To appear in *The Routledge Handbook of Mindshaping*, ed. Tad Zawidzki

Psychotherapy as a folk-psychological practice: Therapeutic mindreading and mindshaping

J.P. Grodniewicz
Copernicus Center for Interdisciplinary Studies
Jagiellonian University
(j.grodniewicz@gmail.com)

Abstract

Most psychotherapeutic approaches are, to a greater or lesser extent, rooted in the theories and principles of scientific psychology. Nevertheless, in-session psychotherapeutic interaction between a therapist and a client is, at its core, a folk-psychological practice. As such, it is based on folk-psychological skills and competencies. But which ones exactly? This chapter argues that, while we may initially be inclined to perceive the practice of psychotherapy as primarily involving sophisticated mindreading on the part of both the therapist and the client/patient, a complete characterization of psychotherapy must give at least the same amount of attention to different forms of therapeutic mindshaping. Using examples drawn from multiple therapeutic traditions, I illustrate how therapeutic mindreading and mindshaping interact. I conclude by highlighting some of the consequences of this perspective for the ethics and politics of psychotherapy.

1. Introduction

Most psychotherapeutic approaches are, to a greater or lesser extent, rooted in theories and principles of scientific psychology. Nevertheless, in-session psychotherapeutic interaction between a therapist and a client is, at its core, a folk-psychological practice. As such, it is based on folk-psychological skills and competencies. But which ones exactly?

One way to think about it is to see psychotherapy (also called “talk therapy”) as, essentially, an exercise in mindreading—the ability to infer other people’s mental states and thus predict how they are likely to behave. Through mindreading, the therapist acquires an understanding of what lies at the core of the client’s mental suffering and—by offering interpretation—helps the client achieve such understanding or “insight” for themselves.

This brings to mind the earliest stages of the development of the original psychotherapeutic approach—psychoanalysis. As Freud—the founder of psychoanalysis—reportedly explained to one of his patients: “[t]he psychoanalyst, like the archaeologist in his excavations, must uncover layer after layer of the patient’s psyche, before coming to the deepest, most valuable treasures.” (Gardiner, 1989, p. 139). The analyst appears here as a master mindreader, whose gaze reaches far beyond what is available to the patient—possibly all the way to the hideaways of the patient’s unconscious. Crucially, throughout the process, the analyst should remain “neutral” (as physically manifested with the patient lying on the couch and the analyst sitting out of sight). As indicated by Hoffer (1985): “[n]eutrality, in a general sense, is the optimal position from which the analyst gathers his data…” (p. 773).

I do not intend to question that mindreading plays an important role in psychotherapy. I think it does. Psychotherapy is the optimal context for sophisticated mindreading; if humans do it at all, it is during the fifty-minute sessions devoted to reflecting on and talking about one’s mental life.

Nevertheless, if we focus solely on mindreading, we will end up with a glaringly incomplete picture of psychotherapy. Already at the beginning of the 20th century, Hungarian psychoanalyst, and close collaborator of Freud, Sándor Ferenczi, “[b]elieving that the psychoanalyst’s personality strongly influences the course of treatment, … criticized the overvaluation of theoretical insight” (Orange, 1995, p. 161). Since then, an increasing number of authors have argued that what is crucial for psychotherapeutic change is not what patients recognize, learn, or understand about themselves but what they experience

and how this experience shapes them (see, e.g., Alexander & French, 1946; Friedman, 1978). This tendency has become even more visible along with the emergence of therapeutic orientations developed as alternatives to psychoanalysis, such as humanistic, Gestalt, Cognitive Behavioral, systemic, or integrative therapies.

In this chapter, I will argue that psychotherapy as it is currently practiced—be it in psychoanalytic/psychodynamic or any other modality—consists of a constant interplay of two general kinds of folk-psychological processes: mindreading (e.g., Baron-Cohen, 1995; Gopnik, 1993; Nichols & Stich, 2003; Gallagher, 2004; Goldman, 2006; Spaulding, 2018) and mindshaping (Mameli, 2001; McGeer, 2007, 2015; Zawidzki, 2013, 2018a). More precisely, in psychotherapy, therapists read (mindreading) and shape (mindshaping) the minds of their clients. Moreover, they create the context in which clients become increasingly able to read (self-mindreading, e.g., Nichols & Stich, 2003; Carruthers, 2013) and shape (self-mindshaping, e.g., McGeer, 2007; Strijbos & De Bruin, 2015; Zawidzki, 2016; Fernández-Castro & Martínez-Manrique, 2021) their own minds. Only by taking all these components into account will we be able to fully appreciate the complexity of psychotherapy as a folk-psychological practice and ask the right questions about its consequences, including the ethical and political ones.

Here is the plan for the rest of the chapter. In Section 2, I specify what I mean by psychotherapy. In Section 3, I discuss the most important aspects of therapeutic mindreading. In Section 4, I focus on therapeutic mindshaping and argue that thinking in terms of mindshaping is indispensable for characterizing what goes on in psychotherapy. Finally, in Section 5, I discuss how therapeutic mindreading and mindshaping jointly contribute to the achievement of therapeutic insight defined as deepening one's self-understanding.

2. What is psychotherapy?

For a phenomenon so prevalent in contemporary Western culture, psychotherapy proves surprisingly difficult to define. This is because what we attempt to capture is not one thing but a family of, nowadays, more than 500 specific approaches (Prochaska & Norcross, 2018), more or less squarely fitting into one of a dozen or so more general traditions. Thus, the best we can do while discussing psychotherapy in general is to focus on integrative or transtheoretical definitions, which attempt to approximate what is true about psychotherapy independently of a specific denomination. I will briefly discuss two such definitions.

