984: 138); however, this, I contend, is the case of both Neoptolemus and Huck Finn.

We have also seen that being integral with one's core commitments is experienced as a necessity; what is at stake is one's identity, thus were the agent to act contrary to such commitments, she would compromise not just her normative apparatus, but herself. Going back to Neoptolemus, it may indeed be possible for some other honorable heroes to be just a little dishonorable for the greater good, but he cannot; being honorable is so central to his moral self-identity that it provides him with the motivation to be always guided by his honor, he cannot just switch it on and off. Conversely, his motivation to trick Philoctetes is weak, because it involves being dishonorable; thus, it does not move him all the way to pursuing his decision. His moral self-identity is the cause of him ultimately acting well. In a similar fashion, we can see that Huck Finn is committed to friendship. He is even less aware than Neoptolemus of the centrality of such commitment, which is what makes his actions look even more mysterious; nonetheless, he is sincerely guided by his wish to care for his adventures companion, and by the gradual development of a sense of "fellowship and trust" (Levy 1964: 385). For this reason, the motivation to turn Jim in is weak; it originates in what they have taught him, not in his core commitments. Thus, he keeps failing to act accordingly. On the other hand, his motivation to save him over and over again originates in his deep-seated commitment to friendship, which is what ultimately explains his actions.

We have seen that those of Huck Finn and Neoptolemus are not instances of *akratic* behavior; that is, they do not show irrational practical reasoning. However, if we exclude that they are cases of inverse *akrasia*, we are left with no explanation as per why they ultimately act the way they do. I have here argued that their motivation to act well is to be found in their moral self-identities; that is, their action issues

<sup>&</sup>lt;sup>10</sup> For a comprehensive analysis of the impact of deontic and consequentialist judgments on behavior, see Bennett (1998).



<sup>9 &</sup>quot;ODYSSEUS: So now, for one short day, follow my lead without a sense of shame. In time to come they will call you the finest man there is" [100].

from a motivation that origins in their deep-seated commitments and volitions. I will end by suggesting that there is a more fitting label for cases such as those of Neoptolemus and Huck Finn, than that of "inverse *akrasia*", which is what Williams (1993) named "moral incapacity".

## 3 Moral Incapacity

So far I have argued that the cause of Neoptolemus' and Huck's actions is their motivation to act according to those deep-seated commitments that are central to their moral self-identities; thus, they are not cases of *akratic* behavior. We still have to make sense of the "cannot" that seems to be doing the work when it comes to the possibility of acting in accordance with their faulty moral knowledge. What is interesting here is the fact that, as explained by Blasi, there seems to be some sort of urge not to betray one's self, which is precisely the source of moral motivation, and it is not exclusively normative. Neoptolemus finds out that he *cannot* trick Philoctetes; not that he *shouldn't*, because, as we have seen, he would have had good reasons to do so. Similarly, Huck Finn cannot turn Jim in; he still thinks that he should, but there is something that prevents him from doing so. In what follows, I will argue that what they experience is what Williams (1993) calls a moral incapacity, where the "cannot" is substantial; it is neither a metaphor, nor a psychological impossibility. I shall first introduce Williams' own account and, then, apply it to the examples considered.

## 3.1 Moral Incapacities as Distinguished from Other Kinds of Incapacities

Moral incapacity refers to situations in which the agent may seriously consider an option, set herself to bring it about, but then, ultimately, find herself "incapable of acting in that direction" (Caprioglio Panizza 2021: 363; Williams 1993). This implies that we may not be aware of what it is impossible for us to do, until we find ourselves in the situation to consider acting as such. Crucially, this implies that moral incapacities are revealing of one's moral character (Williams 1993; Winch 1965); that is, it is not only a discovery of what you ultimately cannot bring yourself to do (a discovery about yourself), but also a discovery of what it is that is so fundamentally important to you that makes it impossible for you to act otherwise. It is a discovery of your moral self-identity; of your identity-giving commitments. Finally, it is a discovery you can account for; you can, to some extent, give reasons for this kind of incapacity (Williams 1993; Caprioglio Panizza 369). But in which sense is a moral incapacity "moral"? Could it not be just a psychological incapacity? And can this really be what happens in the alleged cases of inverse *akrasia* considered here?

