

Expanding the Active Mind

Jan Slaby^{a*}

^aInstitute of Philosophy, Freie Universität Berlin, Berlin, Germany

jan.slaby@fu-berlin.de

*corresponding author

Expanding the Active Mind

What I call the *active mind* approach revolves around the claim that what is “on” a person’s mind is in an important sense *brought on* and *held on to* through the agent’s self-conscious rational activity. In the first part, I state the gist of this perspective in a deliberately strong way in order to create a touchstone for critical discussion. In the second part, I engage with two categories of our mental lives that seem to speak against construing the mind as active. First, I discuss affectivity, in particular emotion, and show that emotional episodes are active engagements. Second, I discuss habitual action, and in particular those manifestations of habit which are initially opaque to the agent. In my responses to both objections, the notion of a practical self-understanding will play a central role. The result will be a qualified defence and expansion of the active mind position.

Keywords: Agency, belief, choice, emotion, habit, self-understanding, Richard Moran

1. Introduction

This paper addresses the bigger picture of philosophical thought about human agency. My aim is to reinvigorate the case for a time-honoured philosophical position about the human mind. At its core stands the claim that the human mind is a matter of self-constitutive agency: what is “on” a person’s mind is in an important sense *brought on* and *held on to* through the agent’s activity; it is her “making up her mind” about what to believe, desire, or intend. So ours is essentially an *active* mind, and the scope of this self-shaping agency encompasses all the major categories of mental comportment, notably its epistemic, emotive, and conative dimensions. In one way or another, this “active mind” thesis, as I will call it for the sake of brevity, has been a staple of the philosophical tradition since Aristotle. It can enlist thinkers such as Kant, Hegel, and Wittgenstein, as well as Heidegger and Sartre among its proponents. It has also found firm adherents in the analytical and post-analytical camps of contemporary philosophy,

with authors such as Anscombe, Hampshire, Korsgaard, McDowell and Moran (to name a few), each of whom elaborate different versions or aspects of it. Obviously, in view of this vast list of names, the spectrum of views collected under the scope of the active mind is rather broad.

In the limited space of this paper, I can only remind readers of the gist and main considerations in favour of the active mind position, after which I will then sketch a strategy to defend an *expansive* version of the perspective. I will show how the view is able to accommodate a range of mental phenomena that might initially seem to fly in the face of its activist upshot. These phenomena seem to be shaped by factors that are not transparent to the agent, let alone by conscious control, and thus make it hard to see how they might fit with the assumption that mental comportment is, at its core, a matter of agency. What the editors of this special issue call the “extended field of action” consists of processes, many of which are inaccessible to the self-conscious agent at the moment of performing an action, for instance habitual actions or actions expressing emotion or other modes of affectivity. Such mental going-ons can have a profound sway on the operative intentions of self-conscious agents. So, how can we accommodate these phenomena into the active mind? By addressing these issues, I hope to instigate a renewed debate between the more analytical currents in the active mind camp and proponents of the phenomenological tradition, not least with those of the latter who have been sceptical towards views of the human mind centred on agency.¹

¹ It is interesting to consider the phenomenological tradition in this regard, as its allegiance is split between differing versions of the active mind thesis. A strong current in phenomenology tends to oppose the Kantian strand in modern philosophy, which champions stronger and normatively inflected versions of the thesis, while another faction – spearheaded by Heidegger, with strong echoes in Sartre and a more balanced uptake in

The paper has two main parts. In the first part, I will state the core of the active mind position in a deliberately strong manner, so as to build a touchstone for the following discussion. Here, I will deal foremost with the epistemic dimension of the active mind, focusing on “active belief”², before closing with reflections on Sartre’s invocation of a mandate of “choice” as being decisive for consciousness. In the second part, I engage with two categories of our mental lives that can seem to present difficulties for the active mind. First, I will discuss affectivity as a potential limit to the active character of human mindedness. I answer it by arguing for the active core of emotional comportment and by also suggesting the notion of a practical self-understanding in order to show how mere affective impingements are taken up within a context of self-conscious activity. Subsequently, I discuss habitual action, notably those manifestations of habit which are opaque to the agent. I draw on considerations by Line Ingerslev to disperse the worry that such manifestations of habit have to be counted as

Merleau-Ponty and also Ricoeur – develops credible and phenomenologically “safer” versions of it. As this text is not about the history of ideas, I remain largely silent about specific authors and their allegiances to versions of my core claim. Let me note, however, that my employment of the term “mind” is more directly aligned with the tradition of reflection as found in post-Wittgensteinian Oxford philosophy (Ryle, Anscombe) and its current adherents such as Richard Moran, Matthew Boyle, also more broadly John McDowell and other so-called post-analytical philosophers. One key feature of this notion of “mind” is that its proponents do not consider it an open question whether animals or pre-linguistic infants do have “minds” in the sense at issue, but rather assume self-reflective capacities that are tied to language as a key prerequisite of the kind of mindedness under discussion.

² Boyle, ‘Active Belief’.

unintentional and thus outside the scope of action proper. As in the case of affect and emotion, the notion of practical self-understanding will play a significant role in accommodating habit within the active mind.

A further, more difficult task lies beyond the scope of this text: to show how the active mind can be constitutively expanded to also encompass aspects of the social and cultural environment. I believe that social factors are always in play as a formative backdrop, enabling structured and intelligibly self-conscious agency in the first place. Therefore, I am receptive to stronger constitutive claims concerning the role of the social and material environment. However, the core claims and central normative outlook of the active mind position can be stated without committing to a specific view about such enabling conditions. While this bigger endeavour – and in particular, developing the position so that it jives with “the socially extended mind”³ – cannot be undertaken here, what I do can clear the way for subsequent studies that situate the active mind more firmly in its formative surroundings.⁴

2. The Active Mind: Big Picture

To have a mind is to *be* a mind, and to be a mind is to be a rational agent – an agent undertaking and living up to epistemic commitments (such as beliefs), undertaking and

³ Gallagher, ‘Socially Extended Mind’.

