

Radoilska, L. Autonomy and Responsibility. Forthcoming (2022) in B. Colburn (ed.) The Routledge Handbook of Autonomy

AUTONOMY AND RESPONSIBILITY

Abstract

This chapter offers a fine-grained analysis of the relationship between autonomy and responsibility in order to address a challenge according to which considering autonomy and responsibility as closely related is misleading since these concepts serve different normative objectives. In response to this challenge, I first explore two criteria of ascription – rationality and control – that autonomy and responsibility seem to share. I then contrast and compare three pairs of autonomy and responsibility conceptions. Examining these pairs rescues the idea that there are normatively significant connections between autonomy and responsibility, albeit that what that connection is and why it matters is highly sensitive to the different understandings of autonomy and responsibility one might adopt. The first pair, self-governance and accountability, posits a notion of core agency as irreducibly valuable. The second, authenticity and attributability, rests on a shared ideal of actively becoming a distinctive self. The third and final, relational autonomy and answerability, derives from the thought that unequal standing impacts heavily on whether and how the criteria of rationality and control are applied in specific cases. This analysis demonstrates that rationality and control are not independent criteria but always work in tandem. Failing to appreciate these conceptual interactions could obfuscate promising pathways for both supporting personal autonomy and challenging unwarranted responsibility ascriptions.

Biographical note

Lubomira Radoilska is a Senior Lecturer in Philosophy, University of Kent, UK. She works on issues at the intersection of philosophy of action, ethics and epistemology. Radoilska is the author of *Addiction and Weakness of Will* (OUP, 2013) and editor of *Autonomy and Mental Disorder* (OUP, 2012).

1. Introduction

This chapter will aim to explore the possible interactions between autonomy and a closely related concept, responsibility. The discussion proceeds as follows. Section 2 looks into the rationale for considering autonomy and responsibility as necessarily related and then expands on a possible challenge. According to this challenge, the assumption of conceptual proximity is misleading since the criteria of rationality and control that autonomy and responsibility seem to share lead to differences with respect to scope and normative significance. In response to this challenge, Sections 3 to 5 contrast and compare three pairs of autonomy and responsibility conceptions. Examining these pairs will rescue the idea that there are normatively significant connections between autonomy and responsibility, albeit that what that connection is and why it matters is highly sensitive to the different understandings of autonomy and responsibility one might adopt. The first pair, self-governance and accountability, posits a notion of core agency as irreducibly valuable. The second, authenticity and attributability, rests on a shared ideal of actively becoming a distinctive self. The third and final, relational autonomy and answerability, derives from the thought that unequal standing impacts heavily on whether and how the criteria of rationality and control are applied in specific cases. The concluding Section 6 highlights some implications for the relationship between rationality and control.

2. Rationality and Control

The conceptual proximity between autonomy and responsibility is often assumed rather than explicitly asserted. This is partly a reflection of the overlap between central cases where both autonomy and responsibility are ascribed. Arguably, this overlap is more than a coincidence: in each case, the ascription criteria at play with respect track aspects – albeit different – of the same feature, acting for reasons of one’s own. So, when we ask whether a person, a choice or an action are autonomous and/or responsible, we are trying to discern the extent to which they display capacities, such as rationality, the ‘reasons’ aspect, but also control, the ‘ownership’ aspect of acting for reasons of one’s own (McAninch 2017; Radoilska 2012).

There are some notable exceptions from this trend: Arpaly (2002), Fisher (2012) and Oshana (2002). All three authors challenge the conceptual proximity picture and argue that autonomy and responsibility follow separate logical paths. If we fail to appreciate that, we end up with unhelpful accounts of both. As Arpaly (2002: 127) observes about approaches that define responsibility in terms of autonomy:

Whenever we ask, “Is so-and-so autonomous?” instead of “Is so-and-so morally responsible?” we expose ourselves to a tendency to equivocate ... or at the very least to be unduly influenced by intuitions about irrelevant senses of ‘autonomy’. Thus, phrasing questions in terms of the agent's autonomy instead of in terms of her moral responsibility serves not to clarify, but to further obscure, the already difficult question of her moral responsibility.