The first comes from Wampold and Imel (2015):

> Psychotherapy is a primarily interpersonal treatment that is a) based on psychological principles; b) involves a trained therapist and a client who is seeking help for a mental disorder, problem, or complaint; c) is intended by the therapist to be remedial for the client disorder, problem, or complaint; and d) is adapted or individualized for the particular client and his or her disorder, problem, or complaint. (p. 37)

This definition has several advantages. First, it highlights the relational or interpersonal aspect of psychotherapy. Psychotherapy happens between two or, in the case of couple or group therapy, more people. Second, the definition is deliberately non-specific, characterizing what psychotherapy is intended to address broadly as a “disorder, problem, or complaint.” Narrowing the list to disorders would unnecessarily restrict our concept of psychotherapeutic care. Finally, the definition highlights that the therapist has to be prepared for their role (“trained”) and willing to adapt the treatment to the needs of particular clients.

Simultaneously, the definition has one major weakness. It tells us nothing about how psychotherapy works.

While still ecumenical and transtheoretical, a definition offered by John Norcross (1990) is slightly more committal in this respect:

> Psychotherapy is the informed and intentional application of clinical methods and interpersonal stances derived from established psychological principles for the purpose of assisting people to modify their behaviors, cognitions, emotions, and/or other personal characteristics in directions that the participants deem desirable. (p. 218)

What happens in all psychotherapies—or so Norcross suggests—is the modification of behavioral, cognitive, and affective patterns. This clarification helps as much as it directs us towards natural further questions: "Which patterns ought to be modified?" "What are the means we use?" and "What are we aiming for, i.e., what is the model or an ideal we are pursuing?" These questions concern, respectively, the diagnosis or case conceptualization, the repertoire of therapeutic methods, and the goal of therapy.

Answers to these questions are what differentiates and individuates various therapeutic approaches.

For example, a representative of the so-called "classical" or "second wave" Cognitive Behavioral Therapy (CBT) developed by Aaron Beck and Albert Ellis, could say that what we aim to modify in therapy are primarily maladaptive patterns of thinking often called "cognitive distortions," e.g., overgeneralization, jumping to conclusion, or all-or-nothing thinking, which underlie various psychological problems from mood (A. T. Beck, 1979) to personality disorders (A. T. Beck et al., 2004). The means by which they are modified include a wide range of cognitive (e.g., cognitive restructuring) and behavioral (e.g., exposure) interventions. Finally, the central goals of therapy are "to provide symptom relief, facilitate a remission of the disorder, help patients resolve their most pressing problems, and teach them skills to avoid relapse" (J. S. Beck, 2011, p. 9).

On the other side of the therapeutic spectrum, what is being addressed in the course of existential psychotherapy is the anxiety resulting from the way in which a patient relates to existential themes such as death, freedom, isolation, and meaninglessness (Yalom, 1980). This happens primarily through the cultivation of therapeutic presence and supporting clients in noticing and confronting various forms of resistance (e.g., ambivalence, shame, control) that they encounter while exploring themselves and their life situation. Finally, the overarching goal of existential therapy is to help clients "gain the clarity needed to more fully realize who they deeply are, the multiple choices available to them, and how to live a life with fuller conscious consent" (Krug & Piwowarski, 2021, p. 562).

These are merely two glaringly simplified illustrations. Neither here nor in the subsequent sections can I hope to provide a comprehensive characterization of the nuances of even one therapeutic approach. Luckily, I do not have to. All I want to achieve is to demonstrate that despite all the differences between and idiosyncrasies of hundreds of its available variants, considered in general, psychotherapy is, at its core, a folk psychological practice consisting of a mix of mindreading and mindshaping.

3. Therapeutic mindreading

Mindreading is commonly characterized as the ability to attribute mental states such as beliefs, desires, and emotions to other agents, in order to understand what they do and why they do it. Historically, we can speak about the two most influential accounts of this ability. According to the Theory Theory (TT) (e.g., Baron-Cohen, 1995; Gopnik, 1993), mindreading is primarily a matter of *theorizing* about other minds, by forming hypotheses about how their mental states relate to and influence each other, perception, and behavior. Simulation Theory (ST) (e.g., Goldman, 2006), by comparison, suggests that we do not primarily theorize about but rather *simulate* what goes on in other minds. Nowadays, many authors agree that mindreading happens by means of multiple processes (see, e.g., Nichols & Stich, 2003; Andrews,

2012; Newen, 2015; Fiebich, 2021). Besides theorizing about and simulating other's mental states, the list includes (but is not limited to) recognition of other's intentions as expressed in their embodied actions (Gallagher, 2004), character trait attribution (Andrews, 2012; Westra, 2021), social categorization and stereotyping (Spaulding, 2018), and production and consumption of narratives (Hutto, 2007).

The view of mindreading I find most promising when it comes to specifying the role it plays in psychotherapy is the so-called "Model Theory" (MT). Originally, the view was proposed by Heidi Maibom (2003, 2009) and Peter Godfrey-Smith (2005) as an improvement on the classical TT. While classical TT assumed that, just like science, folk psychology depends on (tacit) knowledge of laws and universal generalizations, Maibom and Godfrey-Smith argued that this assumption is false both about science and folk psychology. Instead, both scientists and mindreaders understand the world through their competency with models that they can apply to understand particular phenomena and situations. For example, when I am at a business meeting, I interpret and predict what people are doing by applying a model of business meetings. That is why I would be perplexed if someone came in speedos, started ranting about their marital problems, or greeted me with a ten-second hug.

The models Maibom and Godfrey-Smith talk about are general-purpose (the same model of business meetings can be applied to multiple agents in multiple real-life situations), and partial (interpreting one's behavior through the lens of a business meeting model affords only a limited outlook a person involved). More recently, Newen (2015), Andrews (2015), and Spaulding (2018) independently took MT a step further, suggesting that, at least sometimes, mindreading involves not only applying a set of general-purpose, partial models, but also *building a specific model*, or simply, *modeling* a person we

try to understand.[1] Importantly, MT allows for a plurality of mindreading strategies, which can feed into a model of a person we are trying to understand.