⁴ Given this moderate ambition, the term “expansion” in this paper’s title is not meant in a technical sense, for instance suggesting something along the lines of an “extended mind” theory. Rather, “expanding” is meant to bring out the centrality and the reach of agency for key dimensions of *individual* mindedness, enlarging the position’s scope to cover not just epistemic comportment but also emotional episodes, habitual action, and potentially other manifestations of sapience.

living up to conative commitments (such as desires), and actively committing to courses of action laid down in intentions to act.⁵ Thus, the agency in question, when it comes to what gets usually and misleadingly called “mental states”, is chiefly the making up of one’s mind about what is true (the activity of believing) and likewise the making up of one’s mind about what is worthy of our pursuit (the activity of desiring, and, built upon that, of intending).⁶ Accordingly, so-called *mental states* are not items in an inner theatre, available as objects of introspective awareness, but are rather the various *active stances* that a sufficiently rational agent can adopt towards the world and herself. Whether this point is made in terms of Kantian spontaneity of the understanding, Heideggerian Dasein as called upon to authentically hold itself to an understanding of being, or Sartrean consciousness as nihilating the given and issuing responsible verdicts on how to go on, it is always a matter of assuming a position of agency that is not an

⁵ A clarification about the notion of “rational” employed in this text is in order. When I speak of “rational agency” or “the rational agent”, I mean no more than the basic conditions of reason-responsiveness and consistency that ensure an intelligible outlook on the world persisting over time so as to give the term “agent” a foothold. The rationality required for this is not only compatible with vastly different thick conceptions of lifeworld rationality, accounting for broad ranges of cultural, historical, and milieu-specific differences, but it also allows leeway for considerable instances of irrationality and individual or group-specific idiosyncrasies. Thus, it would be wrong to blame the present account for promoting a form of reductive rationalism in the mold of rational choice theory or instrumental rationality, as some readers of earlier drafts had feared.

⁶ Throughout this exposition I will take many hints from Richard Moran, whose encompassing formulation of the position in *Authority and Estrangement* lays the foundation for what I will do in the following.

option but a requirement for mindedness.⁷ A similar point can be made in the more pedestrian terms of deliberation and rational accountability, as has happened in more recent articulations of the active mind, for instance by Richard Moran and Matthew Boyle.

2.1 Active belief and practical knowledge

Moran's account is mainly framed as a discussion of self-knowledge and the first-person perspective, but by implication applies to the full scope of the active mind. On this view, *belief* functions as a central illuminating instance. Belief also figures prominently, but with a different emphasis, in Matthew Boyle's work on this topic.⁸ In view of this, I will begin by talking about belief, before the purview of the discussion is expanded to other types of attitudes.

Consider being asked what you believe about some non-trivial matter. Do you really respond to this by "looking inside", as it were, to see what it is that you believe about the theme in question? It is much rather the case that you look out into the world, toward the theme in question and to the relevant facts about it, and make up your mind about them. You consider the issue, you deliberate, for instance by actively recalling what you know about the theme, by weighing available evidence, and eventually

⁷ This aligns the present account with an important current of post-phenomenological thought that is labeled "enactivism". However, there is a marked difference in emphasis, as enactivists in their attempts to develop the de-facto agentic character of the mind usually stress the continuity of animal and human mindedness and therefore give short shrift to higher-level reflective capacities. An important exception to this trend, and thus much closer to what is developed here, is the work of Alva Noë, see e.g. Noë, *Action in Perception*.

⁸ Boyle, 'Active Belief'; Boyle, 'Making Up Your Mind'.

arriving at a conclusion on how the case lies. You thereby actively shape your attitude towards the topic. And when you *avow* your belief, you thereby express a commitment that includes, among other things, a readiness to defend your view when challenged, assent to what follows from your belief, and so on. All this is a far cry from just finding oneself saddled passively with some mental content or other. And while there is content in play here, it is foremost the facts “out there”, not some mental items “in here”, which are decisive and thus in the focus of what must accordingly be your active orientation.⁹ In this manner, a self-ascription about a given topic meets what Moran, following Gareth Evans, calls the *transparency condition* insofar as it is made by consideration of the facts about the theme itself. Belief is transparent to the world in this way; it is the ability to let one’s epistemic stance be controlled by what is the case.¹⁰

Insofar as I am *active* on these occasions – I actively *make up my mind* – we deal with a form of immediate, practical self-knowledge. In the sense at issue, self-knowledge about what is on my mind is, in a crucial respect, similar to self-knowledge regarding action more generally. I know what I am doing not by observing my deeds, which would be oddly alienating, but by *performing* them. And in the performing of my actions, there is no distance between my doing and my knowing such that my self-knowledge of agency is in this sense immediate, that is, not based on evidence.¹¹ This is a peculiarity of the first-person perspective. I do not gain knowledge about my own mind empirically through observation, but I know directly, without recourse to evidence, what is on my mind: *I know it*, one might say, because *I’m doing it*. Moran elaborates this feature as follows: in the central cases, first-person authority derives

⁹ See McDowell, *Mind and World*.

¹⁰ Moran, *Authority and Estrangement*, 92–93.

¹¹ See Anscombe, *Intention*; Thompson, ‘Anscombe’s *Intention* and Practical Knowledge’.

from first-person *authorship* of the attitudes in question. Yet it would be imprecise, although on the right track, to say that I actively *create* my attitudes. It is more to the point to say that these attitudes *are* my activities; at issue is not their “origin” (which might indeed be elusive), but my *holding* them: attitudes such as beliefs exist only insofar and to the extent that I actively commit to what is true. This shows vividly why talk of *mental states* is so misleading: what is at issue are active stances, rational postures actively adopted by an agent in a situation.¹² The self-knowledge at issue is distinct from observational knowledge because it is *practical knowledge* – knowledge which inheres in the activity of self-determination, and inheres in it by virtue of its being something I do.¹³ Accordingly, when I sincerely state what I believe, I do not *report* my mental state, but *avow* my attitude.¹⁴ First-person authority is, in this sense, a matter of *authorship* of one’s attitudes.¹⁵

Of course, besides practical knowledge of my mental agency, I can also obtain observational knowledge of my mental comportment. In terms of achieving this knowledge, I am in no fundamentally different position from a third-person observer. What is important to see is that such observational knowledge must be secondary vis-à-vis deliberative knowledge because I would not be an agent with mental capacities at all were it not for my ability to deliberate about and actively commit to what is true, right, or proper in the core cases.¹⁶

¹² Note that this does not amount to doxastic voluntarism: I cannot believe at will because I am committed, qua my capacity for belief, to only believe what I deem true. See Moran, *Authority and Estrangement*, 51–55.