According to Oshana (2002: 279), the opposite move, positing that responsibility entails autonomy, shares the same bleak prospects:

Responsible agency demands a capacity for social participation and involvement, but insofar as the submissive person is not socially and epistemically incapacitated (by his own psychology or by the activities of those he follows) we can continue to consider him a competent subject of moral address, even though we cannot describe him as autonomous...

In a similar vein, Fisher (2012) proposes a distinction between two target concepts of agency that might be at work when we try to define freedom of action. The first is responsibility-identification and the second, autonomy-identification. According to Fisher, this distinction has been ignored by both proponents and critics of the thesis that identification with one's own motives is the defining feature of free actions. Both sides have failed to appreciate that the contested feature is fundamentally underdescribed. It can serve a more capacious agency concept, such as responsibility that includes unreflective and weak-willed behaviours. Alternatively, it can ground a more demanding agency concept, such as autonomy that excludes such behaviours.

Importantly, all three authors draw a contrast between responsibility and autonomy by exploring the different ways in which they engage the capacities at the heart of the conceptual proximity picture, rationality and control. This contrast is least pronounced on Fisher's account: autonomy is more demanding than responsibility because it sets out a higher standard of rationality. To meet this standard, an agent should act in a self-governed way, that is out of a true or real self rather than merely on reasons or motives of their own. By contrast, responsibility is consistent with failures of rationality which, in this context, are closely associated with failures of self-control. As Fisher (2012: 174) puts it:

When one exhibits weakness of the will, one acts freely and is morally responsible; indeed, one is criticisable irrational for the weak-willed behaviour. Thus, an agent's being morally responsible for a bit of behaviour is perfectly consistent with an agent's falling short of certain norms of rationality (as well as morality) in so behaving.

Yet, the contrast need not be interpreted as one of degree. Nor should we expect that rationality and control always pull in the same direction. To see how this might work, let us briefly consider Oshana (2002) and Arpaly (2002).

Oshana (2002) articulates two separate conceptions of rationality. The first, instrumental conception refers to the set of abilities required to fulfil, revise or abandon one's preferences. It turns on whether a person is capable to avoid inconsistencies within and across their plans, and to identify and pursue appropriate means for realising them. The second conception, however, requires normative competence with respect to moral reasons. A rational person in this sense appreciates moral considerations as authoritative reasons for action. These two conceptions are different in kind: an instrumentally rational person may not be at the same time normatively competent; and vice versa, responsiveness to moral reasons may be enacted in the absence of sound means-to-end reasoning. Since autonomy is interpreted primarily in terms of self-determination, self-directedness and self-management over the course of one's life rather than control over specific decisions and choices (Oshana 2002: 273-77), it requires only instrumental rationality. This is not the case with responsibility. As indicated by the passage quoted earlier, it rests entirely on normative competence, modestly understood. That is to say, acting for morally wrong reasons is consistent with such a competence to the extent that the agent can, in principle, understand that these reasons are morally wrong.

What's more, autonomy and responsibility turn out to align with two independent notions of control. While autonomy requires diachronic self-control in terms of comprehensive self-governance along the three dimensions listed above, responsibility only implies the so-called guidance control over particular actions (Fisher and Ravizza 1998). And as we saw, guidance control can be satisfied even by weak-willed actions, which demonstrate loss of self-control. Such actions are still reasons-responsive in the required, modest sense.

Arpaly (2002) widens further the gap between self-control and responsiveness to reasons. Whether a person acts against their better judgment is deemed as irrelevant to questions of rationality and moral responsibility. For weak-willed agents might be tracking reliably moral reasons without realising it. Like Huckleberry Finn who finds himself unable to turn in his new friend Jim even though he firmly believes that this is his duty, such agents are morally praiseworthy and by no means irrational. Moreover, the lack of self-control displayed in these instances of 'inverse akrasia', as the phenomenon has come to be known, is consistent with authenticity. This is because when a person acts against their better judgement, this action may represent what they truly care about better than the judgement they forego. Reflective endorsement is not a reliable guide to either responsiveness to moral reasons or authenticity. Such a person, however, would not be self-governing. As Arpaly (2002: 122) puts it: 'they do not control themselves..., on the contrary, their selves control them'.