Williams (1993) does distinguish between moral incapacities and other kinds of incapacities. Generally, from a 3rd person perspective, they are all alike; but they are significantly different in the way the agent comes to the conclusion that she cannot act. To begin with, he claims that moral incapacities are unique in their being "expressive of, or grounded in, the agent's character or personal dispositions"



(60); but this could still be true also for psychological impossibilities. Let us imagine Clara, an incurable arachnophobic; she is so afraid that if she sees a spider, she instantly freezes. We can say that it is part of her character that of being uncontrollably frozen in front of spiders and among her set of dispositions that of being easily startled by sudden crawly movements. Let us imagine that Clara is in a rush and she has to get to her scooter quickly; but, alas, there is a spider that is crouched right on her scooter keys. Now, she can try to convince herself that she has to move the spider away from the keys so that she can take them and go, she can even decide that she will try, but then, after taking the broomstick, she is likely to reach the conclusion that she cannot, after all, do it. The very possibility that the spider may crawl around after she tries to move it away from the keys is terrifying.

What is the difference between Clara's arachnophobia and Neoptolemus' commitment to honor, or Huck Finn's commitment to friendship? According to Williams, what he labels "a psychological incapacity" is very much like a physical incapacity (63); in a physical incapacity to A, under no condition will the world contain me A-ing. Let us employ his own examples: Rambo cannot physically lift 500kg and if ever there was the possibility for him to do so (say, he could under hypnosis), then it would have not been a physical impossibility in the first place. In the same way, if I cannot psychologically do A, under no condition will the world contain me A-ing; that is, according to Williams, the fact that I cannot work out in 10 seconds the product of two numbers of 5 digits each is like Rambo's not being able to physically lift 500kg. The world will never contain me doing that. Now, it is clear that Williams is working with a definition of "psychological" incapacity that is exclusively cognitive: if we apply this to the arachnophobia case, Clara not being able to remove the spider is like Rambo's not being able to lift 500kg, or like not being able to work out in 10 second the product of two numbers with 5 digits; she could not do it even if she tried to, and if she managed to do so through self-deception, then it is not true that it was impossible for her to remove the spider in the first place. This is odd; if Clara were to hypnotize herself into seeing a frog instead of a spider, she would have no problems removing it from the keys. The issue in the spider case is that she cannot remove it because she knows it is a spider, not because she has some sort of cognitive limit like in the case of the calculation. For this reason, I believe that we should distinguish between cognitive incapacities, such as the mathematics one, and psychological incapacities, such as the arachnophobia case<sup>11</sup>. Cognitive incapacities are much more obviously similar to physical incapacities, while psychological incapacities still look way too close to what Williams deems as moral incapacities. Let us see if we can make better sense of their difference.

There are some incapacities that Williams labels as "other" incapacities that he recognizes as being very much linked with moral incapacities, to the point that the difference is subtle, but crucial. What both other incapacities and moral incapacities have in common is the fact that the agent is unable to do A if she *knows* that she is A-ing (1993: 63). So far the arachnophobia case perfectly fits the description. He

Williams uses the example of vertigo (1993: 64) as one that is not "psychological" in the way he intends it, but as belonging to "other" kinds of impossibilities.



then adds that in the case of moral incapacity  $^{12}$ , as opposed to that of other incapacities, it is not true that I would fail to  $\phi$  even if I tried. Quite the contrary, if I tried I may well succeed; "[t]he moral incapacity is revealed in the fact that for the appropriate kinds of reasons, I will *never try*" (ibid., my italics). This is the difference. If we assume that phobias are action-impeding, then Clara will indeed not succeed in getting close to the spider, even if she tried. But this is fairly trivial; what is crucial from a moral point of view are, according to Williams, the *reasons* she has to stop trying, or to avoid trying at all. In the case of moral incapacity, something morally important is at stake; and that is why the agent concludes that she cannot  $\phi$ . Cases of moral incapacity are not cases of moral phobias, because in cases of moral incapacity,

[i]t is the process of deliberation that actually bears the weight [...], and it is at the centre of the moral incapacity. For the same reason, the fact that an act would be (in my view) disloyal or shabby is a consideration for *me* in deciding not to do it. The fact that an act would disgust me can be such a consideration. But where the act is so disgusting that I cannot do it, then the fact that it is disgusting does not function in that way: the question of deciding not to do it does not come up, or is cut off; or, if I do decide to try, the incapacity will make me fail. [...] In the case of moral incapacity, my deliberative conclusion not to do the act, reached on the basis of these totally decisive considerations, just is the conclusion that I cannot do it (64-65).

