Pre-print. Please cite published version: Ballesteros, V., Fernández-Castro, V. & Núñez de Prado-Gordillo, M. (2025). It Doesn't Feel Like Myself: A Mindshaping View on Self-Illness Ambiguity. *Routledge Handbook of Mindshaping* (eds. T. Zawidzki & R. Tison), pp. 473-485. Routledge.

It Doesn't Feel Like Myself:
A Mindshaping View on Self-Illness Ambiguity

Virginia Ballesteros[1], Víctor Fernández-Castro[2] & Miguel Núñez de Prado-Gordillo[3]

[1]Department of Philosophy, University of Valencia

[2]Department of Philosophy I, University of Granada

[3]Department of Philosophy, University of Rijeka

Notes on contributors

Virginia Ballesteros, PhD, is Assistant Professor in the Department of Philosophy at the University of Valencia. Her research focuses on the philosophy of psychiatry, the philosophy of psychedelics, and their intersection.

Víctor Fernández-Castro, PhD, is Associate Professor in the Department of Philosophy I at the University of Granada. His research focuses on the philosophy of cognitive sciences and the philosophy of psychiatry.

Miguel Núñez de Prado-Gordillo, PhD, is a postdoctoral researcher in the Department of Philosophy at the University of Rijeka. His research lies at the intersection of the philosophy of mind, the philosophy of psychiatry and clinical psychology.

Acknowledgements: Virginia Ballesteros was supported by the Government of the Valencian Community (grant CIGE/2023/008) and the Spanish Ministry of Science and FEDER, UE (grants PID2021-126826NA-I00, PID2023-151949NA-I00 and PID2023-151071NB-I00). Víctor Fernández Castro was supported by the Spanish Ministry of Science and FEDER, UE (grant PID2021-126826NA-I00). Miguel Núñez de Prado-Gordillo was supported by the Croatian Science Foundation (grant HRZZ-IP-2022-10-1788), the Dutch Research Council (grant VI.VIDI.195.116), and the Spanish Ministry of Science and FEDER, UE (grant PID2021-126826NA-I00).

Abstract: The phenomenon of *self-illness ambiguity*, concerning whether an agent's actions and experiences reflect some psychiatric condition or their "authentic self"— or both —has attracted increasing attention in the last years. Different views of self-illness ambiguity range from realism about the self-illness divide, which views it as a pre-existing border to be *discovered,* to constructivism, which claims that the boundary is *constructed*. Both face important challenges; the former related to the undue reification of self and illness, the latter to the risk of becoming a subjectivist account of the distinction that neglects how embodiment and others' perspectives impose constraints on our self-narrative possibilities. Our contribution explores how mindshaping approaches offer key conceptual resources to retain the best of both views. Specifically, the mindshaping perspective yields a constructivist view of the self-illness divide according to which the distinction does not capture pre-existing facts independent from our (self-)interpretation practices; however, this perspective also views mental concepts and narratives as strongly dependent on social norms and dynamics that, while contestable, cannot be changed at will. The resulting picture avoids an "anything goes" view of the self-illness distinction, retaining the realist insight that not only self-creation, but also self-discovery is crucial for addressing self-illness ambiguities.

Keywords: Philosophy of Psychiatry; Self-Mindshaping; Self-Management; Self-Narrative

1. Introduction

In her contribution to *Mad Pride: A Celebration of Mad Culture,* Louise C. discusses with mental health professionals what it means to be mad for her:

> My thoughts on being mad are thus: people tell me I am mad. I have been certified as mad. I take tablets for mad people. You told me that I'm mad. […]. Hey, look, mad or not I am a person, a human being, and being dumped by my husband of ten years has broken my heart. I do crazy things, you know. I take various quantities of class B's. I keep washing my hands. I talk to animals. Some days, most days, I hear voices, saying 'Louise, you are mad!' These voices appear to be coming from a gremlin in my brain. But you told me, it's my voice, my thoughts. It's my personality, you said. Try to have a more positive sense of yourself, you told me. Have

> higher self-esteem, you told me. Well, I am trying. (Curtis et al., 2000, 29–30)

Similarly, Mad Pride founder Pete Shaughnessy recounts his involuntary hospitalization as follows: “The freedom I had found for six short weeks was now over. Everything I had felt, loved, over that time was, I was told, ‘all a delusion.’ Time to conform.” (Curtis et al., 2000, 20). These examples illustrate the difficulties encountered by people who experience mental health problems — or others around them — when trying to establish whether their actions, thoughts, and experiences reflect their *authentic* selves (i.e., who they *truly* are) or some underlying psychiatric condition. Developed in the 1990s by social scientists working on the relation between self, illness, and medication (Karp, 1992), the notion of ‘self-illness ambiguity’ aims to capture this puzzling experience (Sadler, 2007). Recent attempts to clarify the phenomenon highlight its huge complexity. To begin with, one may question whether self-illness ambiguity (hereafter, SIA) is something necessarily detrimental; in fact, some argue that it may be a tool for promoting positive changes (Drożdżowicz, 2023). Furthermore, we might also resist an either-or formulation of the phenomenon: as the examples above and others from the mad and neurodiversity literature illustrate, we may understand some of our actions and experiences as reflective of conditions we can eventually identify with (Jeppsson, 2022; Chapman, 2023).