Of course, such elaborate modeling is not always necessary. As pointed out by Godfrey-Smith (2005, p. 10), there is a difference between "folk psychology on the freeway and in the lawcourt." An equally good comparison is one between a freeway and a therapy room. I do not have to build a complete person model each time I try to predict the behavior of a driver changing the lane in front of me. However, a therapist *has to* build a detailed model of their patient—with particular emphasis on the nature and origin of their mental suffering—which is later used to guide the therapy process. In psychotherapy, such modeling has a name: "case formulation."[2]

Case formulation is "a process for developing a hypothesis about, and a plan to address, the causes, precipitants, and maintaining influences of a person's psychological, interpersonal, and behavioral problems…" which "includes consideration of within person factors, such as the person's learning history, style of interpreting information, coping style, self-concept, core beliefs, and basic, axiomatic assumptions about the world." (Eells, 2015, p. 16). To put it bluntly, the goal of case formulation is to figure out the nature of the patient's struggle and whether and how therapy can help. Case formulation may, but does not have to, include diagnostic formulation, spelled out in terms of a psychiatric diagnosis, such as the ones we find in the Diagnostic and Statistical Manual (DSM) or International Classification of Diseases (ICD). Crucially, case formulation is an ongoing process—the model of a patient amidst their struggle ought to be constantly tested and updated.

[1] While Newen and Spaulding explicitly characterize their views as developments of what has been proposed by Maibom and Godfrey-Smith, Andrews does not. Nevertheless, in her (2015) paper, she is very explicit that understanding people involves building their models.

[2] Recently, Anya Plutynski (2024), without invoking MT or any other account of folk psychology, argued that therapeutic case formulation has much in common with *scientific* modeling. While Plutynski is certainly on the right track, I believe that this claim becomes most persuasive once we recognize that: (1) in general, in the spirit of MT, understanding people (including oneself, see Section 5 and Grodniewicz (ms)) can be seen as a form of modeling, and (2) psychotherapeutic case formulation is a special case of understanding people (clients/patients).

What is particularly interesting about psychotherapeutic case formulation is the mix of theoretical knowledge and folk psychological skills it requires. Therapists can and sometimes do use psychometric tools such as questionnaires and inventories. Moreover, they typically use concepts developed on the grounds of their specific therapeutic tradition. That is why psychoanalytic case formulation will be spelled out in terms of developmental issues, defense mechanisms, and transference reactions (e.g., McWilliams, 1999), while CBT case formulation will mention automatic thoughts, core beliefs, moods, behaviors, physical reactions, etc. (e.g., Kuyken et al., 2009). Nevertheless, the vast majority of data therapists operate on comes from ongoing elaborated mindreading. This is nicely illustrated by Fritz Perls—the founder of Gestalt Therapy:

> ...everything the patient does, obvious or concealed, is an expression of the self. His leaning forward and pushing back, his abortive kicks, his fidgets, his subtleties of enunciation, his split-second hesitations between words, his handwriting, his use of metaphor and language, his use of 'it' as opposed to his use of 'you' and 'I', all are on the surface, all are obvious, and all are meaningful. (Perls, 1981, p. 76)

But the role of therapeutic mindreading is not limited to case conceptualization. Constant careful mindreading is necessary at every stage of the psychotherapeutic process. That is why, in transcripts of psychotherapeutic sessions, we frequently encounter descriptions of significant behaviors accompanying what people are saying: [*with an audible sigh*], [*after a pause*], [*looking away*], [*with a chuckle*], picked out by attentive therapists: "I noticed you sighed heavily while talking about how your brother treated you. How did you feel when speaking about it? Would it be fair to say that there is a certain amount of resignation you experience about what is going on?"

This is most striking when therapists notice that an emotional expression does not match the content of what the client is saying. For example, if a client smiles when speaking about something painful, a therapist may draw their

attention to this fact and examine the role the smile played for the client. Due to their learning history, some patients display a happy-go-lucky attitude and avoid manifesting negative emotions. In such cases, a therapist will often inquire into the discrepancy: “You smiled when you were telling me about the diagnosis you received, but I imagine receiving such news must be painful.” Sometimes, a therapist may even use self-disclosure to highlight the discrepancy: “When you were telling me about the diagnosis, I saw you smile, but at the same time I felt this heaviness in my chest and I thought ‘Gee, this is terrible news.’” In doing this, the therapist reveals not only their own experience, but a folk psychological model they apply while interpreting a given situation; in this case, a model which predicts feeling shocked, saddened, and scared after hearing an unfavorable medical diagnosis.

Naturally, what is important in psychotherapy is not only mindreading performed by a therapist. Equally, if not more important, is both the self- and other-related mindreading performed by a client. The authors of Mentalizing Based Therapy (MBT) (Bateman & Fonagy, 2010) go as far as to identify improving one’s mentalizing (another name for mindreading) skills as the main goal of therapeutic work. This is—they hypothesize—most important in the case of borderline personality disorder, associated with the breakdown of mindreading abilities, especially amidst distress. I will come back to this topic in Section 5. But now, once we have some outlook on how therapists use mindreading to understand their patients and identify their problems, we should look at how they use mindshaping to help them change.

4. Therapeutic mindshaping

In Section 2, I said that all therapeutic approaches answer three questions, albeit differently: “What patterns (of behavior, thinking, affect) need to be modified, and what developmental trajectories attended to and corrected?”, “By what means?”, and “Towards what end?” This immediately brings to mind the three components of the mindshaping model. As stated by Zawidzki (2018a, p. 739):

> I define mindshaping as a relation between a target mind (the mind being shaped), a cognitive mechanism (the proper function of which involves shaping that mind), and a model that the mindshaping mechanism works to make the target mind match.

