

**Christine Straehle, University of Ottawa and University of Hamburg**  
**christine.straehle@uottawa.ca**  
**Vulnerability, Autonomy and the Limits of Epistemology**

Vulnerability has often been described as an inherently human characteristic. In this ontological sense, humans are vulnerable due to the embodied nature of human lives. A second way of conceptualizing vulnerability has proposed it as a constraint to individual autonomy. The argument then is that those who are vulnerable are not able to take decisions about the course of their lives since their decision-making capacity is hampered by dependency on the acts of another. Accordingly, morally problematic vulnerability arises in interaction with another. This is possibly particularly problematic for accounts of relational autonomy.

In this paper, I explore the concept of vulnerability with these two uses in mind. I suggest that a neglected aspect of individual vulnerability is the limits it may pose for epistemological capacity. In particular, I am interested in how specific circumstances of individual vulnerability may challenge the role individual introspection can play in leading autonomous lives. I want to argue that individual vulnerability can jeopardize the role introspection is usually assumed to play in deciding what course we hope to give our lives. In some cases of vulnerability, individuals may not be able to endorse and take responsibility for choices they make due to the fact that they don't trust their introspective knowledge. In this way, vulnerable individuals lose the authoritative grip on their introspective self-knowledge. Yet self-knowledge is a vital condition of individual autonomy: without self-knowledge, individuals lack the internal requirements of personal autonomy.

#### INTROSPECTION

I understand introspection as the capacity to access our intentional states and to reason about them. Without introspection, it is very difficult if not impossible to know one's wishes and hopes. I should note at the outset that I am not interested in the question of whether or not there is such a thing as introspection, or whether or not, as Hume famously put it, that whenever he introspected, he only came across a range of perceptions. Instead, I will discuss introspection as an epistemological method, as a way to know and to learn about ourselves. In a first instance, a few comments are warranted about what kind of knowledge introspection can yield.

Introspection the way I construe it here serves at least two different functions: first, it helps us access our intentional states; our thoughts, wishes and desires. In this vein, introspection is a vehicle to self-knowledge, understood as "inwardly directed attention" (Goldman 2006, 246). Questions that are asked in this context are: what do I feel? What do I think about x? how does x affect me?

Knowledge arrived at through introspection is *exclusive* to the person who engages in introspection: "Introspection is the capacity to inspect the, metaphorically speaking, 'inside' of one's mind. Through introspection, one knows what mental states one is in: whether one

is thirsty, tired, excited, or depressed.” (Steup, 2018) It is then often characterized “as the operation of a mental self-scanning or self-monitoring process” (Schwitzgiebel 2012, 30).

Many philosophers think that introspection is a *privileged* way of accessing our intentional states, thus generating authoritative self-knowledge since “no one other than oneself can be the subject of one’s intentional states” (Macdonald 2007, 369). Introspection is then taken to be a more immediate way of accessing knowledge about ourselves, less error prone than other modes of accessing knowledge, such as perception since *prima facie*, it is not clear how we can discount my knowledge that I have a headache. In contrast, we can dispute somebody’s perception of knowledge in the world. Introspective knowledge thus seems to be more secure, “epistemologically different from, and better than, other ways of knowing about the world.” (Smithies 2012, 260).

Later on, I will discuss to what extent the privileged and authoritative understanding of introspective knowledge may be challenged by certain situations of vulnerability. All I want to say here is that introspection is necessary to have access to our mind, and to the functioning of our mind. Introspection allows us to answer the question of what we want, think and feel, but also to reflect and deliberate on our desires, thoughts and feelings. This is to say that through introspection, we can access not only our first-order thoughts about our intentions, but also our second-order thoughts. Introspection helps answer the question of *what* we want, wish and hope, and also *why*. In this sense, through “endorsement, rejection or revision” introspection allows to appropriate some intentions as *ours*, as expressive of ourselves.<sup>1</sup>