SIA might appear at an experiential (sometimes also called ‘phenomenological’ or ‘unreflective’) or conceptual (‘reflective’) level (Bortolan, 2022; De Haan, 2023; Dings and Glas, 2020; McConnell and Golova, 2023). At an experiential level, phenomena characteristic of certain conditions may feel alien or be experienced in different degrees of *mineness*, as Louise’s voice hearing suggests. But even when behaviors and experiences feel fully one’s own pre-reflectively, doubts may still arise about how best to conceptualize them and balance competing person-focused vs. illness-focused *narratives*. Do all the “crazy” things that Louise does and experiences reflect who she really is, her natural way of reacting to her husband abandoning her? Or should these be best understood as Pete’s psychiatrists understand his – as the result of mere delusions?

This chapter focuses on the conceptual aspect; specifically, on how different ‘self’ and ‘illness’ narratives may clash when trying to make sense of

ourselves and how best to understand these conflicts. We argue that a mindshaping approach to folk psychology (McGeer, 2007; 2021; Zawidzki, 2008; 2013) offers key conceptual resources to understand the role of competing (self-)narratives in addressing SIA. Specifically, a mindshaping approach can help us articulate the *constructivist* view that the self-illness divide is at least partially dependent on our interpretative practices and the various norms, expectations, and practical interests associated with them – while also preserving the *realist* intuition that not "anything goes" when it comes to distinguishing self from illness. Our concepts and narratives are dependent on social norms and dynamics that, while contestable, cannot be changed at will, nor can we therefore freely decide where to draw self-illness boundaries.

The structure of the chapter is as follows. Section 2 addresses the phenomenon of SIA and the contrast between realist and constructivist perspectives. Section 3 examines how the mindshaping framework offers key tools for developing the constructivist perspective, focusing on its view of agency and the role of competing narratives about an agent in self-mindshaping. Section 4 explores how the mindshaping perspective can still accommodate realist insights concerning the relative role of self-construction and self-discovery in addressing SIA. Finally, the conclusion discusses a possible limitation of our approach and suggests directions for future research.

2. Realism, Constructivism, and Self-Illness Ambiguity

The phenomenon of self-illness ambiguity—or, more generally, the ambiguity over how best to understand an agent's relation between their problems and themselves (De Haan, 2023)—has attracted increasing attention in recent years, following a renewed interest in the impact of psychiatric experiences and diagnoses on the configuration of the self and self-related phenomena, i.e., authenticity, self-discovery, or self-management (De Haan, 2023; Dings and Glas, 2020; Strijbos and Slors, 2020; Tekin, 2019). Moreover, this ambiguity touches on fundamental philosophical questions about the nature of the self, (mental) illness, and self-knowledge (Dings and Glas, 2020).

The self-illness divide has been approached from various stances, ranging from *strong realism* to strong *constructivism* (Jeppsson, 2022; De Haan, 2023; McConnell and Golova, 2023; Wilkinson, 2023). The core tenet of the former is that clear boundaries exist between self and illness; disambiguating whether

certain behaviors, emotions, or thoughts can be properly attributed to one or the other involves discovering relevant facts, whether from a first- or third-person perspective. Constructivist stances, on the other hand, challenge the idea of a pre-existing boundary: the self-illness divide is not ready-made, but constructed. No fact independent from our interpretive practices could resolve the ambiguity. Depending on how one spells out such theses, realist and constructivist stances might appear as totally antagonistic views. Yet, we believe that both stances have valuable insights, and that a mindshaping approach allows us to articulate a constructivist stance on the self-illness divide while accommodating certain realist intuitions. First, let's explore each stance, beginning with realism and then moving to constructivism.

Overt defenses of a realist stance on the self-illness divide are rare in the literature on self-illness ambiguity. However, realist assumptions are traceable in some texts. One example could be John Sadler's (2007) characterization of "good clinicians" as "trained to engage the patient in regarding the self as a third-person-like object of observation: the patient, in observing the self, separates out facets of the self and begins to differentiate self from illness" (p. 121). This passage suggests there exists a boundary between self and illness, and that learning about the relevant matters of fact is at least part of what is necessary to disambiguate them, by observing the self as a "third-person-like object".

Realist assumptions about the self-illness divide are more explicitly reflected in the biomedical model of mental disorders (Dings and Glas, 2020; De Haan, 2023). For years, biopsychiatry has promoted a reifying view of mental illnesses, treating them as natural kinds — discrete pathological entities of a biological nature, clearly demarcated both from normalcy and from each other (Haslam, 2000). For instance, consider how Peter Kramer, a biologically-oriented psychiatrist, describes his patient Margaret, who suffered from depression:

> The emotions … were externally imposed, by the onset and then the persistence of an illness. … The guilt – that was not Margaret, nor the indifference either … A disease had robbed her of feelings that were properly hers and imposed alternative ones. Margaret was the one who experienced those depressive feelings and reported them in psychotherapy; but they did not arise from, or had only weak roots in, the psychology that was hers in health. (Kramer, 2005, 25)

Here, mental illnesses are regarded as being as foreign to the self as infections are to the body. The "real" self is understood to be free of mental symptoms, and we should be aware of telling them apart from who we really are (see Karp, 1992; Dings, 2020; Jeppsson, 2022; De Haan, 2023).