To recap, a closer look at the roles played by rationality and control in the ascriptions of autonomy and responsibility, respectively, reveals that the latter two concepts may have different rationales and serve different normative agendas. This upshot, however, is only a first step toward a comprehensive analysis. Arguably, both autonomy and responsibility are essentially contested concepts (Gallie 1956), each with some competing conceptions that are

neither reducible to a more fundamental one, nor easy to reconcile. A contrast in terms of rationality and control cannot do justice to the complex interactions across such conceptions as it tends to paint our target concepts as more homogenous and unified than they can plausibly be. A finer-grained analysis is required to address this task. The following three sections will expand on this thought by contrasting three pairs of cognate conceptions of autonomy and responsibility in turn. The analysis builds on two distinctions. The first is between three families of autonomy conceptions explored in Radoilska (2012): value-neutral (autonomy as self-governance), substantive (autonomy as authenticity) and relational (autonomy as equal personal standing within a community). The second is between three kinds of responsibility put forward in Shoemaker (2015): accountability, attributability and answerability.

3. From Core Agency to Self-Governance and Accountability

Both self-governance and accountability rely heavily on intuitions about what we may term core agency. This is because it helps identify some formal structures and processes as opposed to the substance of choices, actions or even whole lives to which questions of autonomy and responsibility may apply. More specifically, the fundamental feature is a distinctive kind of activity (as opposed to passivity) at the heart of agency. When activity in this sense takes place, its exercises are deemed as both autonomous, that is self-governed, and things for which an agent may be aptly held to account. So, with respect to scope, self-governance and accountability are not expected to come apart.

This becomes apparent in McAninch (2017) where the notion of agential activity is deployed to elucidate why any conduct resulting from it is both self-determined and responsible. The neat parallel between the two ascriptions is underpinned by a characteristic called 'capacity

for normative avowal': being agentially active commits us to either avow our motives as reason-giving or, conversely, to disavow them as inadequate.

The links between self-governance and accountability are even tighter in Colburn (2010). According to Colburn, 'Autonomy consists in deciding for oneself what is valuable, and living one's life in accordance with that decision' (2010: 19). This requires voluntary choice, that is not non-voluntary choice made in a context where acceptable alternatives are available. Voluntary choice, however, is also ground for responsibility as accountability: If I choose voluntary in the sense specified above, I thereby assume substantive responsibility for the consequences of my choices. Here, both autonomy and responsibility derive directly from the same core of agency: voluntary choice. They look like two sides of the same coin.

Nevertheless, questions about autonomy concern primarily longer periods. This is because the perspective of self-governance – including responses to factors that may undermine it – is that of an agent, whose interests transcend the realisation of each of their individual choices. In contrast, questions about responsibility target specific actions. The perspective from which they arise is that of a moral community that perceives these actions as problematic in some way. Being ready to account for such effects of one's voluntary choices can also be traced back to self-governance; however, the core agency feature is now locally ascribed to particular actions rather than globally to a life as a whole.

These observations partly address the concern about possible difference in scope raised in Oshana (2002): self-governance and accountability are indeed coextensive; however, ascribing them appears to have a different point. Hence, core agency is assessed globally in the case of self-governance and yet locally in the case of accountability.

Following on this thought, it is worth exploring if the ways in which core agency supports ascriptions of autonomy as self-governance may also vary from the ways in which it supports ascriptions of accountability. To see if this is so, let us look into a couple of relevant accounts on both sides of our contrast-and-compare analysis.

Starting with self-governance, its significance within value-neutral accounts, such as Radoilska (2012) can be defined as that of an independent source of normativity. That is to say, there is strong (though defeasible) assumption that one's choices and actions deemed as autonomous in this sense ought to be protected from interference by others to the extent that they do not interfere with the autonomous choices and actions of others. Here, core agency or active self-determination underpins the normative significance of having some discretionary sphere where our freedom to make our own mistakes may not be challenged. This sphere is meant to give substance to the close conceptual connection between two central dimensions of autonomy highlighted in Feinberg's classical analysis (1986: Ch. 18), a capacity for self-determination safeguarded by a right to self-determination.