In the case of psychotherapy, the target mind is the mind of a client/patient, the mechanism is whatever underlies and is responsible for the effectiveness of interpersonal stances, methods, and interventions used in the process, and the model is whatever ideal of mental health or well-being a given therapeutic tradition employs.

In contrast to mindreading view, which construes folk psychology primarily as an epistemic (predictive and explanatory) activity, the basic idea of mindshaping is that our folk psychological efficacy and the success of our social coordination result from the fact that—by means of mechanisms such as imitation, pedagogy, production and reception of narratives, and norm enforcement—we "mold" or "shape" each other. Folk psychology is a regulative practice (Mameli, 2001; McGeer, 2007, 2015; Zawidzki, 2013, 2018a).

Crucially, the mindreading-mindshaping debate does not concern the question of *whether* our social cognition involves mindreading *or* mindshaping, but which has the phylogenetic and ontogenetic primacy over the other (this is explicitly admitted by champions of mindshaping, e.g., McGeer, 2015; Zawidzki, 2018a; for discussion see, e.g., Spaulding, 2018; Peters, 2019; Westra, 2021). While mindreaders argue that we are efficient mindshapers, thanks to a prior, epistemic ability to correctly ascribe mental states to ourselves and others, mindshapers claim that it is the other way around.

Here, fortunately, I do not have to take a stance in the primacy debate. This is because I focus exclusively on psychotherapy, which is a unique form of human interaction involving all our most sophisticated folk-psychological skills. In particular, it is a context where speed and cost efficiency do not play such a role—there are typically ample amounts of time and mutual attention participants are ready to devote to one another. Moreover, the stakes are so

high that participants are willing to invest significant resources to ensure the success of their joint venture. Thus, whether mindreading or mindshaping is phylogenetically or ontogenetically primary will neither be reflected in nor affect what occurs in psychotherapy.

The role of mindreading in psychotherapy has been discussed in the previous section. We will now look at examples of therapeutic mindshaping.

Some neobehavioral therapeutic approaches, such as Functional Analytic Psychotherapy (FAP), explicitly conceptualize much of what is done in psychotherapy as shaping client's behaviors:

> the FAP therapist shapes behavior by differentially reinforcing approximations to more useful interpersonal behaviors. The therapist reduces counterproductive client behaviors by either failing to reinforce them, establishing competing behaviors that are more useful, or occasionally punishing them when they are harmful to others. (Follette et al., 1996, p. 625)

A good example could be a therapist responding "I really appreciate your honesty" after the client says that they did not like the previous session—a response intended to reinforce the client's honest sharing of difficult information and feedback.

This is a behavioral incarnation of a more general idea of "corrective experiences," first introduced by psychoanalysts Alexander and French (1946) and picked up by most therapeutic traditions (Castonguay & Hill, 2012). According to this idea, in therapy and beyond, the client should have an opportunity to experience a familiar type of event or relationship in a novel and unexpected fashion. For example, clients raised in prudish environments may expect their interlocutors to be uncomfortable and dismissive when the conversation turns to sex and intimacy. Each time they experience a therapist's receptive response to such topics, new expectations and assumptions about what is acceptable, normal, and important are being shaped.

An even more general therapeutic mechanism is modeling. It taps into our basic mindshaping mechanism of imitation and concerns the stance assumed and consistently displayed by a therapist. For example, therapists working within classical Cognitive Behavioral Therapy (CBT) or Rational Emotive Behavior Therapy (REBT) will, by their own in-session behavior, attempt to model an attitude of inquisitiveness and curiosity towards the contents of thoughts and beliefs of their clients, necessary for spotting maladaptive thinking patterns (J. S. Beck, 2011). Rogerian therapists will attempt to model unconditional acceptance toward the clients just as they are (Rogers, 1946). Compassion Focused (CFT) therapists will attempt to model a care-based, compassionate stance (Gilbert, 2010), and Gestalt therapists—an unremitting focus on the present moment, on what one feels and experiences right now (Perls, 1981). The list can go on.

Borges and Koening (1983) sum it up perfectly (and almost explicitly in mindshaping terms) in their discussion of modeling in group therapy: “The best model is often the therapist. Modeling may occur spontaneously or may be deliberately employed… The modeled behaviors may be used to shape norms, to inhibit or disinhibit behaviors, and to produce independent behavior in the group” (p. 133).[3]

To go even further, in some therapies, e.g., Schema Therapy, the therapist assumes the role of a good caretaker whom the client may have never encountered in the past. Through this process, quite tellingly called “limited reparenting,” “the therapist helps patients go back into that child mode and to learn to get from the therapist, and later from themselves, some of what they missed” (Young et al., 2006, p. 129). To put it bluntly, some of the mindshaping that should have but did not occur in the relationship with one’s primary caregivers can later occur in a relationship with a therapist.

[3] An important, additional dynamic of group therapy and couple therapy is mutual shaping occurring between clients.

This was anticipated in psychoanalytic approaches of the second half of the 20th century, often grouped under the label of object relations and intersubjective theories. English psychoanalyst Donald Winnicott famously suggested that therapists for their clients—just like mothers for their infants—should create a "holding environment," in which one's "true self" can safely develop (Winnicott, 1965). Similar ideas underlie Balint's concept of "psychological mothering" (Balint, 1985) and Kohut's theory according to which "in the treatment situation a 'good object' is provided for the patient in the form of a therapist who will be internalized, and thus mitigate or repair deficits in the self-structure resulting from inadequate early parenting" (Buckley, 1994, p. 519).