Realism has several virtues. Firstly, it can be liberating by allowing individuals to externalize negative aspects (behaviors, emotions, beliefs) they do not wish to identify with. A well-known example is postpartum depression: for many women, the realist, biomedical narrative of postpartum depression has been freeing, helping them relinquish feelings of guilt and inadequacy by recognizing it as a medical condition (Wardrope, 2015). In this view, the depressed mood and other symptoms following childbirth do not reflect who they really are and want to be, and thus they should not be blamed for their mental disorder[1]. Biomedical narratives can thus be seen as a means to do justice to attributions of reduced moral and/or legal responsibility to individuals suffering from mental illness—indeed, this seems to be the idea behind the programs and campaigns that seek to fight stigma against mental health populations (Jorm et al., 2000)[2]. In some cases, mental illness may serve as a mitigating or exonerating factor for responsibility, as related behaviors may be difficult or impossible to control.

Furthermore, realism aligns with the subjective experience of (self-)estrangement that often accompanies mental symptoms, as these symptoms tend to disrupt our everyday (self-)experience and (self-)narratives. It also fits with the sense of “being ourselves again” that usually follows symptom resolution. As Margaret expressed upon recovering from depression: “I said to Gregory, ‘I’m better.’ How could he know what I mean? I kept repeating, ‘No, look. *I’m* back.’” (Kramer, 2005, 15). Moreover, realism accommodates well the

[1] An underlying assumption here is that what can be attributed to our "true" selves is what we can be held responsible for. This reflects a significant aspect of the self, often emphasized across the realism-constructivism spectrum and in broader philosophical debates. However, equating the self solely with what we are responsible seems reductionistic. As it has been argued (Dings & Glas, 2020) the self is multidimensional, including aspects over which we have limited control or agency. As such, the relationship between illness attribution and blame/responsibility mitigation is likely more complex and nuanced than suggested by biomedical narratives.

[2] While biomedical explanations may reduce the tendency to blame individuals with mental illness, they do not necessarily improve other negative attitudes toward psychiatric populations and are often associated with pessimistic views about recovery (Kvaale, Haslam, and Gottdiener 2013).

possibility of *self-discovery*, – i.e., that we may come to find out how we *really* are – vs. how we take ourselves to be. It is common to reflect, “I’ve learned something new about myself” or realize, “I’m not the person I thought I was,” – which also indicates that self-discovery involves the possibility of being wrong.

Despite its strengths, realism faces significant challenges, particularly in identifying the boundaries between self and illness. One major issue arises when realism involves a reified view of self and illness, as seen in biomedical approaches. Regarding mental illness as a discrete entity contradicts scientific findings (Hyman, 2010; Kendell and Jablensky, 2003). At best, evidence suggests that normalcy and pathology lie on a continuum, without clear-cut boundaries between "normal" mental functioning and mental symptoms. Therefore, if we identify the self with "normal" functioning, the distinction between self and illness becomes unclear. Admittedly, one could still argue that selves and illnesses are distinct but have fuzzy boundaries. However, this weakens the realist position. If the boundary is inherently unclear, there is no already-given demarcation to track when distinguishing self from illness.

Moreover, even if we assume there is a clear boundary between self and illness, significant epistemological challenges remain. As the broader issue of self-ambiguity (Dings and Glas, 2020) and common self-knowledge difficulties show, identifying which behaviors and experiences truly belong to us is not a straightforward task. In the context of SIA, Sophia Jeppsson (2022) notes that we can get stuck in a never-ending questioning: “Every time I think the thought ‘maybe my previous thought was just a mental illness symptom’, this *new* thought can be doubted the same way” (p. 297). Jeppsson argues that even if a boundary exists, trying to *discover* it often leads to futile, never-ending doubt. Instead, she suggests that our best approach is to deliberate on what we want to identify with and *construct* our own narrative accordingly.

The rejection of a ready-made boundary between self and illness is a core tenet of constructivism. As with realism, this idea can be spelled out in different ways. One of the most straightforward constructivist approaches is that of Jeppsson (2022), who addresses the self-illness divide without reifying either (see also Rego, 2004; Wilkinson, 2023 for similar views). According to Jeppsson (2002), self and illness are not entities with predefined boundaries susceptible to being discovered; rather, they are constructed as we “ponder and deliberate —with more or less help from clinicians, friends, and family—about what to

identify with" (p. 309). Whether something reflects one's "true self," one's psychiatric condition, or both, is, to some extent, something that we may decide. In our narratives, we have some leeway to decide what to identify with and what to externalize into illness, by interpreting our life story, actions, reactions, desires, beliefs, etc., in a way that is intelligible and meaningful. While Jeppsson does not claim that these constructions are without limits, she places significant emphasis on individual deliberation and decision.

One *prima facie* virtue of Jeppsson's approach is that it avoids the (presumably fruitless) search for facts to resolve the ambiguity. Instead, it can empower individuals by giving them the ability to determine how to conceptualize themselves. We can choose which narrative to endorse through deliberation about what is most beneficial for us. As Jeppsson (2022) states, "since there is no fact of the matter—no border between self and illness which is just there, independently of where we would like to put it—we should evaluate constructivist solutions to the self-illness ambiguity by ethical standards" (p. 305). Another strength of her account is that it promotes a radically person-centered approach to self-disambiguation in therapeutic contexts. Instead of rigidly classifying behaviors and experiences as illness- or self-based once and for all, as traditional diagnostics often do, her approach allows for a more flexible, context-sensitive evaluation of when to internalize or externalize particular traits or behaviors[3].