Taken to its limits, the idea of self-governance as independent source of normativity would serve to delineate a range of choices and actions for which an agent does not have to account at all, in so far as they are self-regarding. In this respect, self-governance breaks away from the initial rationality requirement that seemed to motivate both autonomy and responsibility ascriptions. A UK court ruling clarifying the scope of a patient's right to refuse treatment, a paradigm case of self-regarding choice, illustrates well this move away from rationality: "A mentally competent patient has an absolute right to refuse to consent to medical treatment for any reason, rational or irrational, or for no reason at all, even where that decision may lead to his or her own death" (Re MB 1997). As the ruling stipulates, on such occasions, the

capacity for self-determination does not have to be exercised in a rational way. Nor does it require backing by reasons that the agent should avow or disavow when challenged by others.

It might be objected that the breakaway from rationality is not as clean as presented. For mental competence includes rational abilities, such as understanding information and making a decision in its light. This is a fair point; however, it does not bridge the gulf between the initial standard of rationality meant to exclude weak-willed actions as non-autonomous by default and rationality as a background condition for self-governance that may result in unreasonable as well as irrational choices. More importantly, whether rational or irrational, self-governed and self-regarding choices are understood as fundamentally discretionary since an agent is not liable to account to anyone for them.

Conceptions of accountability based on core agency take a different path. While activity as opposed to passivity is still the central normative feature as in theories of autonomy as self-governance, it is interpreted in a way that helps revisit the initial control requirement but not the rationality one. The underlying reasoning is clearly articulated in Smith (2005: 251):

When we praise or criticize someone for an attitude it seems we are responding to something about the content of that attitude and not to facts about its origin in a person's prior voluntary choices, or to facts about its susceptibility to influence through a person's future voluntary choices. More specifically, it seems we are responding to certain judgments of the person which we take to be implicit in that attitude, judgments for which we consider her to be directly morally answerable.

On this view, some rational connection to a person's evaluative judgments is the mark of core agency, not voluntary choice. Thus, the activity which grounds accountability is expressed in the reasons that make an action or attitude morally relevant. Importantly, these reasons do

not have to be apparent to, let alone avowed or endorsed by the agent who is held to account. Forgetting a friend's birthday, to use one of Smith's own examples, is not something one can do for a reason. Yet, it is something for which a person is aptly held accountable when it indicates that they do not consider their friend's feelings as important enough. There is a prima facie rational connection between an offending evaluative judgment – even though in all likelihood it has never been explicitly made by the agent – and the unwitting conduct for which they are called to account.

On other accountability conceptions, however, rationality is not interpreted as alternative to control. For instance, Portmore (2019) argues that cases, such as being held responsible for forgetting a friend's birthday clearly indicate the kind of control involved in what we termed core agency. Looking at attitudes for which accountability is apt, say, taking for granted somebody's friendship, it becomes apparent that the relevant feature they have in common with standard examples, such as voluntary choices and actions is to be 'receptive and reactive to reasons' (2019: 26). According to Portmore, this feature demonstrates why we should not be looking for any other control than the one already embedded in the standard of rationality applicable to attitudes: they can be formed, sustained or abandoned in light of new considerations that come to the agent's attention, such as the fitting resentment they have provoked in others. That such corrective exercises of agency may not be under our direct voluntary control is beside the point. They still can be shaped in the light of reasons even though this might require greater effort over time.

To recap, both self-governance and accountability rest on the same notion of core agency and so remain coextensive albeit that the former is typically ascribed globally and the latter, locally. Moreover, core agency supports these ascriptions in a different way. With respect to

self-governance, it leads to identifying a category of actions for which an agent is not liable to account precisely because they are autonomous, in addition to being self-regarding. This is the upshot of seeing core agency through the lens of control and contracting rationality to a threshold capacity that need not be exercised on any occasion. With respect to accountability, in contrast, core agency is conceived as responsiveness to reasons that agents might exercise unbeknown to them. This form of rationality is sometimes interpreted as superseding control; alternatively, it is used for identifying more clearly the relevant kind of control.