¹³ See Anscombe, *Intention*.

¹⁴ See Moran, *Authority and Estrangement*, 83–93.

¹⁵ See also Hampshire, *Freedom of the Individual*, 97.

¹⁶ See Moran, *Authority and Estrangement*, 150.

The claim that believing, in particular, is an activity seems to fly in the face of conventional wisdom in the philosophy of mind. A dominant view is that beliefs are *standing* attitudes in the sense of temporally persisting mental *states*, as opposed to mental *acts*, such as judgments. A consideration against this standard picture of belief can be generated from a closer look at the assignment of epistemic responsibility that comes with belief attribution.¹⁷ We usually hold those to whom we attribute beliefs rationally responsible for heeding the epistemic standards of believing, that is, we demand that they be ready and able to adjust their beliefs in the light of new evidence and that they are capable and sufficiently willing to justify their beliefs upon request. Now, upon closer inspection, it becomes clear that the responsibility we attribute to believers immediately concerns these agents' *holding* of beliefs, not merely their (putatively rare) acts of *acquiring* beliefs. If we inquire into why someone believes that P, we demand to know what speaks in favour of P *right now*, regardless of how, or how long ago, the agent has come to acquire the belief in question. If everyday interactional practice is any indication, then we can say that a cognitive agent must be ready *throughout* to defend their beliefs. Thus answerable to rational challenges, an agent's "holding their beliefs" seems quite clearly to be a self-conscious agentic stance, something that we rightly consider to fall within the agent's active powers, and in this important and strong sense, something they do. This consideration mitigates against the notion that beliefs are somehow passively stored, as a cognitive possession. If beliefs are "had" or "possessed", then they are so in the sense of being actively held – an activity of the agent, albeit one different in kind from putative "acts of judgment".¹⁸

¹⁷ See Hieronymi, 'Responsibility for Believing'.

¹⁸ Obviously, much more could be said here. A deeper diagnosis of what goes wrong in accounts that adhere to the standard picture of belief as inert, standing attitudes is provided

2.2 Consciousness and the Mandate of Choice

In light of this discussion of active belief, we are in a good position to better grasp the scope and strength of the active mind. At a central place in Moran's account, a key thought from Sartre's *Being and Nothingness* figures prominently: "Choice and consciousness are one and the same thing."¹⁹ What Sartre's flashy proclamations concerning the active nature of consciousness amount to, according to Moran, is the claim "that no attitude or impulse apprehended by reflective consciousness has any right to continued existence apart from one's free endorsement of it"²⁰. While indeed radical, this claim is both focused and reasoned as it allows ample room for heteronomy and alienation while still being unremitting in its intended scope. It applies to the full gamut of non-pathological mental life:

When I am reflectively aware of some attitude or impulse of mine, I am thereby made aware that its persistence in me (as a "facticity") is not a foregone conclusion stemming from the inertia of psychic life, and in particular that its counting as a reason for me in my current thought and action is my affair.²¹

Sartre was adamant to construe even episodes where the agent seemingly "does nothing" – situations where someone coasts along in the thrall of whatever impulses may happen to arise – as a form of agency. Sartre's point was that even in those cases, the active mind must be seen as having a say in determining the subsequent course of action or conduct. A self-conscious agent is *tasked* with taking a stance so that whatever they do, not do, or merely "let happen" must count as an exercise of agency.

by Boyle, 'Active Belief', who helpfully clarifies the kind of activity at issue in believing and contrasts it with the standard picture of the putatively punctual "act" of judging.

¹⁹ Sartre, *Being and Nothingness*, 595.

²⁰ Moran, *Authority and Estrangement*, 140.

²¹ Ibid.

Moran agrees, and reframes the point in terms of deliberation. Everything that I mentally undergo *thereby* becomes subject to my assessment as to whether *it should* figure as a reason in the course of my ongoing accountable conduct.²² In other words (and this is central to Sartre), from the perspective of the agent, choice is never merely an option. Choice is forced upon the agent. This is the decisive difference between the first-person position and the third-person vantage point. Of course, as participants in social interaction, we are in the position to *also* assess and cast judgment upon *another person's* mental impulses, in case these are brought to our attention. In these more distant third-personal encounters, we always have the possibility to opt out and withhold judgment.²³ I can let someone else off the hook, as it were, in the sense of remaining noncommittal as to whether I approve or disapprove, accept or reject, endorse or dismiss what a putative mental content is bringing to attention. But I cannot similarly opt out from taking a stance in my own case. A fundamental fact about the first-person perspective is that it does not allow distancing in this manner. To be a subject entails this “irreplaceability”, since it means that the activity in question and the responsibility for it must not only be *someone's* by default or merely *the same* someone in both cases, but it must also be, on pain of destroying the logical coherence of the agentive perspective, irrevocably *yours* or irrevocably *mine*. There is a mandate of authentic ownership to being an agent, or *Je-Meinigkeit*, to use Heidegger's term.²⁴

²² See Moran, *Authority and Estrangement*, 142.

²³ This need not apply to all interpersonal encounters, but only to those that allow a specific social distance which is reflectively acknowledged. Besides these, there are forms of more intimate interpersonal engagement which do not, or at least not initially, allow such distance to take hold (see, e.g., De Jaegher and Di Paolo, ‘Participatory Sense-Making’).

²⁴ Heidegger, *Sein und Zeit*, 41.

Moran sums this point up in the Sartrean terms of self-consciousness and commitment:

That is, self-consciousness is figured here as forcing the issue of some kind of commitment that the person cannot shift on to another, with the understanding that, like commitments of other kinds, it remains in effect when the person alienates himself from it or otherwise fails to fulfill it.²⁵

One potential misunderstanding should be cautioned against at this point. The idea that choice is forced upon the agent might still allow for a weak reading along the following lines: all that an agent's activity in relation to a given mental episode could amount to is an assessment in the sense of either approving or disapproving of what mentally sways her. On this reading, it could be a normal, even exemplary case of what is at issue here, to be in the thrall of a strong impulse while vehemently but ineffectively "disapproving" of it, yet without any further effect. If this were all that the agency under discussion amounts to, we would end up with a disappointingly weak position. The agent could as well be an opinionated bystander beholding mental stirrings without being able to change their course. While we can grant that occasions of being overwhelmed and rendered helpless *might* and *do* occur, for instance if an agent is momentarily shaken by exceptionally strong affects, the account sketched here cannot allow these to be more than relatively rare extremes. Instead, the position under discussion demands that the agent's deliberate capacity, once exercised, makes a substantive difference. By subjecting an inclination or impulse to rational assessment, that inclination or impulse gets suspended in its force, deprived of both its activating sway and inferential legitimacy until a considered verdict is reached. In practice, this suspension and transformation of mental going-ons can take various forms, including modes of