However, some critics worry that this approach could lead to a subjectivist, "anything goes" view of the self-illness distinction, where individuals unrestrictedly choose between self- or illness-narratives (Dings and De Bruin, 2022; Tekin, 2022; De Haan, 2023; McConnell and Golova, 2023). This raises questions about the criteria for making such decisions and how to maintain a sense of authenticity. Decision-based constructivism faces challenges in specifying how narrative construction happens, who participates in it, and how to balance deliberate self-construction with the potential for error in self-

[3] Realism, however, also seems to be supported by practical reasons, e.g., when biomedical narratives are promoted to reduce stigma. However, this externalizing strategy is also available for constructivism. The key difference between realism and Jeppsson's constructivism lies in what justifies endorsing self- vs. illness-based narratives; whereas for realism this depends on — presumably universal — factual discoveries about the mind, constructivism suggests that different narratives are primarily justified by their practical consequences for particular individuals in particular context.

narratives. McConnell & Golova (2023) address this by distinguishing between "shapeable" aspects of the self — such as values, plans, and self-narratives — and more "essential" aspects that configure embodied and pre-reflective (self-)experiences, like genetic and other deep-seated dispositions. Addressing SIA requires an ongoing balance between self-shaping — which allows for explicit deliberation and decision-making — and self-acceptance, which at least partly involves a process of self-discovery.

Furthermore, social actors play an irreducible role in the process of (self-)narration. Our narratives are shaped and encountered by others who may accept, question, or reject them, making self-narratives often co-authored with significant others, mental health professionals, etc. (McConnell and Golova, 2023; De Haan, 2020; 2023; Tekin, 2022). This friction is not just common but necessary for properly grounded self-narratives. Sometimes, others — especially those close to us — may be better positioned to determine whether an experience or behavior should be externalized or internalized, as they have an intimate, second-personal perspective on our mental states (Borgoni & Pinedo, forthcoming). Narrative co-authoring, therefore, involves a balance between first- and third-person (or second-person) authority, with others' perspectives sometimes placing constraints on our self-narrative possibilities (Strijbos and De Bruin, 2015; De Haan, 2020; 2023). This is why introspection may not be the best route for discovering such constraints; instead, learning from our interactions with the world and others can provide valuable self-knowledge[4]. For instance, a person with depression might create a narrative where their need for social isolation is regarded as external to their true self, as a symptom of their newly developed mental disorder. This might be challenged by a parent, who could point out that their desire for isolation has always been part of their

[4] Jeppsson (2022) highlights that the same issues with introspective attempts to distinguish self from illness also arise when deciding which narrative to adopt. Mental difficulties can obscure or bias our ability to properly deliberate and ponder which narratives to endorse (De Haan, 2020; Bortolan, 2022; Tekin, 2022). However, Jeppsson notes a key difference between introspective and deliberative strategies in how they address doubt. Introspectivists seek an indisputable inner fact to justify whether certain aspect belongs to the self or illness, but since I may always doubt whether my own introspection is affected by illness, this spirals into a "never-ending questioning". Constructivists like Jeppsson, on the other hand, argue that decisions can be more or less justified based on how the deliberative process unfolds and how well it meets practical demands. While decisions may be revisited, their adequacy doesn't depend on accessing some elusive, indisputable fact. The process of questioning remains open-ended but not intrinsically never-ending.

personality, as they often spent long hours playing alone as a child and seemed to enjoy it. Similarly, self-narrative accounts of identity have imposed certain constraints, such as Marya Schechtman's (2007) "reality constraint," which demands consistency with generally accepted understandings of reality, and the "articulation constraint," which requires individuals to coherently explain their narrative when appropriate and address legitimate questions concerning their reasons for acting and reacting in certain ways (see also Dings and De Bruin, 2022; Tekin, 2022).

In sum, although constructivism about SIA provides several theoretical and practical advantages over realist accounts, placing too much emphasis on individual deliberation in the construction process fails to account for how embodiment and social dynamics constrain our self-interpretation (Dings and De Bruin, 2022; Tekin, 2022). A proper constructivist account must balance the roles of self-creation and self-discovery in addressing ambiguity. In the following sections, we outline how self-mindshaping accounts of metacognition, self-interpretation, and agency provide a sound framework for this balance. Specifically, we argue that the mindshaping perspective avoids reifying the self, illness, and their boundaries while resisting an overly subjective or "anything goes" account of self-construction processes in setting the divide.

3. Self-Mindshaping and Self-Construction

To understand how the mindshaping view can capture the strengths of constructivism, we need to examine key aspects of mindshaping in relation to the first-person perspective. Mindshaping views (McGeer, 2007; 2021; Zawidzki, 2008; 2013) have been proposed to address fundamental issues about socio-cognitive abilities and practices. The core idea is that these practices — such as predicting, explaining, and justifying behavior — rely on agents' ability to act and think in ways that align with the norms and values of their social environment. The main function of social cognition is to form and acquire norms and patterns of behavior that regulate our social interactions and facilitate cooperation by making behavior more transparent to others. Thus, humans predict others' behavior because they regulate their own behavior according to socially mediated normative expectations. Natural selection likely favored the development of mechanisms that enable humans to learn and internalize these social norms, such as automatic conformity, imitation, and natural pedagogy.

Moreover, complex social dynamics require regulative practices to maintain normative structures, such as monitoring and correcting counternormative behavior by expressing reactive attitudes or asking for justifications. A crucial aspect of our socio-cognitive abilities is using these regulative strategies to help others conform to the appropriate norms and values. In fact, a crucial aspect of our socio-cognitive capacities is using these regulative strategies to make others conform to the appropriate norms and values.