4. From Selfhood to Authenticity and Attributability

Neither autonomy, nor responsibility are only ascribed to exercises of core agency, such as choices, actions or attitudes. They can also be ascribed directly to agents as selves. Conceptions focussing on the kind of person that may be deemed as autonomous or the character traits for which a person may be held responsible both face an apparent tension between activity and passivity. On the one hand, personal identity is a given, over which the agent has little direct control. On the other, ascriptions of autonomy and responsibility are fundamentally communicative, they are forms of moral address aiming to acknowledge and engage this agent's deep self as the ultimate source of relevant motivations.

Some philosophers urge us to acquiesce to this tension and abandon the underlying assumption that only active self-determination can ground both personal autonomy and moral responsibility. For instance, Buss (2012) draws attention to the normative significance of the distinction between health and sickness with respect to extreme emotions. According to Buss, this distinction does not track fuller or more reflective exercises of agency as opposed to lesser ones. Instead, it reveals a passive, non-agential and purely causal role that the agent

plays in the formation of their own intentions when the ensuing actions are autonomous and/or responsible. As Buss (2012: 691) puts it: 'To govern ourselves, we must express our identity as nonagents. The self-determination that distinguishes us as autonomous agents is self-determination in the passive mode'.

Others, however, investigate this tension in order to separate out apt from inept ascriptions of autonomy to and responsibility for one's self. Katsafanas (2016) is a case in point. On his account, autonomy is a capacity to direct one's actions in the light of self-given principles and values. It includes a distinctive self-understanding: to be autonomous, agents have to see themselves as choosing and acting on good reasons. Character, by contrast, is understood as a set of unchosen dispositions that incline a person to act in a certain way. Whenever my character determines the outcome of a choice of mine independently of what I would see as a good reason, I am not autonomous in virtue of my own character's influence. This becomes apparent in cases where a person realises the insidious causal role played by dispositions of theirs, such as irascibility. Once such a discovery is made, the agent can no longer see themselves as acting on good reasons.

Not all character influences need diminish autonomy in this way. As Katsafanas (2016) points out, acting in character may not affect self-determination any more than ordinary circumstantial factors. However, appreciating the normative interest of autonomous selves as opposed to autonomous exercises of agency requires a further dialectical step. For the ambition is to uncover the appeal of autonomy as an ideal of authenticity, a relationship to oneself that at the very least excludes such experiences as loss of intelligibility and alienation (Moran 2001). This ideal goes beyond remaining true to oneself, the conception of authenticity at work in Arpaly (2002). As we saw earlier, being true to oneself does not

preclude experiencing one's own motives as unintelligible. In contrast, a commitment to shaping one's own character by engaging in specific practices and cultivating related dispositions offers a more promising ground. This can be fleshed out in terms of character planning that aims to remove as much as possible any covert influences that affect us precisely in virtue of our being unaware of them (Colburn 2011). Alternatively, it can be devised as a character-building exercise meant to diminish and ultimately outgrow what an agent perceives as akratic distractions (Radoilska 2012). In either case, an authentic self will not be autonomous in the sense of being effortlessly forged by our voluntary choices.

Likewise, responsibility as attributability goes beyond voluntary control since its focal point is a person's deep self. As Shoemaker (2015: 48) stipulates: 'To be an attributable agent is to be worthy of agential admiration/ disdain for attitudes (volitional and non-volitional) in virtue of their expressing one's deep self'. The underlying motivation for this kind of responsibility ascriptions stems from a deep-seated interest in who our fellow agents are morally speaking. Are they considerate, trustworthy, courageous or kind? Conversely: Are they selfish, conniving, reckless or fickle? On their own, accountability ascriptions would not speak to that interest. For, we may still wish to know what kind of person an agent is even when we do not – or no longer – care to hold them to account for anything they do.