²⁵ Moran, *Authority and Estrangement*, 142.

habitual, sometimes nearly “unthinking” acquiescence, so that we do not have to assume a counter-intuitive phenomenology of stepwise explicit rational assessment.²⁶

Deliberation is a capacity to *effectively* commit, that is, a capacity to actively form and hold one’s attitudes – not merely to appraise and assess them from a distance. Moran elaborates:

This is a perfectly homely assertion of one’s freedom. It is what is exercised in the undramatic situations of making up one’s mind about some matter: I decide that my conclusion was hasty or based on a failure of imagination, and that conclusion itself is thereby pulled out of circulation, either temporarily or for good; I decide that I can’t keep looking for my keys and holding everyone up, and I thereby alter my intention; I decide on reflection that my eagerness to tell an embarrassing story about my ex-wife actually reflects more badly on me than on her, and my desire changes. To take oneself to be in a position to ask and answer this sort of question about one’s belief or intention is to take oneself to be in a position to *make* something true in one’s answering it.²⁷

Moran is right, I think, in making this sound like a mundane, indeed homely, capacity, something that is in play unremarkably throughout our lives. Yet we also have strong intuitions about the limits of such agentic autonomy, especially with regard to sustained, character-defining attitudes, many of which are found in the emotive spectrum. Accordingly, we have to find a way to accommodate various scenarios where it seems as if the agent is much less actively involved in determining what is on their mind.

To this end, one preliminary consideration goes as follows: it is a categorical fact about a capacity that it need not be exercised on all occasions where it could be

²⁶ This problem recalls the McDowell-Dreyfus debate on whether experience is conceptual all the way down and whether this would require the actual involvement of concept tokens in conscious experience, as Dreyfus seemed to think; see the contributions in Schear, *Mind, Reason, and Being-in-the-World*.

²⁷ Moran, *Authority and Estrangement*, 145–146.

exercised. So it is possible to be capable of deliberation, and to do it often enough to qualify as a rational agent, while still failing to exercise this capacity on a certain number of occasions and for various different reasons – for instance arrogance, time pressure, carelessness, or self-delusion.

With these considerations, we have moved towards the wider field of phenomena that need to be accommodated in order to ensure the broad applicability of the active mind. The “extended field of action”, as mentioned in the introduction, consists of the broad sphere of habitual conduct and also encompasses a good deal of an agent’s affectivity. We will now turn to these focused challenges to the active mind.

3. Expanding the Active Mind

An obvious problem for the active mind is presented by affective experience. It can seem that emotions and other affective phenomena, while clearly mental, are not things the agent *does*. I will invoke a consideration that will rehabilitate a sense of agentive involvement in emotion. Habitual action is another potential stumbling block for the position under discussion. In a considerable number of cases where I act habitually, I do something that I don’t intend *actively* and *reflectively*, and to the extent that I do intend it, I intend it under a description that does not capture the specific way of *how* I am doing what I do.²⁸ The problem posed by affectivity is in one key respect similar to the problem posed by habitual action, as affective phenomena can manifest orientations towards world, self or others that seem alien, or partly alien, to the conscious agent. Insofar as these orientations prompt us to act on them, they seem to undermine agentive autonomy.

²⁸ See Ingerslev, ‘On the Role of Habit’

3.1. Emotion and the Active Mind

There is a simplistic, yet almost proverbial image of emotions as a person's being in the thrall of a force from without that shuts down reflection and deliberation, replacing them with near-instinctual frenzy and thoughtless upheaval. On this view, emotions are alien impulses, and thus neither something an agent does nor something they can reasonably endorse through reflection and deliberation. Much work in 20th century philosophy of emotion has demoted such a reductive picture. A more adequate view understands emotions as complex forms of engagement with the world, revolving around a matter of concern which the agent apprehends affectively and upon which the agent usually acts out of a fitting evaluative orientation. In their temporal unfolding, emotions might come with phases of relative heteronomy, and especially the onset of episodes of, say, grief, anger, shame, or intense fear can temporarily suspend the active capacities of the agent.²⁹ On many occasions, however, such passionate transfixion is relatively short-lived, whereas the emotional episode itself often endures. Grief, while often kicked off by a shock-like state, is only truly intelligible as a longer-term process of grappling with the loss of a loved-one. Anger might sometimes be eruptive and overwhelming in the short term, but often stabilizes subsequently into a broader pattern

²⁹ I cannot go into the intricate debate about the proper sense in which emotions can indeed be understood as “passive”. Jean Moritz Müller has recently illuminated this issue by bringing out that emotions must be understood as forms of reason-responsiveness which renders them “spontaneous” (in a roughly Kantian sense), instead of “receptive”, as many other views championing emotional passivity seem to imply. Despite significant differences in the details, Müller's clarification supports the gist of my approach as it understands emotions to be *activities* that contrast sharply to perceptions as paramount exercises of receptivity (see Müller, ‘The Spontaneity of Emotion’).

of confronting an opponent in light of an offense, probably leading to a longer-term “revenge project” that occupies the emoting agent for quite some time. Shame, while potentially catastrophic and momentarily erosive of an agent’s composure, often gives way to a longer sequence of milder embarrassment interwoven with a conscious effort to regain control over the situation, and so forth. Accordingly, in emotional episodes of these relatively enduring varieties, to which many of our more weighty real-life emotions conform, the agent is very much in the picture as, in a sense, performing or “managing” a complex sequence of engagement with the world.³⁰

While it is interesting to see how the overall emotional episode affects the agent’s rational comportment, it would be wrong to say that the self-constitutive agency required for deliberation is suspended or shut down. What is clear, however, is that the situation here looks different from the case of belief, where epistemic norms directed at truth and truth-conducive evidential standards clearly have a more central role. Emotional episodes bring to bear various strategies of placing oneself in a certain light, of letting the world show up under specific aspects, sometimes in ways complicit with the overall emotive orientation, which may lead to a self-confirmatory circle. Jealousy, for example, is notorious for leading the agent onto a path of finding ever more evidence for the jealous suspicion at the base of the attitude. Some instances of fear tend to spiral into dreadful, self-stabilizing loops, as if the agent were wearing