The idea that social dynamics shape our minds to make us more predictable is closely linked to how these dynamics also underpin our agency. Regulative strategies, like expressing reactive attitudes such as resentment, indignation, or gratitude, express our consideration of others as accountable agents capable of meeting a normative ideal. Now, reactive attitudes usually do their job correctly when the recipient of the attitude acknowledges that their behavior has been subjected to normative review.[5] This creates a prospective element to actions like blaming or reproaching, where the recipient recognizes their behavior is subject to review and may adjust it in future contexts. When an agent reacts to the frustration of a norm, for example, by reproaching or expressing disapproval, the recipient usually understands that their behavior is being and can co-react by excusing or modifying their action in subsequent contexts.[6] These co-reactions are meaningful because reactive attitudes urge the norm-deviant to evaluate their own behavior and mental states in light of the expectations generated by social norms and values. We become socially competent beings when we respond to the mindshaping influences of others by reviewing our behavior, self-regulating future actions, or by providing reasons that justify our actions and reactions.

A key aspect of developing and exercising our agency is, thus, cultivating self-oriented psychological capacities aimed at interpreting, knowing,

[5] Of course, in particular instances, reactive attitudes may induce norm-conformity even if the target does not recognize the normative evaluation behind the attitude. For instance, someone might conform to norms simply to avoid social punishment, without understanding the moral significance of their actions. However, recognizing the normative review reflected by the reactive attitudes is necessary for understanding the agent's norm-conforming behavior as a genuine instance of rule-following rather than as a mere conditioned response.

[6] This may involve contesting the norm itself or the expectations typically associated with it. As McGeer (2008; 2015) points out, such acts of normative contestation represent one way in which an agent can demonstrate reason-responsiveness or sensitivity to the normative demands of others (see Section 4).

understanding, evaluating, and regulating our own cognition and behavior (McGeer, 2007; Strijbos and De Bruin, 2015; Zawidzki, 2016; Fernández-Castro and Martínez-Manrique, 2021). These capacities include regulating ourselves according to self-narratives, self-concepts, stereotypes, or individual avowed decisions. For example, Zawidzki (2016) argues that self-conceptualization plays a crucial role in self-interpretation. Conceptualizing ourselves as 'parents,' 'mates,' 'doctors,' or 'teachers' involves a commitment to act in line with the expected behaviors and cognitive profiles associated with these roles. These self-concepts function as social models that provide behavioral and cognitive prescriptions reinforced not just by social pressures but by affective states like social emotions that encourage adherence to these expectations (Zawidzki, 2016, p. 484). Another important self-mindshaping capacity lies in public self-ascriptions of mental states (e.g., beliefs), derived from individual reflection on the external world, whose content our behavior should abide by. Finally, narratives (Hutto, 2008; Zawidzki, 2013; Fernández-Castro and Martínez-Manrique, 2021) also serve as external models that help us interpret others' actions, offering coherent explanations and contextualization of their behavior; but also, they can provide a framework for understanding our own actions and generate expectations on how to respond to the environment. As Fernández-Castro and Martínez Manrique (2021) put it:

> our behavior makes sense for us when it is coherent with our self-narrative, and we regulate it according to the expectations this narrative sets up, responding to others' behavior, personality, and context. We are able to bring our personal narrative into line with our deeds and our deeds into line with our narratives. (p. 154)

Now, self-oriented mindshaping capacities align with a fundamental claim of constructivism. Advocates of the mindshaping view argue that regulative strategies do not necessarily require accessing pre-existing mental states as understood by classical approaches to metacognition. For example, following Moran's (2001) argument for the transparency of self-knowledge, McGeer (2008; 2015) claims that many belief self-ascriptions result not from accessing our own inner psychological states via introspection, but from reflecting on how the world is from our perspective and pondering what to believe in light of the available evidence (see also Strijbos and De Bruin, 2015). Similarly, Zawidzki (2016) claims that "self-interpretation is not, primarily, in the business of

discovering truths about independently constituted mental states. Rather, it helps constitute the mental states of self-attributors by shaping them to fulfill self-attributions" (p. 480). Thus, self-mindshaping capacities are better understood as *self-creating* devices, i.e., as controlling cognitive and behavioral trajectories by setting certain constraints and prescriptions, ultimately giving form to the behavioral and mental states that figure in such trajectories (Fernández-Castro and Martínez-Manrique, 2021, pp. 146–47).

It certainly does not follow from the fact that our metacognitive capacities are not based on state detection or that there are no facts involved in determining what an agent believes or desires. When an agent avows a particular mental state, part of their agency lies in recognizing what their *actual* cognitive, emotional, and behavioral trajectories are, and actively regulating them according to social norms and scripts linked with such avowal. However, while recognizing such factual elements, the mindshaping perspective remains consistent with constructivism regarding SIA in its emphasis on the self-creating or self-shaping role of self-interpretation. Might the recognition of the role of self-mindshaping in SIA answer the challenges that constructivism must address? Could this version of constructivism account for the intuitions behind realism?

4. Self-Mindshaping and Self-Discovery

From the mindshaping perspective, our capacities for self-narration and self-conceptualization do not require accessing discrete mental phenomena; in this sense, the view seems to be a natural ally of constructivism and its rejection of reified accounts of mind (Fenici and Zawidzki, 2021). At the same time, however, it provides us with tools to understand the realist intuition that, in a substantive sense, there are aspects of our mental life that we *discover*, and that this discovery is an important part of the process someone goes through when constructing self-illness boundaries.