These observations help clarify why attributability may not be subject to the same constraints as accountability. They do not, however, address but merely highlight the tension between activity and passivity at the heart of responsibility for character. This challenge is taken up by Strabbing (2016) who argues that the underlying puzzle can be solved by looking into attributability for weakness of will. The negative moral appraisal that weakness of will seems to attract in terms of attributability goes against the grain of standard ascriptions where some

virtue or vice gets attributed to an agent because of their acting and feeling in character, that is in accordance with their own practical identity. Yet, weakness of will or acting and feeling against one's better judgment occurs when an agent contradicts core aspects of their practical identity. So, on the face of it, weakness of will should not be attributable since it does not display the capacity to act in accordance with one's practical identity. If accepted, this conclusion would apply to any character traits that deviate from an agent's self-understanding. This in turn would both reduce the scope for attributability and undermine its distinctive rationale.

On Strabbing's view, this unwelcome upshot can be avoided as it derives from an unduly restrictive interpretation of the relevant capacity, acting in the light of one's own practical identity. This capacity does not have to be exercised to mark out dispositions that are rightly attributable to an agent. Suffice that the agent has that capacity in reserve since not exercising it may be just as telling of their character. Returning to Strabbing's test case, weakness of will, we can see that the akratic preference is also valued albeit it less than the better, forgone alternative. This evaluative stance is attributable to the agent: it is their own even though it conflicts with the practical identity they endorse.

To recap, authenticity and attributability face an implicit tension between activity and passivity due to the notion of deep self which is central to both. For instance, having a character can enhance autonomy in one respect while undermining it in another. To resolve this tension, some conceptions of authenticity have moved away from the ideal of being true to oneself to endorse that of character building. Similarly, attributability looks either unwarranted or superficial kind of responsibility if limited only to acting and feeling in character. Broadening its scope to include conflicted and even disavowed but nevertheless

recurrent actions and attitudes links attributability to an agent's own evaluative stance. Thus, responsibility attaches to a distinct rational activity albeit such that it often falls beyond this agent's direct control.

5. From Equal Standing to Relational Autonomy and Answerability

As argued earlier, self-governance and accountability track a distinctive relationship to one's choices and actions whereas authenticity and attributability are aimed at a distinctive relationship with oneself as a character. The third and final pair of autonomy and responsibility conceptions concern the set of relationships an agent has with others. The underlying approach is well-established in the autonomy literature where the constitutive role of equal standing has long been acknowledged (Mackenzie and Stoljar 2000; Oshana 2006). Both self-governance and authenticity can be significantly enhanced or, conversely, impeded as a result of ongoing interactions with meaningful or powerful others. Moreover, seeing oneself as a source of projects worth sticking to in the face of adversity is not something agents can achieve by themselves. Thus, an individual's self-understanding required for authenticity and self-governance rests on social foundations.

This argument has been taken further to suggest that relational autonomy captures fully the target concept; therefore, less helpful, if not muddled conceptions, such as authenticity and self-governance should be discarded. For instance, Garnett (2014) articulates a purely social or relational view premised on one relevant feature only: the absence of rule from another. The three constraints – legal status, self-worth and critical reflection – that jointly determine whether this feature is present within a person's life all track their capacity to resist attempts at domination by others within intersecting social institutions. In Garnett (2017), this

approach to autonomy is applied to revisit the assumptions that underpin such notions as self-government and authenticity. According to Garnett, the only warranted interpretation is negative: not being subject to another's will. Philosophers who try to build a more substantive interpretation of autonomy do so by extrapolating what a person's relationship with themselves should look like from the interpersonal case. This prompts a category mistake: self-government denotes a relationship between a thing and itself rather than between that thing and *its* self (2017: 14). The latter, illicit move leads to positing an authentic self and then misinterpreting inner constraints to self-determination on the model of external constraints, such as domination, deceit or coercion.