³⁰ See Goldie, *The Mess Inside*, Ch.4. Besides drawing on the work of Peter Goldie, who has championed the view of emotions as enduring episodes or processes that connect various elements into a complex whole (see Goldie, *The Emotions* and Goldie, *The Mess Inside*), I take up hints from various authors who have proposed agency-based approaches to emotion. For instance, Griffith and Scarantino ‘Emotions in the Wild’ present a powerful naturalistic account in line with considerations from ethology, evolutionary biology and social psychology. Slaby and Wüschner ‘Emotion and Agency’ outline a perspective more in line with the phenomenological tradition by drawing on Sartre and Merleau-Ponty.

tinted glasses that highlight dangers everywhere. Also, and more vexingly, emotions tend to invade higher-order reflective capacities so that the agent may consciously endorse even quite irrational first-order attitudes as putatively warranted. Sartre spoke of reflection on one's emotive consciousness as "accessory" and "after the fact"³¹. Goldie, in an epistemological key, noted that emotions tend to "skew the epistemic landscape"³².

Relatedly, it makes sense to remember Heidegger's observation in *Being and Time* that what he calls moods or attunements disclose the world to the agent, but does not tend to do so in a straightforward mind-to-world direction. Instead, attunement discloses in the mode of "turning away".³³ That is, emotions seem to "allow" the processing of reality only in limited, strategically selected doses, often so as to suit the agent's specific purposes. In the grip of shame, individuals tend to avert their gaze from the situation at hand to not have to face the reality of the shameful occasion. The same can be true in many instances of fear, when one allows oneself only brief glimpses at the feared object, so as to prevent the full load of dread from arising. Such evasive and piecemeal emotive disclosure takes many forms, and it would be wrong to just dismiss these temporary limitations in the emoter's capacity for the real as unreservedly irrational. Emotions adhere to standards other than those of belief – they come with their own framing of reality and the agent's position in it.³⁴

³¹ Sartre, *Sketch*, 61.

³² Goldie, 'Emotion, Reason, and Virtue'.

³³ See Heidegger, *Sein und Zeit*, 135.

³⁴ This is part of why cognitive theories of emotion are so problematic from a phenomenological vantage point: while such views are correct to emphasize that emotions have formal objects, they misconstrue the peculiar way in which emotions disclose a part of reality and thus fail to capture the unique characteristic of affective intentionality.

Despite these peculiarities, the involvement of agency at the heart of emotional comportment is not in dispute. Here, Sartre is again instructive in his identification of emotions with strategic actions that aim at a “magical transformation” of the world. If reality is too hard to bear because it poses too many obstacles, a strategic denial or selective uptake in the manner of a “small drama” that the emoter performs can be the appropriate response. At issue is the specific situated perspective of an agent on the cusp of being overwhelmed, nearing breakdown. Accordingly, emotions, as Sartre construes them, are feats of theatrical make-believe capable of fooling others, but also, and sometimes foremost, oneself. Sartre makes this clear: “If emotion is play-acting, the play is one we believe in.”³⁵ They are stances to which we have truly committed. If Sartre is on the right track, then many emotions might surely be criticized as irrational. It is clear that they fail to meet the standards of rational deliberation (although it is notoriously difficult to determine the adequate standards of rationality in the case of emotion, as emotive comportment is itself partly constitutive of what is important or worthy of pursuit).³⁶ But more importantly, emotions do not haphazardly fall short of the relevant rational norms; they do so in specific, circumscribed ways that make sense in their own right, as they have their own strategic logic. This logic originates from the projection of a “magical” reality, which endows the episodes with their own manner of contextual significance. Robert C. Solomon has done a lot to elaborate this point by discussing the “surplus reality” projected into the world and enacted through emotion, thus rendering emotional deviations from standard types

³⁵ Sartre, *Sketch*, 41.

³⁶ A noteworthy and to date still highly relevant analysis of this specifically emotive rationality is Helm, *Emotional Reason*.

of rationality not only intelligible but also potentially prudent (in some regards), given the circumstances of and options available to the agent.³⁷

What all this shows, I think, is that agency is at the centre of emotional episodes and that whatever else is true of human emotions is additional and adjacent to emotion's agentic core. The agent remains – if not outrightly *in charge* – definitely at the centre of the picture for much of an emotional episode's unfolding, calling the shots in a complex sequence of engagement with the world. However, this agentic position is both *precarious* (at any point, the agent might be temporarily overwhelmed again by another affective impulse) and *uncomfortable* (given the responsibility and blameworthiness that comes with assuming this position). Thus, as Sartre never tired to point out, it can be beneficial for the agent to disclaim any truly active involvement in emotional engagement, settling for *mauvaise foi* instead. Moreover, what is specific about the agency involved in emotion is the way in which it can bring the practical orientation of the agent as a whole under the sway of a decidedly one-sided evaluative tendency (sometimes, of course, this enthralling one-sidedness is perfectly apt, even from a less partial rational vantage point). When such a tendency reigns, the agentic perspective as such, the capacities of the agent to reflect upon and assess what goes on with and around them, are modified in ways that may elude the agent. That can create the impression that emotions cast a passivizing spell upon the subject. Now, while transparency in Moran's sense is not threatened – after all, even my exaggerated jealousy is eminently world-directed and finds fresh evidence for my partner's infidelity everywhere I look – the deeper roots of my affect-tinged outlook as such remain

³⁷ See Solomon, *The Passions*.

opaque. *What is it about me that makes me so prone and willing to see the world in this particular way?*

What this shows is *not* that there is no agency involved, but only that the agent at the centre of the active mind is not a neutral, disengaged figure without qualities. The agent is a situated, flesh-and-blood individual with a concrete history; a person with concerns, character traits, dispositions. And these concerns, traits, and dispositions come to bear in emotional comportment. The active mind view does not require the agent to be capable of achieving full insight and self-transparency with regard to these conditioning factors; rather, it demands that the manifestation of those elusive regions of the self raise a concern or pose an earnest question to the agent, instead of remaining a matter of indifference. Here again, the first-person perspective inherently demands that the agent adopt a *stance* on whatever is effectively manifest within it. This does not mean that the agent has to find a substantive and satisfying answer that would illuminate the impulse or orientation in question.