The reason behind Clara's failure to remove the spider is quite simply the fact that she is afraid of spiders, which has nothing to do with moral considerations. In fact, Clara is not committed to any moral judgment on spiders, she is just plain scared, and the fact that she cannot but freeze is not a conclusion she reaches every time she encounters a spider, it just happens. Let us now consider a moral case. Imagine that Clara promised her sister Frances to be present at a family dinner; Frances wants to talk about her career dreams, but since their parents can be extremely nasty when it comes to judging Frances' choices, she asks for Clara's psychological support. Clara then comes to know that, on the very same evening of the dinner, there is a concert all her friends are going to, which is surely going to be a lot of fun. She realizes she would much rather go to the concert than to that dinner, so she seriously contemplates the possibility to lie to her sister in order to go to the concert instead. Let us also imagine that, a bit like Huck Finn, Clara is not particularly susceptible to the requirements of morality; so, the fact that she promised something does not, in itself, constitute a reason for her to stick to it. She thus proceeds to elaborating the perfect excuse to miss the dinner, but when it comes to rehearsing it, she realizes that something important is at stake here – she is not the kind of person who leaves her sister alone, to be roasted by their parents. She does not conclude that she should not do it, but, rather, that she cannot do it.

It is not so clear what Williams means by "these totally decisive considerations"; at the end of the paper he claims that deliberation need not occur consciously, thus the moral incapacity may take the form of both a decision and a discovery (1993: 66). Clara may have thought of herself as perfectly capable of leaving her sister

<sup>&</sup>lt;sup>12</sup> That is, of "pure" moral incapacities (Williams 1993: 63). He recognizes that there may be mixed



alone, yet discover that, actually, she is not. She can go to the concert, but doing so would fall outside of her moral map; she sees it as a possibility, contemplates it, but then concludes that although there is a sense in which she can do it, she ultimately *cannot* do it<sup>13</sup>.

# 3.2 On the Difference Between "incapacity" and "impossibility"

Could Huck Finn and Neoptolemus have behaved differently? In a way, yes. They had the options to, in one case, hand Jim over and, in the other, go along with Odysseus' plan. Those options were physical – they had the physical ability to do so – but also practical; as a matter of fact, they seriously consider them. In Williams' terms, they could bring them about, but, for some morally relevant reasons, they will never try. It is in this sense that they can, in principle, act accordingly; but they come to the conclusion that, actually, they *cannot*. This is pretty evident in Neoptolemus' case, but we can see a similar pattern in Huck Finn. When the two slave hunters ask Huck what color is the skin of the man who is travelling with him, he automatically replies "white", even if he had been questioning what to do until a moment before. Not only that; he lies to the two hunters so that they would not double-check his response, which puts him in an extremely dangerous position; he knows what is at stake - his freedom and safety – yet he cannot but act to save his friend. In a way, he can: he has plenty of other options – but there is a sense in which, given his good heart, he cannot. In order to make better sense of this latter point, it may help to consider why these cases are instances of a moral incapacity and not of a moral *impossibility*.

Silvia Caprioglio Panizza has recently distinguished among three ways through which a moral impossibility can manifest itself (2021): (i) as instances of "the unconceivable", where an option is unconceivable to the agent insofar as it does not emerge as an option at all (2)<sup>14</sup>; (ii) as instances of "the unthinkable", where the agent may conceive of an option, but find it incomprehensible, not due to a cognitive deficit, but rather "due to the inability to make sense of something that appears beyond any category through which we understand and discuss morality" (2)<sup>15</sup>; and (iii), cases of what Williams (1993) labeled "moral incapacity", where the agent may seriously consider an option and set herself to bring it about, but then find herself incapable of acting accordingly (2).

I believe that (iii) is a substantially different phenomenon compared to (i) and (ii). In both "the unconceivable" and "the unthinkable" the element of the impossible is doing the work, while when it comes to moral incapacity, the crucial element is precisely the fact that, for the agent, acting in a morally dubious way is perfectly *possible*. This is why, as we have seen, Williams takes the "I can't" as a deliberative

<sup>&</sup>lt;sup>15</sup> See Frankfurt (1998) and Gaita (2004).



<sup>&</sup>lt;sup>13</sup> It is in this sense that "[a] moral incapacity [...] is one with which the agent is identified" (Williams 1993: 68). Of course one can have a moral incapacity with which does not identify anymore, but as soon as she set out to change it, it is already not a moral incapacity anymore, "not necessarily in the sense that he can now do the thing in question, but in the sense that if he cannot, it is no longer a moral incapacity, but rather one that is merely psychological" (1993: 69).