Reducing authenticity and self-governance to relational autonomy is not however the only dialectical option once we acknowledge the significance of social standing. A promising alternative emerges from recent work on how external constraints to agency might translate into inner impediments. Thus, Kennet and Wolfendale (2019) identify and explore the effects of deprivation and discrimination on both a person's agency and the social ascriptions of responsibility it incurs. Self-control is at stake in both cases. In the former, the lack of what Kennet and Wolfendale term 'moral security' affects two core dimensions of self-control, long-term perspective and sense of one's own agency. Morally secure agents enjoy social recognition: overall, they are treated with respect and their personal projects, acknowledged as worth pursuing. Such agents have the right external conditions to constitute themselves as 'autonomous, self-governing individuals' by exercising self-control (2019: 38). In contrast, self-control is unlikely to pay off for morally insecure agents since their desired outcomes are not reliably connected to what they do: no matter how hard a person living on the breadline tries to save up, they will not lift themselves out of poverty. Moreover, seeing one's own efforts as futile with respect to significant projects over time would shift one's sense of

agential efficacy to what's doable in the short term. So, in the absence of moral security, agents are likely to end up with a shrunk scope for self-control. This impacts on their self-governance and authenticity since the projects they get stuck with are often not of their choosing.

The loss of autonomy across all three dimensions – equal standing, action and character – is paralleled by widespread negative moral appraisal. For the effects of deprivation and discrimination on the very rationality of self-control are ignored. Instead, morally insecure agents are unfairly held responsible for what looks to outsiders like poor exercise of self-control. The latter is misinterpreted as the main cause for the predicament in which these agents find themselves. Kennet and Wolfendale (2019) see this upshot as a major argument against conceptions which embrace the irreducibly social nature of responsibility ascriptions. For such cases as moral insecurity reveal that these ascriptions may perpetuate relationships of oppression. A theory of responsibility that does not take the social practices of holding responsible as its starting point would stand a better chance to correct and resist their shortcomings.

The conception of responsibility at issue here is answerability. It points to a communicative exchange where individuals or groups are called to offer reasons for various exercises of their agency that have been singled out as morally significant by others. In standard cases, that significance is *prima facie* negative. For answerability is understood, in its inception at least, as a response to a conversational challenge, such as 'Why did you do it?' (Hieronymi 2004; Macnamara 2015). The underlying assumption is that all participants in the answerability exchange are of equal moral standing. Consequently, agents become answerable for apparently undermining the equal standing of others.

Still, Kennet and Wolfendale's challenge cannot be readily dismissed. Equal standing might remain unfulfilled desideratum in answerability as a social practice as opposed to normative theory. As their study on the effects of deprivation and discrimination shows, the conversational challenge might be picking on the same groups and individuals. Whether intended or not, relentless moral scrutiny of this kind would eventually undermine the equal standing of the affected agents adding to the social stigma that they are already face.

This challenge, however, can be met by expanding the social core of answerability rather than giving up on it. For instance, a process-oriented conception according to which holding someone responsible is itself a morally relevant action can address the answerability gap we saw opening earlier (Radoilska 2021). This is because on this conception, the relevant communicative exchange is no longer interpreted as a response to a one-off conversational challenge but an ongoing process where the challenge itself is subject to moral attention in the same way as the exercises of agency on which this challenge picks up.

6. Concluding remarks

The preceding discussion offered a fine-grained analysis of the relationship between autonomy and responsibility. The three pairs of cognate conceptions that we explored showed that, once specified, autonomy and responsibility ascriptions are intimately connected. In each dimension we considered – core agency, selfhood and equal standing – questions of autonomy and responsibility turned out to be mutually illuminating. For although counterpart ascriptions of autonomy and responsibility are, overall, coextensive, they serve different normative objectives. The contrast is especially striking with respect to core agency. Ascriptions of autonomy as self-governance are meant to exclude a set of significant choices and actions from the scope of accountability, the cognate responsibility conception. The