Instead, the claim is that the stance of reflection, being the stance from which the status of some “psychic given” *as* a reason may be constituted or suspended, makes it the case that my response to that “given” can now be understood in terms of the person’s responsibilities.³⁸

The *reach* of reflection, and thus of self-constitutive agency in the sense at issue, might exceed the scope of what is currently graspable by the agent as a definitive, endorseable determinant of her stance. Yet reflection as such, and so her *taking* of a stance, are forced upon her.

At this point, it might help to address the extent to which the present account is a normative one. That Moran’s position has a strong normative thrust is undeniable, and

³⁸ Moran, *Authority and Estrangement*, 148.

this surely carries over into what is presented here. But it is important to clarify what this means. What has been developed so far aims to bring out why certain normative demands are *rightfully* placed on most adult human subjects. Indeed, what is so fascinating about, for instance, Sartre’s writings on self-consciousness as choice, is their appeal to the very agentive core of their readers’ conscious perspectives. In this respect, such a philosophy might be compared to a pep talk, spurring a somewhat feeble, potentially self-forgetful agent back into earnest attempts at responsible self-determination. In line with this, the “normative atmosphere” that the present article might emanate amounts to the following: not even trying to take charge of one’s situated existence is in any case much worse than trying and ultimately failing to come to terms with an elusive, somewhat alien, or potentially obnoxious aspect of one’s mental life. The account is then normative in a performative and evocative sense, but also by crucially bringing into view what makes us fitting addressees of such normative demands in the first place.³⁹

In light of this, one might read the considerations offered here as giving shape to a version of the familiar notion of a practical perspective or self-understanding, roughly in the sense outlined by Charles Taylor.⁴⁰ Simplifying a bit, one can say that the self-

³⁹ Whether this robustly normative inflection prevents the account from expressing “homely” truths about human subjects, as one reviewer suggested it does, is debatable. That we are open to normative demands and often heed these calls is not in doubt, yet obviously our emotive attitudes hover right at the (inevitably disputed) border between the voluntary and involuntary. In view of this, one might contend that there are indeed few homely truths to be found in this region.

⁴⁰ See Taylor, ‘Self-Interpreting Animals’. This is certainly a position that entwines a normative outlook with a phenomenology of situated embodied existence.

understanding, as the agent's partly elucidated, but always concretely lived "thick" perspective on both the world and herself, realizes the dimension of manifest agency in an expanded sense. This means that not only are her actions and activities, namely what the agent intends to do under a description, embedded within her self-understanding, but – and this is crucial – so are those going-ons which initially merely impinge upon the agent and make their presence felt *within* her self-understanding. So one could say that a person's self-understanding is the region shared by bona fide agency and all those going-ons which are, at least initially, passively undergone. A practical perspective is such that it is *modified by* and *tasked to respond to* any impulse or "given" that the agent finds herself confronted with – by either taking it up, integrating, and running with it, or else dismissing, blocking out, and disregarding it; or, if all else fails, enduring, somehow putting up with it, and waiting it out. A lived self-understanding is the modicum of agency even where effective intending is temporarily suspended due to overwhelming impulses. It is, to use a formulation suggested by Line Ingerslev, *the medium in which the foreign and alien can become mine*.⁴¹

In this admittedly schematic manner, the position under discussion can accommodate all sorts of more robustly "alien" impulses without giving up the core claims of the active mind. Yet it is important *not* to construe the actual practice of adjusting one's self-understanding as a matter of stimulus and response, which would be overly simplistic. Rather, most of what affects the agent by impinging on her is *thereby* already couched within the terms of the practical perspective. The "deliverances of receptivity", to use McDowell's phrasing, arrive such that they modify a practical

⁴¹ See Ingerslev, 'On the Role of Habit', 492.

outlook, and so they are in turn always already met with an attitude that accommodates them in one way or another.⁴²

The notion of a practical self-understanding as the substantive, worldly, situated, and historicized manifestation of a first-person perspective – the full embodied agent, in short – also helps us deal with the other major consideration that threatens the strength of the active mind position: habitual action.

3.2. Habit and the Active Mind

From a general vantage point, habitual actions might seem to be less of a stumbling block for the position advanced here. In fact, they might even lend some additional credibility to the active mind view. Habitual action, as long as it is within the immediate range of truthful intentional self-ascription on the part of the agent, can help answer the common objection to the effect that while they are happening, we don't seem to reflect and deliberate very much about our intentions and our doings. A proponent of the active mind can say that habit takes over in cases where the best course of action is by and large already settled, as long as the self-conscious agent *can* re-enter the picture when the situation demands it. On the condition that explicit reflection, re-consideration, or re-commitment of the intentions in question kick in at appropriate times, all is well for the active mind. Habitual action passes Anscombe's test for intentional action if the agent can, upon being interrupted in performing such an action, truthfully and without

⁴² Note how this intelligent intertwinement of receptivity and spontaneity, experience and action, or passivity and activity within an expansively construed embodied self-understanding is something that links the views of Taylor and McDowell quite closely with Merleau-Ponty's phenomenology. See Slaby & Wüschner, 'Emotion and Agency' for more specific hints in this direction.

recourse to evidence, answer the question “Why are you doing X?”.

However, the issue is more complicated for a significant range of habitual actions where this test simply fails. Line Ingerslev has discussed these cases. Her example is a routine of unthinkingly greeting two colleagues very differently each morning at work, where the particular manner of doing so is not clear to the agent, but can be readily observed by others. Here, the agent seems not to be aware of what he or she is doing, as the intentional description available to her is too thin to capture the relevant details of her action. Must we accordingly disqualify these cases from the domain of action proper and classify them as mere behaviour instead? This would be disappointing, as it would lead to a significant reduction of the range of intentional action on the part of regular agents. Ingerslev thus sets herself the following task: “We want to understand how the involuntary aspects of habit are more than just impersonal happenings, since it goes on in my name”⁴³.

Ingerslev’s solution to secure the status of these habitual actions as intentional is instructive for the view advanced here. She invokes the notion of an agent’s self-understanding, but crucially points to the necessity of construing the self-understanding *diachronically* instead of merely *synchronically*. If by self-understanding we just mean whatever the agent is capable of avowing *right now*, at the very moment of performing her habitual action, we would have thinned out the idea of a self-understanding beyond recognition. Ingerslev accordingly objects to what she calls the *Synchronicity Assumption*, which she thinks is implicit in a lot of work in contemporary action theory: “The full self-understanding afforded by an agent’s awareness of her habitual behavior is immediately available by the time of the awareness”⁴⁴. Against this assumption, she

⁴³ Ingerslev, ‘On the Role of Habit’, 490.