A form of therapeutic mindshaping most extensively discussed in philosophical literature is the production and reception of narratives (Hutto & Gallagher, 2017; Hutto, 2023). Hutto (2023) carefully examines Narrative Therapy (White & Epston, 1990) and suggests that such treatment "can shape our thinking" (p. 60) about who we are and who we want to be. To this, we can add other forms of high-level, language-based mindshaping, such as the insistence on the client's use of change-eliciting language in Motivational Interviewing (Magill & Hallgren, 2019), identification and clarification of personal values, which are supposed to guide one's further behavior in Acceptance and Commitment Therapy (Hayes et al., 1999), guided meditations in which clients are supposed to imagine themselves as having qualities ascribed to inanimate objects (e.g. stillness in "mountain meditation" popular in various mindfulness-based treatments (Segal et al., 2018)), Fixed-role therapy (Kelly, 1955) based on creating a description of a character with some desirable features, whom the client is supposed to later "enact," and many more.

This overview is necessarily sketchy—detailed documentation of all forms of therapeutic mindshaping could easily fill a monograph. Hopefully, it is sufficient to establish that the folk psychological practice of psychotherapy consists of at least as much mindshaping as mindreading.

5. Therapeutic self-understanding

A domain of therapeutic work where the interplay of mindreading and mindshaping is particularly striking is the acquisition of therapeutic insight (Marková & Berrios, 1992; Lacewing, 2014). In (Grodniewicz & Hohol, 2023), I suggested that we can identify insight with the deepening of one's self-understanding. Self-understanding, in turn, can be seen as a form of objectual understanding, i.e., understanding of a phenomenon or a subject domain consisting in grasping coherence making relations between constituents of a body of information about this phenomenon (Kvanvig, 2003; Zagzebski, 2008). In the case of self-understanding, the target phenomenon is oneself, and the body of information is one's self-concept, i.e., whatever we bring to mind while thinking about ourselves (Neisser, 1997).

Importantly, as suggested by Dellsén (2020), objectual understanding of a phenomenon can be seen as a form of modeling it. This sits well with the Model Theory of mindreading introduced in Section 3, especially with Newen's (2015) version, according to which we understand people, including ourselves, by building their (more or less complete) person models. Where I part ways with Newen, is his characterization of a person model as: "A unity of properties or features that we represent in memory as belonging to one person" (Newen, 2015, pp. 2–3). An unorganized "unity" will not do. What is crucial for understanding is grasping coherence-making relations (Kvanvig, 2003) or modeling dependency relations (Dellsén, 2020) between these features, properties, and other information we have about ourselves. It is not enough, e.g., to believe that I have problems at work *and* overuse alcohol to understand (or have an insight into) my drinking problem. I have to grasp that I have problems at work *because* I overuse alcohol.[4]

Be it as it may, what is crucial from the point of view of the current discussion, is a peculiar dynamic at the core of deepening one's self-understanding

[4] For a detailed discussion of self-understanding as a kind of objectual understanding see (Grodniewicz, ms).

through psychotherapy. Psychotherapy is simultaneously an epistemic process of self-exploration and a regulative process of self-transformation.[5]

The goal of therapeutic self-exploration is to assess our current situation and the potential for improvement (cf. Fileva & Brakel, 2023). The epistemic capacities employed for this task are the ones discussed in a vast and ever-growing literature on self-knowledge (Gertler, 2021).[6] The goal of self-transformation, in turn, is to move towards what is desirable. It happens, at least in part, by means of self-mindshaping (cf. McGeer, 2007; Strijbos & De Bruin, 2015; Zawidzki, 2016, 2018b; Fernández-Castro & Martínez-Manrique, 2021).

Proponents of self-mindshaping highlight the unreliability of self-knowledge—when asked about our mental states, we commonly confabulate and deceive ourselves (Strijbos & De Bruin, 2015). Some go as far as to claim that self-interpretation is not optimized for truth but utility, measured primarily as our ability to coordinate with others (Zawidzki, 2018b). From this, they infer a specific version of the primacy thesis: self-interpretation has "primarily constitutive rather than an epistemic function" (Zawidzki, 2016, p. 489).

Once again, at least while speaking about the specific context of psychotherapy, I am inclined to dodge the primacy problem. Regardless of whether the constitutive or epistemic aspect of self-interpretation is phylogenetically or ontogenetically primary, the kind of self-exploration that takes place in psychotherapy consists of a constant interplay between both. Without some amount of self-mindreading (even if not perfectly reliable), we would not be able to foster motivation for change or design a plan for pursuing it. As pointed out by Westra (2021, pp. 8223–8224): "...when seeking to effect a change in a person's mental states that will cause them to behave in a predictable way, it helps to know what that person's mental states are, and to be able to predict

[5] For a related discussion of this dynamic see Strijbos and Jongepier (2018).
[6] While some believe that how we get to know ourselves is symmetrical to how we get to know others (e.g., Carruthers, 2011), most others disagree, and focus on the distinctiveness and privileged status of knowledge about oneself.

how their minds might change in response to different actions." This is what therapists do through case conceptualization, and what clients/patients do while deepening their understanding of themselves and their mental struggles through psychotherapy. For example, unless I recognize that there are some depressive patterns present in how I think about and interact with the world (e.g., by withdrawing from activities that used to bring me joy), I will not gain insight into my depression. Consequently, I will fail to see the reason to seek help in order to change my predicament.

At the same time, many self-ascriptions that take place in the context of psychotherapy have what McGeer (1996) calls "forward looking truth conditions," and thus should be considered examples of self-mindshaping. If therapy brings me to the conclusion that I am a strong, capable person ready to face life's adversities, this might be more of an aspiration than a recognition. It is a reflection of how I am choosing to perceive myself from now on, a perspective I still need to fully integrate[7] with my updated self-understanding and—more importantly—my actions.