From the mindshaping point of view, our deliberative powers are fragile, and are often easily hijacked by our more 'subversive inclinations' (McGeer, 2008; Strijbos and De Bruin, 2015). Indeed, as McGeer points out, social and moral development isn't solely focused on improving our ability for deliberate avowal, but also on learning how to manage our tendencies or proclivities. These inclinations can be identified with McConnell & Golova's (2023)

“essential” aspects of the self. Beyond specific psychiatric conditions, we all exhibit character weaknesses, emotional tendencies, and behaviors that often clash with our self-narratives. This is why some of the skills that we need to exercise our agency involve actively regulating our behaviors and experiences to align with the normative expectations associated with our own values and self-concept. As McConnell & Golova (2023) admit, it is not always clear what the limits of self-shaping and self-narration are; after all, we can also work on changing our most deep-seated embodied dispositions. For instance, someone could change their deep-seated, problematic tendency toward jealousy by systematically training themselves—e.g., by exposing themselves to jealousy-triggering contexts, employing relaxation techniques while preventing habitual responses, etc. However, even relatively unchangeable aspects of our tendencies can be incorporated into different narratives to align them with our values and design strategies that avoid their unwanted consequences. For instance, jealousy can be understood as a sign of a highly possessive and envious person or as an expression of insecurity. Both narratives account for an aspect of the self that resists regulation, but the second may be more in line with our values and open possibilities for managing complex situations, such as seeking self-validation and enhancing self-esteem outside of romantic relationships.

Thus, agency also involves a process of self-discovery of aspects of our mental life that often clash with and force changes in our self-narratives. For the mindshaping view, the exercise of self-interpretation and self-regulation abilities does not ignore the fact that there are aspects of our mind that are not transparent to us. Often, we find ourselves feeling, experiencing, or behaving in ways that feel alien. However, this does not mean that these mental states and behaviors are not subject to revision, interpretation, or active regulation, in addition to being potentially assimilated by our own narratives. As such, self-mindshaping capacities imply the possibility of reinterpreting the limits of those feelings, experiences, or behavior.

Now then, could this mindshaping type of constructivism address the criticisms of individualism and relativism that we discussed in section 2? In addressing constructivism about SIA, the question of *who constructs what* arises. As discussed above, some forms of constructivism seem to imply that it is individuals who have the capacity to draw the line between self and illness:

“solving the ambiguity might mean that we draw a line and decide that traits, behaviors, thoughts, and emotions that fall on one side should be considered illness symptoms and those that fall on the other side part of myself” (Jeppsson, 2022, p. 295). However, the mindshaping perspective offers a more nuanced position regarding the scope of the individual’s ability to draw the self-illness divide. This is due, in part, to the fact that there are important aspects of our self-regulative and agential capacities that strongly depend on social interactions.

First, the self-narratives and self-conceptualizations that come into play in our self-interpretation are permeated by social aspects. Narratives and concepts are social models that we acquire already structured in a certain way, so the decision to fit certain self-narratives or self-conceptions can be socially biased. In fact, even the conception we have of illness is already shaped by the social dynamics themselves. In this sense, for example, it is far from obvious that an individual immersed in a culture that promotes narratives that depict mental illness as a discrete biological phenomenon different from the self could choose *not* to externalize certain behaviors and experiences. That is, we cannot know a priori whether this individual would be able to break free from prevailing social scripts and decide that those behaviors and experiences that are socially treated as psychiatric symptoms alien to the person actually reflect their true self.

Second, self-mindshaping capacities are not only dependent upon specific social models (e.g., being a “parent”), but are also subject to social dynamics that can influence the agent’s strategy for addressing SIA. Even if we decide to interpret the relation between self and illness according to a particular self-concept or narrative, the resulting self-ascriptions will be subject to public scrutiny and strategies of regulation by our social peers, which can raise difficulties in maintaining the self-narrative. Finally, decisions regarding the ambiguity between self and illness are often collective ones, involving family, therapists, and close ones. Those individuals play a fundamental role in drawing the boundaries between self and illness, such that the decision-making process itself is also socially mediated, rather than a product of individual decision-making (see De Haan, 2023).

As for the ‘anything goes’ challenge, the appeal to social dynamics as a source of scrutiny for our own self-conceptions and narratives serves as a corrective to the subjectivism that might concern those skeptical of

constructivism. In this sense, a mindshaping approach sheds light on the reality and articulation constraints of Schechtman's (2007) narrative account of identity (see section 2): general assumptions about reality and expectations about the logical articulation of our narratives are enforced through mindshaping mechanisms, thus limiting our self-narration possibilities.

A key concern with focusing on individual decision-making when choosing a narrative is that it may obscure the patient's suffering, or even hinder recovery by endorsing self- or illness-based narratives that are harmful or detrimental to the person’s integrity. However, as we have seen, from the mindshaping perspective, narratives ultimately require the agent to take responsibility for a range of behaviors, perceptions, and cognitions that are subjected to scrutiny. This scrutiny is, in an important respect, public — due to the social regulation by others — but also personal, as the subject's own values and norms also regulate their experiences and ways of coping with the environment.