<sup>&</sup>lt;sup>14</sup> "I can only choose within the world I can see, in the moral sense of "see"" (Iris Murdoch 1970: 37).

conclusion, but one that is voluntary. A moral incapacity is an act of agency, while a moral impossibility is not. As extensively depicted by Caprioglio Panizza through the case of veganism (2020), the phenomenology of those who refuse to eat or participate in any way in animal exploitation is one of impossibility: there are multiple options, but the vegan agent does not consider them as actual possibilities, because they are either unconceivable or unthinkable for her. It is in this sense that the agent does not really have a choice. On the other hand, in cases of moral incapacity the agent does have a choice and also feels the alleged reasonableness of other possibilities, but she chooses to live up to the "cannot". This choice may be experienced through different forms: as an active endorsement of it, like Neoptolemus, or as an active surrender to it, as it seems more likely the case of Huck Finn. In all these declinations, agency is present, because the agent has in fact other possible options in front of her, but chooses, or discovers she deeply wishes to choose, to live up to the urge not to betray herself that is expressed in the "cannot". Cases of impossibility that have a moral relevance seem more close to moral phobias; that is, just like Clara the arachnophobic, the vegan agent avoids consuming animal products for the very reason that they are animal products – they appear un-eatable, un-exploitable ("affordances", Caprioglio Panizza 2022) and the thought of consuming them is either not present or experienced as disgusting, unacceptable. The line between moral and psychological reasons is incredibly blurred in moral impossibilities. Moral incapacities, on the other hand, are uniquely moral because the outcome is not due to a psychological incapacity but, rather, to the agent's active endorsement of her core moral commitments.

#### 4 Conclusion

The original question was: why do Neoptolemus and Huck Finn act the way they do, notwithstanding their original choices and faulty moral knowledge? Arpaly & Schroeder's (1999) answer is that they are inverse akratics; Holton's (manuscript) answer is that they are not inverse akratics because they are, rather, confabulators about their moral reasons. Neither of these answers are exhaustive in addressing the question. In this paper I argued with Holton that the cases in question should not be understood as showing akratic behavior (par 1) and, continuing where Holton left it, that they act the way they do due to their motivation to be integral with themselves (par 2); conversely, they fail to act in accordance with their original choice and faulty moral knowledge due to a lack of motivation to do so. Now, it is clear that the two cases are different. Neoptolemus knows that he is committed to honor; but he still sees the appeal in tricking Philoctetes in order to pursue the easiest path. Faced with his own actions, he realizes that he cannot pursue them further: honor is much more central to his moral self-identity than he originally believed. Tricking Philoctetes is not only wrong (he should not do it) but, first and foremost, would imply betraying himself; thus, he cannot do it (par 3). The fact that tricking Philoctetes would be wrong does not motivate him to avoid doing so, while feeling the threat of self-betrayal does. Thus, Neoptolemus' case is much more accurately understood as a case of moral incapacity, where the incapacity is grounded in his



moral self-identity. Huck Finn possesses faulty moral knowledge ("slavery is just and I am doing something wrong protecting Jim"), so he has all the "shoulds" in the wrong place; however, they are not motivating, because, first and foremost, he is a good friend and this is what motivates him. He is not aware of how much the value of friendship shapes his moral self-identity, nor he seems aware of knowing that he values friendship in the first place – he is, after all, an incredibly troubled boy who just wants to go on adventures with a trustworthy companion – but what ultimately motivates him to act is precisely this commitment to friendship, notwithstanding the appeal of the other options he has. Thus, also Huck Finn's case is much more accurately understood as a case of moral incapacity, where the incapacity is grounded in moral self-identity. These kinds of incapacities are not moral phobias; that is, agents are not disgusted or impulsively pushed away from other options. What it is that does the job is moral self-identity; the fact that notwithstanding the attractiveness of other courses of action, the agent decides that she cannot ultimately betray her moral self.

**Acknowledgements** I would like to thank Maria Silvia Vaccarezza and the Philosophy staff affiliated to the Northwestern Italian Philosophy Consortium (FINO) for comments on early versions of the paper.

Funding Open access funding provided by Università degli Studi di Genova within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a>.

#### References

Aristotle, Eudemian Ethics. Oxford's World Classics.

Aristotle, Etica Ncomachea. Economica Laterza.

Aristotle, Retorica. Biblioteca Filosofica Laterza.

Arpaly, N. 2000. On acting rationally against one's best judgment. *Ethics* 110 (3): 488–513. https://doi.org/10.1086/233321.

Arpaly, N., and T. Schroeder. 1999. Praise, blame and the whole self. Philosophical Studies 93: 161-188.

Bauer, K., V. Somogy, and M. Corinna, eds. 2017. *Dimensions of practical necessity "Here I Stand. I Can. do no Other.*" Basingstoke, UK: Palgrave McMillan.

Bennett, J. 1998. The act itself. Online. Oxford: Oxford Academic. https://doi.org/10.1093/019823791X.001. 0001.

Bergman, R. 2002. Why be moral? A conceptual model from developmental psychology. *Human Development* 45 (2): 104–124.