To be sure, introspection is not the only way to acquire self-knowledge. We could imagine that we achieve knowledge about ourselves through discussion with others, who tell us what they think of us, how they perceive us and our actions in the world. Some philosophers doubt the idea of epistemic autonomy that my take on introspection may suggest. Elizabeth Anscombe, for instance, questions introspection as a valuable epistemological method since “it is a rather doubtful one, as it may consist rather in the

---

<sup>1</sup> “Situations in which a subject focuses on her thoughts, even introspectively, are not merely ones in which she simply becomes aware of her thoughts by noticing them.[...] a subject may be actively reflecting on the content of her first-order thought in order to evaluate that thought: deliberation seems to be one such case. Normally subjects engage in this process in order to make up their minds about what they are to think or believe. The ability to engage in it requires being able to reflectively attend, in current, conscious thought, to one’s own thought contents, and to know what those contents are, in order to accept, reject, or adjust those contents and effect rational transitions between them. This critical stance involves thinking second-order thoughts about one’s first-order ones, presenting those thoughts as present ones and deploying contents that present the contents of the first-order ones as the contents of the first-order ones, and so as candidates for endorsement, rejection or revision in the light of other first-order contentful states.” (Macdonald, 2007, 362)

elaboration of a self-image than in noting facts about oneself" (Anscombe 1975, reprinted in Cassam 1994, 157). Instead, Anscombe believes that "much of our knowledge of ourselves is acquired through interaction with other people, whose testimony provides one with an insight into truths about oneself which one may have no other way of knowing. Since we perceive others, and they perceive us, by means of 'outer sense', it would be quite legitimate to claim that the outer senses are at least as important as introspection in the acquisition of self-knowledge." (Cassam 1994, 4).

In a broader epistemological critique, McMyler argues that epistemic autonomy is overwrought and has not actually ever existed. Similar to critiques of individual autonomy, McMyler argues that epistemological accounts should be revised to include the *relational aspect* of accessing knowledge; since interpersonal relations generate epistemic knowledge, including self-knowledge, there is not much point in searching for the conditions of introspection if it is understood to require epistemic autonomy (McMyler 2011).

I accept that my broader claim about the challenges of vulnerability to introspective capacity may be taken to assume epistemic autonomy. Yet the fact that we acquire self-knowledge *also* through our relations with others is not problematic for my project here since there are important differences between the self-knowledge we gain through interaction with others, compared to the kind of self-knowledge accessible through introspection. First, what we may call *relational* self-knowledge is necessarily evidence-based (Macdonald 2007, 369). Others observe us in our interactions with the world, our deliberations with them and others, and draw conclusions about our state of mind, our intentions and goals. In contrast, the kind of distinctive self-knowledge that introspection can yield, authoritative and privileged, is not evidence-based since it focuses on our attitudes and intentional states: "one is actively reflecting on one's independently antecedently present intentional state" (Macdonald 2007, 362). Introspection thus focuses on acquiring a different kind of knowledge about ourselves than evidence-based self-knowledge seeks: whereas evidence-based reflection aims to verify "whether or not what one is thinking is true", introspection helps to determine "what one is thinking [...] This is what introspection is naturally suited to determining" (Macdonald 2007, 361).

This account of introspection as distinct from other epistemological methods to acquire self-knowledge is not challenged by those philosophers who question any idea of epistemic autonomy. Instead, introspection and relational epistemic models can all be conducive to self-knowledge.<sup>2</sup> To put this differently, it is uncontroversial that introspection is not the

---

<sup>2</sup> See here: "when I think about my first-order intentional states in an epistemically direct or immediate way, I authoritatively know the contents of my first-order (reviewed) states because I am better placed than others to know them, but my knowledge is neither incorrigible nor infallible. Because, when reflecting on such a thought, the thought on which I am reflecting is distinct from and independent of the reflecting one, my thought about what I am currently, consciously thinking is not contextually self-verifying and my knowledge is not infallible. And my claim to know what I am thinking is defeasible: others, who know my thoughts in an evidence-based way, may sometimes for this reason be in a

only way of achieving self-knowledge, if we accept that part of self-knowledge aspires to truth claims: in some cases, we may want to get things right *as the person we want to be* and to do so, we may need the help of others.<sup>3</sup>