Even though I have only managed to scratch the surface of this complicated problem in this section, it seems undeniable that the acquisition of therapeutic insight is not purely an epistemic matter of sophisticated self-mindreading. While deepening our self-understanding through psychotherapy, we simultaneously shape our new selves. Self-exploration triggers self-transformation, and vice versa. As with other elements of the process of psychotherapy, we will not be able to give a full account of psychotherapeutic insight unless we view it simultaneously in light of therapeutic mindreading and mindshaping.

[7] This "full integration" might be what many authors refer to as "emotional insight" (see, e.g., Strachey, 1934; Richfield, 1954; Ellis, 1963; Poland, 1988). As interesting as it is, I will have to leave this topic for another occasion.

6. Conclusion

In this chapter, I argued that psychotherapeutic practice relies on a constant interplay of different folk-psychological processes. While we may be initially inclined to perceive this practice as primarily involving a sophisticated mindreading on the part of the therapist, a complete characterization of psychotherapy must pay at least the same amount of attention to all forms of therapeutic mindshaping.

This brings us to the verge of yet another discussion—one about the ethics and politics of psychotherapy. For many years now, within critical approaches to mental health care, we have been asking: What is our yardstick for "normal" and "healthy"? To what extent is the way in which we think about these concepts determined by the current cultural, political, and economic context? Does it enable and encourage us to critically reappraise it, or rather petrify the status quo? (see, e.g., Foucault, 1988; Frank & Frank, 1991; Illouz, 2008; Watters, 2010; Frances, 2013).

The present discussion gives us a solid footing for future critical inquiries into the ethics and politics of psychotherapy. Psychotherapy is a form of mindshaping. Within the process, clients' minds are molded to match certain models. Where do the models come from? Who chooses them and based on what criteria? How much say do clients have in what model they will be shaped into?

Perhaps one of the most striking historical examples of abuse and aberration resulting from arbitrary choices of a target model was "conversion therapy," designed to shape homosexual and bisexual individuals—considered "abnormal"—into the likeness of "normal" heterosexuals. A long shadow of shame for these harmful practices rests on the psychiatric and psychotherapeutic communities to this day.

A less extreme example is the emergence and proliferation of the "ideal" of emotional control that was tightly connected with the social changes taking place at the turn of the nineteenth and twentieth centuries, especially with the

emergence of a new managerial class (Illouz, 2008). The belief that started gaining currency around this time was that a professional and competent person is one who keeps their emotions in check at all times. This triggered the boom of an entire industry—fed by the newest achievements of psychological sciences—designed to help people shape themselves into the new model of workplace professionalism.[8] Simultaneously, the imposition of the ideal of emotional control exacerbated gender inequalities—with women being considered "more emotional" and thus "less professional"—and contributed to the situation in which, nowadays, many men have to go to therapy to "reconnect" with their emotions, and thereby undo the problematic consequences of the cultural conditioning they received.

Models we choose to shape each other are rarely "neutral" or "objective." This is well illustrated by the constant succession of new models of "health" and "normalcy" throughout the history of psychiatry. Given that—as I argued in this chapter—psychotherapy is a folk-psychological practice involving not only mindreading but also mindshaping, choosing the models we rely on in this practice is not only a scientific but also an ethical and political decision.[9]

References

Alexander, F., & French, T. M. (1946). *Psychoanalytic therapy; principles and application* (pp. xiii, 353). Ronald Press.

Andrews, K. (2012). *Do Apes Read Minds?: Toward a New Folk Psychology.* The MIT Press. https://doi.org/10.7551/mitpress/9283.001.0001

[8] One of the newest incarnations of this trend is the so called "McMindfulness" (Purser, 2019)—the use of a trivialized version of selected elements of the mindfulness tradition as tools to achieve success in neoliberal capitalist economy.

[9] I am grateful to Grzegorz Gaszczyk, Rémi Tison, Anna Urbanek, and Tad Zawidzki for many helpful comments and suggestions. Special thanks to Jan Topczewski for his invaluable help in gathering examples of mindreading and mindshaping across psychotherapeutic traditions. This research was supported by the National Science Centre, Poland (grant number: 2022/47/D/HS1/00923).

Andrews, K. (2015). Pluralistic folk psychology and varieties of self-knowledge: An exploration. *Philosophical Explorations*, *18*(2), 282–296. https://doi.org/10.1080/13869795.2015.1032116

Balint, M. (1985). *Primary Love and Psycho-analytic Technique*. Maresfield Library.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind* (pp. xxii, 171). The MIT Press.

Bateman, A., & Fonagy, P. (2010). Mentalization based treatment for borderline personality disorder. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *9*(1), 11–15. https://doi.org/10.1002/j.2051-5545.2010.tb00255.x

Beck, A. T. (1979). *Cognitive Therapy and the Emotional Disorders*. Penguin.

Beck, A. T., Freeman, A., & Davis, D. D. (2004). *Cognitive therapy of personality disorders, 2nd ed* (pp. xx, 412). Guilford Press.

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond* (2nd ed). Guilford Press.