Blasi, A. 1984. Moral identity: Its role in moral functioning. In *Morality, moral behavior and moral development*, ed. W.M. Kurtines and J.J. Gewirtz, 128–139. New York: Wiley.

Blasi, A. 1983. Moral cognition and moral action: A theoretical perspective. *Developmental Review* 3: 178–121.

Blasi, A. 1999. Emotions and moral motivation. Journal for the Theory of Social Be-Haviour 29: 1–19.



Blasi, A. 2004. Neither personality nor cognition: An alternative approach to the nature of the self. In *Changing conceptions of psychological life*, ed. C. Lightfoot, C. Lalonde, and M. Chandler, 3–26. Mahwah, NJ: Erlbaum.

Blasi, A. 2005. Moral character: A psychological approach. In *Char-acter psychology and character education*, ed. D.K. Lapsley and F.C. Power, 18–35. Notre Dame, IN: University of Notre Dame Press.

Cagnoli Fiecconi, E. 2018. Aristotle on the structure of Akratic action. Phronesis 63 (3): 229–256.

Caprioglio Panizza, D. 2020. If veganism is not a choice: The moral psychology of possibilities in animal ethics. *Animals* 10 (145): 1–12. https://doi.org/10.3390/ani10010145.

Caprioglio Panizza, S. 2021. Forms of moral impossibility. *European Journal of Philosophy* 2021: 1–13. https://doi.org/10.1111/ejop.12651.

Caprioglio Panizza, S. 2022. The ethics of attention: Engaging the real with Iris Murdoch and Simone Weil. Routledge.

Darnell, C., L. Gulliford, K. Kristjánsson, and P. Paris. 2019. Phronesis and the knowledge-action gap in moral psychology and moral education: A new synthesis? *Human Development* 2019 (62): 101–129. https://doi.org/10.1159/000496136.

Davidson, D. 1980. How is weakness of the will possible? In Essays on actions and events, ed. D. Davidson, 21–43. Oxford: Oxford University Press.

Frankfurt, H.G. 1971. Freedom of the will and the concept of a person. *The Journal of Philosophy* 68 (1): 5–20. https://doi.org/10.2307/2024717.

Frankfurt, H.G. 1998. Rationality and the unthinkable. In *The importance of what we care about*. Cambridge: Cambridge University Press.

Gaita, R. 2004. Good and evil. London: Routledge.

Harmon-Jones, E., and C. Harmon-Jones. 2007. Cognitive dissonance theory after 50 years of development. *Zeitschrift Für Sozialpsychologie* 38 (1): 7–16.

Holton, R. Inverse akrasia and weakness of will. MIT, manuscript.

Hursthouse, R. 1999. On virtue ethics. New York: Oxford University Press.

Irwin, T. 1986. Aristotelian actions. Phronesis 31: 68-89.

Karniol, R., and D.T. Miller. 1983. Why not wait?: A cognitive model of self-imposed delay termination. Journal of Personality and Social Psychology 45 (4): 935–942. https://doi.org/10.1037/0022-3514.45.4. 935.

Kleist, C. 2009. Huck Finn the inverse akratic: Empathy and justice. Ethical Theory Moral Practice 12: 257–266.

Kristjánsson, K. 2013. Aristotelian motivational externalism. Philosophical Studies 2013 (164): 419–442. https://doi.org/10.1007/s11098-012-9863-1.

Lapsley, D.K. 2008. Moral self-identity as the aim of education. In *Handbook of moral and character education*, ed. P. Nucci. Larry and Darcia Narvaez. New York: Routledge.

Levy, L.B. 1964. Society and conscience in Huckleberry Finn. Nineteenth-Century Fiction 18 (4): 383–391.

Monroe, K. 1994. "But what else could I do?" Choice, identity and a cognitive–perceptual theory of ethical political behavior. *Political Psychology* 15: 201–226.

Murdoch, I. 1970. The sovereignty of good. London: Routledge.

Sofocle, Filottete. Salento Books.

Taylor, C. 1989. Sources of the self. Harvard University Press.

Vigani, D. 2016. Moral judgment and motivation: Making sense of mixed intuitions. *Ethical Perspectives* 23 (2): 209–230.

Wiggins, D. 1978. Weakness of will, commensurability, and the objects of deliberation and desire. *Proceedings of the Aristotelian Society* 79: 251–277.

Williams, B. 1993. Moral incapacity. Proceedings of the Aristotelian Society 93: 59–70. https://doi.org/10. 1093/aristotelian/93.1.59.

Winch, P. 1965. The universalizability of moral judgments. The Monist 49 (2): 196–214.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