#### INTERNAL REQUIREMENTS OF AUTONOMY

So far, then, I have argued that introspection is necessary to have access to self-knowledge about our intentional states. Earlier, I claimed that some kinds of vulnerability challenge introspective capacity; this poses problems for the internal requirements of autonomy. The definition of autonomy I apply here is that of *personal* autonomy as self-authorship (Raz 1987). This is to say that we are autonomous to the extent that we can decide on the shape we want to give our lives, and, most importantly for my purposes here, that we can ‘develop a conception’ (Raz 1987, 385) of ourselves :

“The ruling idea behind the ideal of personal autonomy is that people should make their own lives. The autonomous person is (part) author of his own life. The ideal of personal autonomy is the vision of people controlling, to some degree, their own destiny, fashioning it through successive decisions throughout their lives.” (Raz, 1987, 369)

To be able to engage in this kind of designing of our own lives, we need to know our intentional states – we need to know what we wish and hope for, and what we find valuable in life. Expanding on Raz’ definition, we can identify three further *internal* requirements for personal autonomy: first, individuals can only really be said to be autonomous if they have or show a disposition to endorse their actions and ends: “This might mean going through the actual process of reflecting on what one might want to do or be, and coming to a conscious decision about it.” (Colburn, n.d. 7)

Second, individuals can only be said to be autonomous if they have arrived at settling on a set of values and preferences through independent reasoning (Colburn 2015). This is not to say that my account of autonomy subscribes to an atomist ideal. Instead, and following Colburn, the independence condition pertains to “the property of the explanation of our decisions about what is valuable.” (Colburn, n.d. 7) This is to say that we need to be ‘able to construct self-selected rationale for action...making choices and confirming them’ (Dworkin 1989: 61). Autonomy in this sense is a ‘property of preference or desire formation.’ (Christman 1989:13).

---

position to show me that I am mistaken about what I am in fact thinking.” (Macdonald 2007, 368)

<sup>3</sup> To give a brief example - think of a father who deeply cares about his daughter. In his mind, he is solicitous, supportive and very engaged in her life. Assume, though, that everybody around him suggest to him that he develops helicopter syndrome: he literally hovers over every one of her friendships and interactions with others, intervenes in all of her decisions and is at great pains to make sure that “she is safe”. His behavior gets to a point that is no longer appropriate if the idea of bringing up children is to allow them to grow into their own personhood. It seems plausible that interventions from outside may help him realize that his actions go against his intentions to be “the best dad there is.”

And finally, an autonomous agent must be the one *responsible* for how her life takes shape: “a life must be attributable to an individual for it to be autonomous” (Colburn, n.d. 9). This attribution requires both, that one is at the root or the cause of the thing authored – that one has “explanatory” responsibility – and that one “has authority and stands in a certain normative relation to the thing authored” – that one has evaluative responsibility (Colburn, n.d. 10).

My claim is that some kinds of vulnerability challenge at least two of these internal requirements of autonomy, the endorsement and responsibility condition. If I understand the account of autonomy as self-authorship correctly, then both endorsement and responsibility condition fundamentally rely on our capacity to access our internal states. If we accept that introspection yields access to second-order thoughts, and that second-order thoughts are those that evaluate our internal states, then it seems plausible to say that the endorsement and responsibility conditions of autonomy rely on introspective self-knowledge. We can only formulate what we may want to do or be if we can actually reflect on our wishes and desires and appropriate them as ours. In other words, and accepting the claim that we can gain some self-knowledge through our interactions with others, we can only satisfy the endorsement condition if we *also* have access to introspective self-knowledge. And we can only count as having explanatory and evaluative responsibility if the values we have set ourselves are indeed based on our intentional states, rather than based on something else. So what kind of vulnerability might pose a problem to autonomy-enabling introspective knowledge?