Borgers, S. B., & Koenig, R. W. (1983). Uses and effects of modeling by the therapist in group therapy. *Journal for Specialists in Group Work*, *8*(3), 133–138. https://doi.org/10.1080/01933928308411743

Buckley, P. (1994). Self psychology, object relations theory and supportive psychotherapy. *American Journal of Psychotherapy*, *48*(4), 519–529. https://doi.org/10.1176/appi.psychotherapy.1994.48.4.519

Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199596195.001.0001

Carruthers, P. (2013). Mindreading the self. In S. Baron-Cohen, M. Lombardo, & H. Tager-Flusberg (Eds.), *Understanding Other Minds: Perspectives from developmental social neuroscience* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199692972.003.0026

Castonguay, L. G., & Hill, C. E. (Eds.). (2012). *Transformation in psychotherapy: Corrective experiences across cognitive behavioral, humanistic, and psychodynamic approaches*. American Psychological Association. https://doi.org/10.1037/13747-000

Dellsén, F. (2020). Beyond Explanation: Understanding as Dependency Modelling. *The British Journal for the Philosophy of Science*, *71*(4), 1261–1286. https://doi.org/10.1093/bjps/axy058

Eells, T. D. (2015). *Psychotherapy case formulation* (pp. xvii, 232). American Psychological Association. https://doi.org/10.1037/14667-000

Ellis, A. (1963). Toward a More Precise Definition of “Emotional” and “Intellectual” Insight. *Psychological Reports*, *13*(1), 125–126. https://doi.org/10.2466/pr0.1963.13.1.125

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

Fiebich, A. (2021). In defense of pluralist theory. *Synthese*, *198*(7), 6815–6834. https://doi.org/10.1007/s11229-019-02490-5

Fileva, I., & Brakel, L. A. W. (2023). Understanding First: A Psychoanalytic Take on Self-Constitution. *Philosophy, Psychiatry, & Psychology*, *30*(3), 195–204. https://doi.org/10.1353/ppp.2023.a908267

Follette, W. C., Naugle, A. E., & Callaghan, G. M. (1996). A radical behavioral understanding of the therapeutic relationship in effecting change. *Behavior Therapy*, *27*(4), 623–641. https://doi.org/10.1016/S0005-7894(96)80047-5

Foucault, M. (1988). *Madness and civilization: A history of insanity in the age of reason* (Vintage Books Ed., Nov. 1988). Random House.

Frances, A. (2013). *Saving normal: An insider's revolt against out-of-control psychiatric diagnosis, DSM-5, Big Pharma, and the medicalization of ordinary life* (pp. xx, 314). William Morrow & Co.

Frank, J. D., & Frank, J. B. (1991). *Persuasion and healing: A comparative study of psychotherapy, 3rd ed* (pp. xviii, 343). Johns Hopkins University Press.

Friedman, L. (1978). Trends in the Psychoanalytic Theory of Treatment. *The Psychoanalytic Quarterly*, *47*(4), 524–567. https://doi.org/10.1080/21674086.1978.11926857

Gallagher, S. (2004). Understanding Interpersonal Problems in Autism: Interaction Theory as An Alternative to Theory of Mind. *Philosophy, Psychiatry, & Psychology*, *11*(3), 199–217. https://muse.jhu.edu/pub/1/article/174925

Gardiner, M. (Ed.). (1989). *The wolf-man and Sigmund Freud*. Karnac Books.

Gertler, B. (2021). Self-Knowledge. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2021/entries/self-knowledge/

Gilbert, P. (2010). *Compassion focused therapy: Distinctive features* (pp. viii, 237). Routledge/Taylor & Francis Group.

Godfrey-Smith, P. (2005). Folk Psychology as a Model. *Philosopher's Imprint*, *5*(6). http://hdl.handle.net/2027/spo.3521354.0005.006

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press.

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16*(1), 1–14, 29–113. https://doi.org/10.1017/S0140525X00028636

Grodniewicz, J. P. (ms). *Understanding oneself*.

Grodniewicz, J. P., & Hohol, M. (2023). Therapeutic Conversational Artificial Intelligence and the Acquisition of Self-understanding. *The American Journal of Bioethics*, *23*(5), 59–61. https://doi.org/10.1080/15265161.2023.2191021

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and Commitment Therapy: An experiential approach to behavior change.* Guilford Press.

Hoffer, A. (1985). Toward a definition of psychoanalytic neutrality. *Journal of the American Psychoanalytic Association*, *33*(4), 771–795. https://doi.org/10.1177/000306518503300402

Hutto, D. D. (2007). The Narrative Practice Hypothesis: Origins and Applications of Folk Psychology. *Royal Institute of Philosophy Supplement*, *60*, 43–68. https://doi.org/10.1017/S1358246100009607

Hutto, D. D. (2023). Narrative Practices in Mental Health: Narrative Therapy and the Fictive Stance. In J. Mildorf, E. Punzi, & C. Singer (Eds.), *Narrative and Mental Health: Reimagining Theory and Practice* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780197620540.003.0004

Hutto, D. D., & Gallagher, S. (2017). Re-Authoring narrative therapy: Improving our selfmanagement tools. *Philosophy, Psychiatry, & Psychology*, *24*(2), 157–167. https://doi.org/10.1353/ppp.2017.0020

Illouz, E. (2008). *Saving the modern soul: Therapy, emotions, and the culture of self-help*. University of California press.

Kelly, G. A. (1955). *The psychology of personal constructs. Vol. 1. A theory of personality. Vol. 2. Clinical diagnosis and psychotherapy* (pp. xxviii, 1218). W. W. Norton.

Krug, O. T., & Piwowarski, T. (2021). Ethical Issues in Existential-Humanistic Psychotherapy. In M. Trachsel, J. Gaab, N. Biller-Andorno, Ş. Tekin, & J. Z. Sadler (Eds.), *Oxford Handbook of Psychotherapy Ethics* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198817338.013.46

Kuyken, W., Padesky, C. A., & Dudley, R. (2009). *Collaborative case conceptualization: Working effectively with clients in cognitive-behavioral therapy*. Guilford Press.

Kvanvig, J. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press.