This response to the subjectivist concern may, in fact, raise an opposite worry: that the mindshaping perspective devolves into some sort of communitarianism, where the community ultimately determines the line between self and illness. We do not think this is the case. While the mindshaping view acknowledges that biological mechanisms during development shape our minds in norm-conforming ways, it also recognizes that many of these mechanisms involve becoming aware of such norms and values — which implies the possibility of contesting and rewriting them (McGeer, 2015). Although peer-regulation dynamics prompt us to review our behavior, experience, or cognition, this need not always result in norm-conformity; contesting the norms and the norm-enforcing practices we are subjected to is a legitimate co-reaction to others’ normative expectations. Psychiatry’s history, where norm-resistance has led to significant normative revisions, is a primary example of this.

In short, while realists are right in noting that certain aspects of our mental life resist self-regulation and, therefore, are discovered rather than constructed, this does not necessarily pose a problem for constructivism beyond acknowledging the existence of frictions that we must accommodate in our own self-interpretations. This merely implies acknowledging that exercising agency is a *situated skill* (Leder and Zawidzki, 2023; Kalis et al., 2024), i.e., that our ability to act on our values, norms, and narratives is constrained by the world

and our embodied, social and psychological makeups. Narratives and self-conceptions are not conjured out of thin air. Furthermore, our narrative choices not only depend on, but also actively shape and transform our social niches in ways that constrain subsequent narrative possibilities. However, this does not mean reifying self or illness, nor does it imply that there is some definitive objective fact separating the two. Which facts are relevant to drawing the self-illness divide will depend on how we exercise our agency and respond to socially mediated normative expectations. That is, while facts may be important when distinguishing self from illness, *which* facts matter is normatively determined.

5. Concluding Remarks and Future Directions

We have argued that a self-mindshaping conception of metacognitive, self-interpretative, and agential capacities affords a sound constructivist understanding of SIA that still accounts for key realist intuitions. However, important questions remain open, particularly concerning its implications for research and therapy. One concern central to our view involves the possibility that the emphasis on the limits of self-narration might obscure the healing and emancipatory nature of constructivism about the self-illness divide. After all, our mindshaping perspective suggests that, no matter how empowering a certain interpretation of the self-illness divide may be for a particular subject, they might not be able to adopt it, due to the social frictions in which they are immersed or to resistant aspects of their condition.

Although a thorough response exceeds the scope of this chapter, our final remarks in section 4 already lay a first stone in that direction: self-mindshaping, as a socially based framework, acknowledges the possible and predictable resistance of prevailing narratives, but also highlights our ability to modify their underlying social dynamics. Recognizing resistance does not mean recognizing the impossibility of change; social reality is modifiable and contestable, and so are our self-interpretation possibilities. At the same time, this resistance helps us remain vigilant to the risks posed by imposed social narratives and the potential harms arising from the authority of certain social agents, including mental health professionals.

References

Borgoni, C., & de Pinedo, M. (forthcoming). Transparency and the second person: Epistemic intimacy in self-knowledge and knowledge of other minds. In B. Winokur & A. Andreotta (Eds.), *New perspectives on transparency and self-knowledge*. Routledge.

Bortolan, A. (2022). Selves hijacked: Affects and personhood in 'self-illness ambiguity'. *Philosophical Explorations*, 25(3), 343–362. https://doi.org/10.1080/13869795.2022.2093393

Chapman, R. (2023). *Empire of normality: Neurodiversity and capitalism*. Pluto Press.

Curtis, T. R., Dellar, R., Leslie, E., & Watson, B. (Eds.). (2000). *Mad pride: A celebration of mad culture*. Spare Change Books.

De Haan, S. (2020). The need for relational authenticity strategies in psychiatry. *Philosophy, Psychiatry, & Psychology*, 27(4), 349–351. https://doi.org/10.1353/ppp.2020.0044

De Haan, S. (2023). What do my problems say about me? *Philosophical Explorations*, 26(2), 159–164. https://doi.org/10.1080/13869795.2023.2205416

Dings, R., & de Bruin, L. C. (2022). What's special about 'not feeling like oneself'? A deflationary account of self(-illness) ambiguity. *Philosophical Explorations*, 25(3), 269–289. https://doi.org/10.1080/13869795.2022.2051592

Dings, R., & de Haan, S. (2022). The role of self-illness ambiguity and self-medication ambiguity in clinical decision-making. *The American Journal of Bioethics*, 22(6), 58–60. https://doi.org/10.1080/15265161.2022.2063436

Dings, R., & Glas, G. (2020). Self-management in psychiatry as reducing self-illness ambiguity. *Philosophy, Psychiatry, & Psychology*, 27(4), 333–347. https://doi.org/10.1353/ppp.2020.0043

Drożdżowicz, A. (2023). Self-illness ambiguity and anorexia nervosa. *Philosophical Explorations*, 26(1), 127–145. https://doi.org/10.1080/13869795.2023.2168033

Fenici, M., & Zawidzki, T. W. (2021). The origins of mindreading: How interpretive socio-cognitive practices get off the ground. *Synthese*, 198(9), 8365–8387. https://doi.org/10.1007/s11229-020-02577-4

Fernández-Castro, V., & Martínez-Manrique, F. (2021). Shaping your own mind: The self-mindshaping view on metacognition. *Phenomenology and the Cognitive Sciences*, 20(1), 139–167. https://doi.org/10.1007/s11097-020-09658-2

Haslam, N. (2000). Psychiatric categories as natural kinds: Essentialist thinking about mental disorder. *Social Research*, 67(4), 1031–1058.