#### VULNERABILITY, AUTONOMY AND INTROSPECTION - I

In the 1944 movie *Gaslight* by George Cukor, Charles Boyer in the role of the husband convinces his wife played by Ingrid Bergman that she is hallucinating and imagining a range of changes around the house, as well as changes in his behaviour towards her. Since then, the term “Gaslighting” has found its entrance into the psychological manuals, describing a form of manipulation “that seeks to sow seeds of doubt in a targeted individual or in members of a targeted group, hoping to make them question their own memory, perception, and sanity. Using persistent denial, misdirection, contradiction, and lying, it attempts to destabilize the target and delegitimize the target's belief.” (APA Diagnostic Manual)

Gaslighting is of course a very specific and dramatic form of vulnerability, in that it is based on the emotional dependence of the victim to its tormentor, a dependence that becomes heightened as the gaslighting goes on. Gaslighting, in other words, is intentionally messing with somebody's mind. It is manipulation of a person's belief about herself and the reliability of her knowledge of her intentional states. The moral wrong of gaslighting resides in the way of “charging someone not simply with being *wrong* or *mistaken*, but being in no condition to judge whether she is wrong or mistaken. The accusations are about the target's rational competence – her ability to get the facts right, to deliberate, her basic evaluative competencies and her ability to react appropriately: her independent standing as deliberator and moral agent.” (Abramson 2014, 8)

This resonates with accounts of introspection as a necessary vehicle for self-knowledge, and the link between self-knowledge and rationality:

“According to the simple theory, there is a necessary connection between rationality and self-knowledge. If one is ideally rational, then one is omniscient and infallible about one’s phenomenally individuated mental states, In other words, introspective self-knowledge is a constitutive ideal of rationality.” (Smithies, 2012, 289).

If this is plausible, I believe that we can expand Abramson’s analysis of the moral wrong of gaslighting more explicitly to the domain of personal autonomy. As I argued, autonomy relies on the internal requirements of endorsement and responsibility. Both of these require individuals to be able to identify and articulate their internal states, and to endorse, revise or reject intentions and desires as valuable or not. If, as in the case of Ingrid Bergman, too many doubts arise while self-scanning, the first step of self-authorship is thwarted. Rather than being able to ask what hopes and wishes she intends to pursue, an agent may be perpetually question whether what she thinks she hopes and desires is *actually* what she wants. Such doubt makes authorship impossible.

Gaslighting cases illustrate one kind of vulnerability that challenges the internal requirements of autonomy, to wit the emotional and increasing intellectual dependence on a manipulative other.<sup>4</sup> I want to expand the case further, however, and suggest that the challenge to introspective autonomy-enabling self-knowledge can arise from other settings of vulnerability.

#### VULNERABILITY, AUTONOMY AND INTROSPECTION - II

In her story “Paradise”, Edna O’Brien recounts the holiday of a radiographer who has met her new and older, very rich new lover when he came into her clinic. She accompanies him to his island some place warm where he hosts several others of his friends, all engaged in the summer activities of the leisurely rich. To all of the residents, including the servants, the fact that she doesn’t know how to swim and that he has hired a swimming instructor imported from the UK to teach her how to swim, is cause for investigation and bemusement. When she takes her lessons very dutifully sometimes several times a day, she overcomes her original panic and fear of that most delicious element that is water. All the while, not being able to swim and hence having a purpose to dedicate her days to allows her to keep apart from the others on their daily excursions by boat, to swim and harpoon. In the evenings, she joins them all for lavish dinners, during which she reflects on their odd constellation – not only does she now swim, but she is also vividly aware of their gazes assessing and comparing her to all the other women who came before her, not lasting past the season. All the while, a proposal is hanging in the air. On the penultimate day of their stay, she is to perform her newfound skills in the water to the entire household.