underlying dynamic is reflected in the relationship between rationality and control, the two criteria that autonomy and responsibility have in common. As we saw earlier, these criteria are not independent but always work in tandem. For instance, a notion of control takes the lead in defining autonomy as authenticity supported by rationality as the capacity to avow or disavow one's motives as good reasons for action. The opposite move takes place in defining its counterpart, responsibility as attributability: rationality as responsiveness to reasons also determines the relevant kind of agential control. Looking into relational autonomy and answerability, it becomes apparent that at least two notions of rationality and control are at play on each side. For instance, when thinking about the more demanding sense of rationality as what one ought to do on pain of irrationality, it becomes uncharitable to say that agents operating under oppression ought to compensate for it by exercising greater self-control. This uncharitable upshot is avoided on a more permissive sense of rationality as what makes sense to do or can be worth undertaking. It makes sense to enhance self-control under oppression. There is however no contradictory relationship with the alternative: not enhancing self-control to compensate for oppression also makes a lot of sense. Failing to appreciate these conceptual interactions could obfuscate promising pathways for both supporting relational autonomy and reorienting answerability expectations to challenge oppression.

References

Arpaly, Nomy (2002) "Varieties of Autonomy," In N. Arpaly, *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press.

Buss, Sarah (2012) "Autonomous Action: Self-Determination in the Passive Mode," *Ethics* 122 (4):647–691.

- Colburn, Ben (2010) *Autonomy and Liberalism*. New York & London: Routledge
- Colburn, Ben (2011) "Autonomy and Adaptive Preferences," *Utilitas* 23(1): 52–71
- Feinberg, Joel (1986) *The Moral Limits of the Criminal Law: Harm to Self*. Volume 3. New York: Oxford University Press.
- Fisher, John Martin (2012) "Responsibility and Autonomy: The Problem of Mission Creep," *Philosophical Issues* 22 (1):165-184.
- Fischer, John Martin and Mark Ravizza (1998) *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Gallie, W.B. (1956) "Essentially Contested Concepts," *Proceedings of the Aristotelian Society* 56: 167–98.
- Garnett, Michael (2014) "The Autonomous Life: A Pure Social View," *Australasian Journal of Philosophy* 92 (1):143-158
- Garnett, Michael (2017) "Agency and Inner Freedom," *Noûs* 51(1):3-23
- Hieronymi, Pamela (2004) "The Force and Fairness of Blame," *Philosophical Explorations* 18: 115-148
- Kennett, Jeanette and Jessica Wolfendale (2019) "Self-Control and Moral Security," In David Shoemaker (ed.) *Oxford Studies in Agency and Responsibility*. Volume 6. Oxford: Oxford University Press.
- Katsafanas, Paul (2016) "Autonomy, Character, and Self-Understanding," In Iskra Fileva (ed.) *Questions of Character*. New York: Oxford University Press.

Mackenzie Catriona and Natalie Stoljar (eds.) (2000) *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*. New York: Oxford University Press.

Macnamara, Coleen (2015) "Reactive Attitudes as Communicative Entities," *Philosophy & Phenomenological Research* 90: 546-569.

McAninch, Andrew (2017) "Activity, Passivity, and Normative Avowal," *Pacific Philosophical Quarterly* 98 (1): 2-24.

Moran, Richard (2001) *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.

Oshana, Marina (2002) "The Misguided Marriage of Responsibility and Autonomy," *The Journal of Ethics* 6(3): 261–280.

Oshana, Marina (2006) *Personal Autonomy in Society*. Aldershot: Ashgate.

Portmore, Douglas W. (2019) "Control, Attitudes, and Accountability," In David Shoemaker (ed.) *Oxford Studies in Agency and Responsibility*. Volume 6. Oxford: Oxford University Press.

Radoilska, Lubomira (2012) "Autonomy and Ulysses Arrangements," In L. Radoilska (ed.) *Autonomy and Mental Disorder*. Oxford: Oxford University Press.

Radoilska, Lubomira (2021) "Circumstance, Answerability and Luck," *The Monist* 104(2): 155–167.

Re MB [1997] 2 FLR 426. <http://www.bailii.org/ew/cases/EWCA/Civ/1997/3093.html>

Shoemaker, David (2015) *Responsibility from the Margins*. Oxford: Oxford University Press.

Smith, Angela (2005) "Responsibility for Attitudes: Activity and Passivity in Mental Life," *Ethics* 115: 236–271.

Strabbing, Jada (2016) "Attributability, Weakness of Will and the Importance of Just Having the Capacity," *Philosophical Studies* 173(2): 289-307