Hutto, D. D. (2008). *Folk psychological narratives: The sociocultural basis of understanding reasons*. MIT Press.

Hyman, S. E. (2010). The diagnosis of mental disorders: The problem of reification. *Annual Review of Clinical Psychology*, 6(1), 155–179. https://doi.org/10.1146/annurev.clinpsy.3.022806.091532

Jeppsson, S. M. I. (2022). Solving the self-illness ambiguity: The case for construction over discovery. *Philosophical Explorations*, 25(3), 294–313. https://doi.org/10.1080/13869795.2022.2051589

Jorm, A. F., Angermeyer, M., & Katschnig, H. (2000). Public knowledge of and attitudes to mental disorders: A limiting factor in the optimal use of treatment services. In G. Andrews & S. Henderson (Eds.), *Unmet need in psychiatry: Problems, resources, responses* (pp. 399–413). Cambridge University Press. https://doi.org/10.1017/CBO9780511543562.032

Kalis, A., Pascoe, J., & Segundo Ortin, M. (2024). Really situated self-control: Self-control as a set of situated skills. *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-024-09989-4

Karp, D. A. (1992). Illness ambiguity and the search for meaning: A case study of a self-help group for affective disorders. *Journal of Contemporary Ethnography*, 21(2), 139–170.

Kendell, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry*, 160(1), 4–12. https://doi.org/10.1176/appi.ajp.160.1.4

Kramer, P. D. (2005). *Against depression*. Viking.

Kvaale, E. P., Haslam, N., & Gottdiener, W. H. (2013). The 'side effects' of medicalization: A meta-analytic review of how biogenetic explanations affect stigma. *Clinical Psychology Review*, 33(6), 782–794. https://doi.org/10.1016/j.cpr.2013.06.002

Leder, G., & Zawidzki, T. (2023). The skill of mental health: Towards a new theory of mental health and disorder. *Philosophy and the Mind Sciences*, 4. https://doi.org/10.33735/phimisci.2023.9684

McConnell, D., & Golova, A. (2023). Narrative, addiction, and three aspects of self-ambiguity. *Philosophical Explorations*, 26(1), 66–85. https://doi.org/10.1080/13869795.2022.2115532

McGeer, V. (2007). The regulative dimension of folk psychology. In D. D. Hutto & M. Ratcliffe (Eds.), *Folk psychology re-assessed* (pp. 137–156). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5558-4_8

McGeer, V. (2008). The moral development of first-person authority. *European Journal of Philosophy*, 16(1), 81–108. https://doi.org/10.1111/j.1468-0378.2007.00266.x

McGeer, V. (2015). Mind-making practices: The social infrastructure of self-knowing agency and responsibility. *Philosophical Explorations*, 18(2), 259–281. https://doi.org/10.1080/13869795.2015.1032331

McGeer, V. (2021). Enculturating folk psychologists. *Synthese*, 199(1–2), 1039–1063. https://doi.org/10.1007/s11229-020-02760-7

Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton University Press.

Rego, M. D. (2004). Externality in psychiatry and the paradox of agency. *Philosophy, Psychiatry, & Psychology*, 11(4), 313–321. https://doi.org/10.1353/ppp.2005.0013

Sadler, J. Z. (2007). The psychiatric significance of the personal self. *Psychiatry: Interpersonal and Biological Processes*, 70(2), 113–129. https://doi.org/10.1521/psyc.2007.70.2.113

Schechtman, M. (2007). Stories, lives, and basic survival: A refinement and defense of the narrative view. *Royal Institute of Philosophy Supplement*, 60, 155–178. https://doi.org/10.1017/S1358246107000082

Strijbos, D., & de Bruin, L. (2015). Self-interpretation as first-person mindshaping: Implications for confabulation research. *Ethical Theory and Moral Practice*, 18(2), 297–307. https://doi.org/10.1007/s10677-015-9579-3

Strijbos, D., & Slors, M. (2020). What kind of 'management' is self-management? A two-dimensional approach to self-management in mental health care.

Philosophy, Psychiatry, & Psychology, 27(4), 355–370. https://doi.org/10.1353/ppp.2020.0046

Tekin, Ş. (2019). The missing self in scientific psychiatry. *Synthese*, 196(6), 2197–2215. https://doi.org/10.1007/s11229-017-1324-0

Tekin, Ş. (2022). My illness, my self, and I: When self-narratives and illness-narratives clash. *Philosophical Explorations*, 25(3), 314–318. https://doi.org/10.1080/13869795.2022.2097300

Wardrope, A. (2015). Medicalization and epistemic injustice. *Medicine, Health Care and Philosophy*, 18(3), 341–352. https://doi.org/10.1007/s11019-014-9608-3

Wilkinson, S. (2023). Psychiatric fictionalism and narratives of responsibility. *Philosophical Explorations*, 26(1), 91–109. https://doi.org/10.1080/13869795.2022.2116473

Zawidzki, T. W. (2008). The function of folk psychology: Mind reading or mind shaping? *Philosophical Explorations*, 11(3), 193–210. https://doi.org/10.1080/13869790802239235

Zawidzki, T. W. (2013). *Mindshaping: A new framework for understanding human social cognition*. MIT Press.

Zawidzki, T. W. (2016). Mindshaping and self-interpretation. In J. Kiverstein (Ed.), *The Routledge handbook of philosophy of the social mind* (pp. 495–513). Routledge. https://doi.org/10.4324/9781315530145