---

<sup>4</sup> Yet note that Abramson makes a convincing case how socially accepted gaslighting is when assessing female behaviour.

“She crouched until the water covered her shoulders, then she gave a short leap and delivered herself to it. Almost at once she knew that she was going to do it. Her hands, no longer loath to delve deep, scooped the water away, and she kicked with a ferocity she had not known to be possible. She was aware of cheering but it did not matter about that. She swam, as she had promised, across the width of the pool in the shallow end. It was pathetically short, but it was what she had vouched to do. Afterward one of the children said that her face was tortured. The rubber flowers had long since come off her bathing cap, and she pulled it off as she stood up and held on to the ladder. They clapped. They said it called for a celebration. He said nothing, but she could see that he was pleased. Her instructor was the happiest person there.”<sup>5</sup>

Immediately afterwards, while planning for the party, he tells her that the next thing she would learn was riding. Later on, when the party is in full swing, she attempts to drown herself.

This passage illustrates several aspects of a different kind of vulnerability that challenges the endorsement and responsibility condition of autonomy. O’Brien here describes the feeling of ‘standing beside oneself’ that often comes with being in the world. We experience ourselves in social contexts not only as protagonists and through our own lens, but also through the eyes of others. This of course sometimes creates conflicting impressions: the children think she is tortured, while the adults clap. Yet another source of conflict may arise from the incongruity between what we think we want – our introspective self-knowledge – and the self we want to be to the outside world. In fact, we may resent how others see us, and find ourselves cast in roles we don’t want, yet still try to inhabit them because they are plausible. We may want to endorse a version of ourselves and take responsibility for it – as O’Brien writes, “this is what she vouched to do” – yet this version may jar with yet another version of ourselves.

In other work, I have discussed the conditions of autonomy – I have argued that in order to be autonomous, we need to be able to define and choose norms along which to live, and that we need to be able to make these decisions without interference that renders them alien to us, so that we can identify with them (Straehle 2017). One important aspect of effective autonomy in my view is the need to be able to propose ourselves to the world as we intend to, with the reasons and values we have adopted, and for the world to recognize us the way we intend to. Much of this has been covered in the literature on recognition and the roots of epistemic injustice: If we are branded as incompetent or irrational by sheer belonging to a biological sex, for instance, then individuals are not able to be self-determining in the way that accounts of personal autonomy demand.

The vulnerability that O’Brien describes, though, is different. It points to the challenge that individuals face while validating intentional states when faced with equally plausible alternative plots along which they could author their lives. . The challenge to the endorsement and responsibility condition of autonomy then stems from the incongruity

---

<sup>5</sup> Edna O’Brien, *Paradise*. London:Faber 2019, p.55.

between two different sets of wants. We can call this vulnerability to incongruous introspective knowledge.

In a telling scene, O'Brien's swimmer, in one of her attempts to get to know her lover, and after realizing that she has had enough to drink to be drunk, asks: "Tell me (...) what interests you?" It was the first blunt question she had every put to him. 'Why, everything' he said. 'But deep down', she said. 'Discovery' he said, and walked away. 'But not self-discovery, she thought, not that.'

The swimmer herself is deeply invested in her own self-discovery – but her search seems thwarted by conflicting introspective results. Note that her vulnerability is of a different kind than that of a person being gaslighted – it is not that she constantly questions her intentional states as *hers* – rather it derives from her introspection yielding results that pull her into opposing directions. Or it may derive from the fact that her second-order thoughts about her intentional states – her reasons why she wants to know herself, say – don't endorse her first-order thoughts that she wants to please and fit in.