⁴⁴ *Ibid.*, 483.

reminds her readers of the – I think rather obvious – reasons for why a self-understanding is a temporally expanded affair that will not be transparent to the agent in all its aspects and layers at any given moment in time. For one, a self-understanding is built up over time; it is reflective of all sorts of experiences that the agent has made throughout her development as a subject and is likely beset by layers of memory that are not readily available to reflection (body memory is a case in point, which Ingerslev discusses extensively). Accordingly, a person's past remains effective in shaping current orientations and intentions, without being retrievable at will in all its relevant aspects by the reflectively conscious agent.

Now, the key point is that cases of habitual actions that display this kind of intransparency can be seen as guided, in part, by such (temporarily) inaccessible aspects of the agent's self-understanding:

Is it the case that something beyond one's formal or friendly greeting behavior can influence one's action while still being part of the reason for my action? If we grant that past experiences can form part of one's reasons for responding in certain unacknowledged ways, we must answer yes to this question. Something beyond the intention that I am non-observationally aware of can influence me and form part of my reason to act in a certain way; thereby exposing me to gaps in my agential motivation while at the same time leaving it open for me to self-surprisingly becoming aware of what I am doing.⁴⁵

So the upshot is a temporally extended view of an agential self-understanding that expands the field of action well into the domain of momentarily intransparent habitual actions. If the specific manner of such actions is brought to the agent's attention, this is a situation where the agent is confronted with a question about herself. There is something in my current doing that eludes me, and this poses an issue for my ongoing

⁴⁵ Ingerslev, 'On the Role of Habit', 489–90.

self-interpretation, which leads me to inquire into my formative past in an attempt to retrieve my reasons for acting specifically in the way I do. Sometimes, no doubt, only very intense and enduring efforts will do the job, such as forms of therapy.⁴⁶

There is one important difference in emphasis between Ingerslev's presentation of her view and what I would countenance as the expanded version of the active mind. Ingerslev has a tendency to construe the situation of self-interpretation prompted by habitual action as one of *discovering* or *rediscovering* who one already is: "In appropriating my own habitual actions as mine, I assume ownership for what has been done by re-committing to who I am as a person. That is, I re-commit to being myself"⁴⁷. Thus, her account seems to reckon not only with a basic stability in a person's self-understanding, but also assumes by and large agreement between the "present" reflective self and whatever the scrutiny of one's habitual actions brings to light. She foregrounds situations of re-commitment, that is, cases where the agent explicitly endorses what they were apparently holding all along. I think this puts the emphasis where it doesn't quite belong. Given the point made above about the unavoidable situation of choice in which the agent finds herself with regard to everything that is "on her mind", it seems more apt to frame enlightenment about opaque aspects of one's self-understanding as cases where one is confronted with a practical question: This is what I seem to have been holding all along, all well and good; is it something that *I should* endorse? Thus, the situation we deal with here is one in which I find myself confronted

⁴⁶ Like Moran, Ingerslev brings up the theme of psychotherapy in relation with her proposal on the elucidation of an agent's self-understanding. Both invoke work by Jonathan Lear, who himself has engaged in instructive ways with Moran's account on these issues. See Lear, 'Avowal and Unfreedom' and Moran, 'Replies'.

⁴⁷ Ingerslev, 'On the Role of Habit', 493. To be fair, Ingerslev does acknowledge the possibility of transformation, but she does so only briefly at the end of her paper (see 495).

with the choice to *either* re-commit *or* vow to change my orientation, depending on my current reflective assessment. Probing into previously occluded dimensions of my self-understanding is not an empirical discovery of something that is given no matter what, but a quest to make up my mind about what it is that I should endorse in light of what I seem to have been endorsing all along. Self-discovery in the first-personal sense is inextricable from renewed reflection on what I *have reasons* to be committed to.

Of course, it might turn out to be hard to change my deeper orientations by means of such renewed reflection, so that all I in fact “take away” from such self-assessments is a piece of factual knowledge about my putative orientations. But the point here – as with the active mind in general – is not that the *success* of responsible self-transformation lies in my power. The point is that I can either endorse or withhold endorsement from “who I am”, and so my self-discovery does not end in a brute given, but, once again, in a situation of choice – a choice which is mine and only mine to make.

4. Conclusion

I have done three things in this paper. First, I stated, in a deliberately strong and streamlined way, the gist of the active mind position, which revolves around the claim that human mental life is at heart a matter of self-constitutive agency. In particular, I have discussed, in a spirit of agreement, considerations by Sartre and Moran on how the self-conscious agent cannot be omitted from the characterization of any mental process or impulse whatsoever and that a situation of choice is inevitable for the bearer of mental states. That is, agency is the fulcrum of the mind as we know it.

Second, I have confronted the active mind position with a challenge based on the assumption that affectivity speaks against the notion that agency is constitutively in play in all major categories of human mindedness. My response drew on considerations

in favour of the agentic core of emotional comportment and on the idea of a practical self-understanding as a kind of mediator between what agents do intentionally and what they undergo or suffer passively. It is the lived self-understanding that brings such impingements into the ambit of the agent's reflective choice, and thus into the scope of her activity.

Third, I have discussed a specific subclass of habitual actions where the guiding intention is inaccessible to the agent at the time of his acting as a further potential objection to the active mind. Drawing on work by Line Ingerslev, I have endorsed her view on how a temporally extended, diachronic conception of a person's self-understanding can widen the scope of intentional action to encompass reasons outside the present ambit of reflective consciousness. I have diverged somewhat from the main thrust of Ingerslev's account, however, in that I have emphasized the principal position of agency from which a person comes to acknowledge her initially opaque reasons, so that these putative reasons, while already effective in habitual conduct, become a matter for renewed endorsement or rebuttal. Acknowledging some previously opaque orientation or inclination as mine does not amount to endorsement or re-commitment, but poses the question of whether *I shall* endorse it – with the acute possibility of finding fault with myself in this regard and thus withholding endorsement upon reflection, committing to self-transformation instead.