Lacewing, M. (2014). Psychodynamic psychotherapy, insight, and therapeutic action. *Clinical Psychology: Science and Practice*, *21*(2), 154–171. https://doi.org/10.1111/cpsp.12065

Magill, M., & Hallgren, K. A. (2019). Mechanisms of behavior change in motivational interviewing: Do we understand how MI works? *Current Opinion in Psychology*, *30*, 1–5. https://doi.org/10.1016/j.copsyc.2018.12.010

Maibom, H. (2003). The Mindreader and the Scientist. *Mind & Language*, *18*(3), 296–315. https://doi.org/10.1111/1468-0017.00229

Maibom, H. (2009). In Defence of (Model) Theory Theory. *Journal of Consciousness Studies*, *16*(6–7), 360–378.

Mameli, M. (2001). Mindreading, Mindshaping, and Evolution. *Biology and Philosophy*, *16*(5), 595–626. https://doi.org/10.1023/A:1012203830990

Marková, I. S., & Berrios, G. E. (1992). The meaning of insight in clinical psychiatry. *The British Journal of Psychiatry*, *160*, 850–860. https://doi.org/10.1192/bjp.160.6.850

McGeer, V. (1996). Is “Self-Knowledge” An Empirical Problem? Renegotiating the Space of Philosophical Explanation. *The Journal of Philosophy*, *93*(10), 483–515. https://doi.org/10.2307/2940837

McGeer, V. (2007). The regulative dimension of folk psychology. In D. D. Hutto & M. Ratcliffe (Eds.), *Folk Psychology Re-Assessed* (pp. 137--156). Kluwer/Springer Press.

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

McWilliams, N. (1999). *Psychoanalytic case formulation* (pp. xv, 240). Guilford Press.

Neisser, U. (1997). Concepts and self-concepts. In U. Neisser & D. A. Jopling (Eds.), *The conceptual self in context: Culture, experience, self-understanding* (pp. 3–12). Cambridge University Press.

Newen, A. (2015). Understanding Others: The Person Model Theory: The Person Model Theory. *Open MIND*. https://doi.org/10.15502/9783958570320

Nichols, S., & Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Clarendon ; Oxford ; Oxford University Press.

Norcross, J. C. (1990). An Eclectic Definition of Psychotherapy. In J. K. Zeig & W. M. Munion (Eds.), *What is Psychotherapy?* Jossey-Bass.

Orange, D. M. (1995). *Emotional understanding: Studies in psychoanalytic epistemology* (pp. x, 226). Guilford Press.

Perls, F. S. (1981). *The Gestalt Approach & Eye Witness to therapy* (5. print). Bantam Books.

Peters, U. (2019). The complementarity of mindshaping and mindreading. *Phenomenology and the Cognitive Sciences*, *18*(3), 533–549. https://doi.org/10.1007/s11097-018-9584-9

Plutynski, A. (2024). How is a Therapist like a Modeler? *Philosophy, Psychiatry, & Psychology*, *31*(2), 151–161. https://doi.org/10.1353/ppp.2024.a930491

Poland, W. S. (1988). Insight and the analytic dyad. *The Psychoanalytic Quarterly*, *57*(3), 341–369. https://doi.org/10.1080/00332828.1988.12021932

Prochaska, J. O., & Norcross, J. C. (2018). *Systems of psychotherapy: A transtheoretical analysis* (Ninth edition). Oxford University Press.

Purser, R. E. (2019). *McMindfulness: How mindfulness became the new capitalist spirituality*. Repeater.

Richfield, J. (1954). An analysis of the concept of insight. *The Psychoanalytic Quarterly*, *23*, 390–408.

Rogers, C. R. (1946). Significant aspects of client-centered therapy. *American Psychologist*, *1*, 415–422. https://doi.org/10.1037/h0060866

Segal, Z., Williams, M., & Teasdale, J. (2018). *Mindfulness-Based Cognitive Therapy for Depression, Second Edition*. Guilford Publications.

Spaulding, S. (2018). *How we understand others: Philosophy and social cognition* (First issued in paperback). Routledge, Taylor & Francis Group.

Strachey, J. (1934). The nature of the therapeutic action of psychoanalysis. *The International Journal of Psychoanalysis*, *15*, 127–159.

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., & Jongepier, F. (2018). Self-knowledge in psychotherapy: Adopting a dual perspective on one's own mental states. *Philosophy, Psychiatry, & Psychology*, *25*, 45–58. https://doi.org/10.1353/ppp.2018.0008

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (Second edition). Routledge.

Watters, E. (2010). *Crazy Like Us: The Globalization of the American Psyche*. Simon and Schuster.

Westra, E. (2021). Folk personality psychology: Mindreading and mindshaping in trait attribution. *Synthese*, *198*(9), 8213–8232. https://doi.org/10.1007/s11229-020-02566-7

White, M., & Epston, D. (1990). *Narrative Means To Therapeutic Ends*. WW Norton.

Winnicott, D. W. (1965). *The maturational processes and the facilitating environment: Studies in the theory of emotional development* (p. 295). International Universities Press.

Yalom, I. D. (1980). *Existential psychotherapy* (pp. xii, 524). Basic Books.

Young, J. E., Klosko, J. S., & Weishaar, M. E. (2006). *Schema therapy: A practitioner's guide*. The Guilford Press.

Zagzebski, L. (2008). *On Epistemology*. Cengage Learning.

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

Zawidzki, T. W. (2016). Mindshaping and self-interpretation. In *The Routledge Handbook of Philosophy of the Social Mind*. Routledge.

Zawidzki, T. W. (2018a). Mindshaping. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.39

Zawidzki, T. W. (2018b). Self-Interpretation as Software: Toward a New Understanding of Why False Self-Conceptions Persist. In P. Pedrini & J. Kirsch (Eds.), *Third-Person Self-Knowledge, Self-Interpretation, and Narrative* (pp. 115–144). Springer International Publishing. https://doi.org/10.1007/978-3-319-98646-3_7