### VULNERABILITY, AUTONOMY AND INTROSPECTION - III

The swimmer's despair reveals itself only when she tries to kill herself. A different kind of circumstantial vulnerability illustrates a third limit to introspection as autonomy-enabling. Imagine the case of Beth, who is diagnosed with cancer. Being a patient illustrates the embodied ontological sense of vulnerability. Previously, Beth thought that she had a clear idea how much treatment she would accept if she were to become seriously ill, and to what pains she would be willing to go to stem the growth of further cancerous cells. Beth may be a scientifically minded person, who has studied statistics and is aware of success rates for different treatments of different cancers. Importantly, and in line with the idea of autonomy as self-authorship, Beth places high value on independence of mind and has led a life aiming to weigh options and risks carefully.

Recall now that I said that introspection as a path to self-knowledge is about determining the attitudinal relationship we take towards the world. I also grant that these attitudes can change over time, such that what we find through introspection may change. Indeed, as I said earlier, we may be corrected in some of our assumptions about ourselves through interaction with others.

Yet how should we think about the case of Beth, who, now and faced with the actual reality of having cancer, wants to hold on to her previous attitude to the world, while *also* act in ways that seem to contradict it. Now, Beth wants to try any kind of treatment possible, even the most experimental ones, while also, concurrently, being critical of these wishes.

Earlier, I said that some authors argue that knowledge accessed through introspection seems to be more secure and less error prone than other ways of accessing knowledge such as perception, testimony or memory. These all include a substantial part of interpretation to provide access to knowledge, we could say. They also demand a substantial amount of trust

– and belief: as MacMyler argues, “one can believe  $p$  but one must still come to one’s own conclusions about whether to believe the speaker.” (McMyler 2012, 149).

The idea here seems to be that perceptual modes of accessing knowledge are thus at least more open to debate than knowledge arrived at through introspection. Introspection in this vein comes across as ‘pure’ because unfiltered, immediate and authentic. The case of Beth, though, suggests that introspection requires a similar mediating effort as perceptual modes of knowledge: the vulnerability that comes from being a patient may demand leaps of *self*-trust in the face of conflicting introspective insight. Beth may question herself about her motivations to undertake treatment, she may worry about taking rash decisions, or ones that she may regret later on. Put differently, her second-order thoughts about her intentional states to want to be reasonable and realistic may be fundamentally at odds with her first-order desires to want to assure survival. To my mind, this case illustrates not only the conflict that O’Brien’s swimmer experienced, but a fundamental incommensurability of the values that introspection may lead to.

How does this relate to the internal requirements of autonomy? In the definition of personal autonomy that I have sketched so far, the idea of self-authorship is built on our capacity to set ourselves on a path of a life that realizes values. To do so, we rely on introspection to access and our internal states – we scan, as it were, our minds to find out what we hope and wish for, and we assess why we hold these attitudes to the world. This is the first step on the path to endorsing values, and to taking evaluative and explanatory responsibility for the life built on them.

In the cases of vulnerability, and particularly in the case of Beth, the cancer patient, introspection can yield that we have one set of internal states that is directing us into one direction of value, which is in conflict with another set. The conflict is heightened, of course, by the risk that is involved in pursuing one course of action against another. Yet we don’t have the option of not acting, although we may understand the choice of suicide in O’Brien’s story as an attempt to resolve the conflict.

## CONCLUSION

Much of the literature on vulnerability portrays it as a feature of human life that should, when morally problematic, ideally be overcome.<sup>6</sup> I have argued that introspection is an integral part to access self-knowledge, and that only if we have access to introspection can we satisfy the internal requirements of personal autonomy as self-authorship. My purpose here was to highlight that some contexts of vulnerability may raise problems to introspection as a mode of accessing self-knowledge. In a world, in which much value is put on autonomous decision-making, to question the conditions of autonomy is vital if we aim to give all agents their moral due. Sometimes, asking people to make decisions, to give consent to medical procedures, for instance, may impose an undue burden. Sometimes, we may find ourselves in situations of high vulnerability that make it difficult if not impossible to determine our attitude to the world.

---

<sup>6</sup> One notable exception is Carla Bagnoli who argues that vulnerability is necessary for individual agency (Bagnoli 2017).