With these considerations, I have attempted to underline the expansiveness of the active mind position, showcasing its wide scope against sceptical injunctions. Now, while not nothing, these results should certainly not be overrated. Arguably, the points developed here have only brought out, in somewhat more explicit terms, what is implicit in the basic operations of a first-person perspective. There is a further and highly significant dimension to the task of expanding the active mind. Besides

embedding self-constitutive agency in the thicket of a lived self-understanding, which might be cautiously called a “subject-immanent” expansion, the active mind also has to be embedded within its wider social and cultural surroundings. Forms of life, social institutions, customs and practices condition, all frame and scaffold the active mind. They endow it with a range of possibilities, but also potentially disown the agent and put limits on agentic autonomy. The philosophical endeavour that goes by the title “Expanding the Active Mind” has barely begun; its full execution would have to include the task, unaddressed in these pages, of showing how – and under what conditions – the social and material environment can be considered to be within the range of practical endorsement on the part of the agent. This is the Hegelian or, probably more so, Marxian vision of seeing the world at large – excluding fellow human beings – as an extension of the capacities of a free and rational agent; or in Marx’s notable words, as “the inorganic body of man”⁴⁸. This is also where the point of contact lies between the present endeavour and much productive recent work on 4E cognition and the extended mind in particular. I have not engaged with these more vexing issues here, which one might claim pose the truly challenging tasks in the quest to expand the active mind.

Acknowledgments

I thank the two anonymous reviewers for their generous, insightful and spirited comments and suggestions. I also thank Laura Candiotta for her thoughtful scepticism expressed in response to an earlier version. Moreover, thanks is due to Kathy Ran for her terrific language editing. Work on this paper was partly facilitated through a grant

⁴⁸ Marx, ‘Economic and Philosophical Manuscripts 1844’, 139, as quoted approvingly in McDowell, *Mind and World*, 118.

by the Deutsche Forschungsgemeinschaft (DFG) as part of the Collaborative Research Center 1171 *Affective Societies* at Freie Universität Berlin.

References

- Anscombe, G. Elizabeth M. *Intention*. Cambridge, MA: Harvard University Press, 1963.
- Boyle, Matthew. ‘Making up Your Mind’ and the Activity of Reason’. *Philosopher’s Imprint* 11, no. 17 (2011): 1–24.
- Boyle, Matthew. ‘Active Belief’. *Canadian Journal of Philosophy* 39, sup. 1 (2009): 119–147.
- De Jaegher, Hanne and Ezequiel Di Paolo. ‘Participatory Sense-Making’. *Phenomenology and the Cognitive Sciences* 6 (2007), 485–507.
- Gallagher, Shaun. ‘The Socially Extended Mind’. *Cognitive Systems* 25–26 (2013): 4–12.
- Goldie, Peter. *The Emotions*. Oxford: Oxford University Press, 2000.
- Goldie, Peter. ‘Emotion, Reason, and Virtue’. In *Emotion, Evolution, and Rationality*, ed. Dylan Evans and Pierre Cruse, 249–267. Oxford: Oxford University Press, 2004.
- Goldie, Peter. *The Mess Inside: Narrative, Emotion, & the Mind*. Oxford: Oxford University Press, 2012.
- Griffiths, Paul E., and Andrea Scarantino. ‘Emotions in the wild’. In *The Cambridge Handbook of Situated Cognition*, ed. Philip Robbins and Murat Aydede, 437–453. Cambridge, UK: Cambridge University Press, 2008.
- Hampshire, Stuart. *Freedom of the Individual*. Princeton, NJ: Princeton University Press, 1975.
- Heidegger, Martin. *Sein und Zeit*. Tübingen: Max Niemeyer, 1927.
- Helm, Bennett W. *Emotional Reason: Deliberation, Motivation, and the Nature of Value*. Cambridge: Cambridge University Press, 2001.
- Hieronymi, Pamela. ‘Responsibility for Believing’. *Synthese* 161 (2008): 357–373.
- Ingerslev, Line Ryberg. ‘On the Role of Habit for Self-Understanding’. *Phenomenology and the Cognitive Sciences* 19 (2020): 481–497.
- Lear, Jonathan. ‘Avowal and Unfreedom’. *Philosophy and Phenomenological Research* LXIX, no. 2 (2004): 448–454.

- Noë, Alva. *Action in Perception*. Cambridge, MA: MIT Press, 2004.
- Marx, Karl. 'Economic and Philosophical Manuscripts 1844'. In *Early Texts*. Trans. David McLellan. Oxford: Basil Blackwell, 1972.
- McDowell, John. *Mind and World*. Cambridge, MA: Harvard University Press, 1994.
- Moran, Richard. *Authority and Estrangement: An Essay on Self-Knowledge*. Cambridge, MA: Harvard University Press, 2001.
- Moran, Richard. 'Replies to Heal, Reginster, Wilson, and Lear'. *Philosophy and Phenomenological Research* LXIX, no. 2 (2004): 455–472.
- Müller, Jean Moritz. 'The Spontaneity of Emotion'. *European Journal of Philosophy* 2021 <https://doi.org/10.1111/ejop.12625> (Early View).
- Sartre, Jean-Paul. *Being and Nothingness*. Trans. Hazel Barnes. New York: Washington Square Press, 1956.
- Sartre, Jean-Paul. *Sketch for a Theory of the Emotions*. Trans. Philip Mairet. London and New York: Routledge Classics, 2002.
- Schear, Joseph K. *Mind, Reason, and Being-in-the-World*, ed. Joseph Schear. New York: Routledge, 2013.
- Slaby, Jan and Philipp Wüschner. 'Emotion and Agency'. In *Emotion and Value*, ed. Sabine Roeser and Cain Todd, 212–228. Oxford: Oxford University Press, 2014.
- Solomon, Robert C. *The Passions: Emotions and the Meaning of Life*. Indianapolis: Hackett, 1993.
- Taylor, Charles. 'Self-interpreting animals'. *Human Agency and Language. Philosophical Papers Vol. I*, 45–76. Cambridge: Cambridge University Press, 1985.
- Thompson, Michael. 'Anscombe's *Intention* and Practical Knowledge'. In *Essays on Anscombe's Intention*, ed. Anton Ford, Jennifer Hornsby and Frederick Stoutland, 198–210. Cambridge, MA: Harvard University Press, 2011.