Beyond Mentalizing: Epistemic Trust and the Transmission of Culture

Peter Fonagy & Elizabeth Allison

To cite this article: Peter Fonagy & Elizabeth Allison (2023) Beyond Mentalizing: Epistemic Trust and the Transmission of Culture, The Psychoanalytic Quarterly, 92:4, 599-640, DOI: 10.1080/00332828.2023.2290023

To link to this article: https://doi.org/10.1080/00332828.2023.2290023

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC

Published online: 14 Dec 2023.

Submit your article to this journal

Article views: 1466

View related articles

View Crossmark data

Citing articles: 5 View citing articles
We explore the interpersonal origins of human culture, arguing that culture emerges as a necessary consequence of our helplessness in infancy, which in turn requires a greater degree of collaboration and social organization than is necessary for other mammals. We propose a model of cultural transmission that depends on a dyadic interpersonal process whose vicissitudes can have a lifelong impact. We explore the role played by imagining subjectively experienced psychological states and processes in others, which we have defined as mentalizing, in the process of cultural transmission, and propose that mentalizing is key to the establishment of epistemic trust—that is to say, an experience
of trust that enables the individual to absorb and use the knowledge they are being offered. We suggest that psychoanalysis can be viewed as a paradigmatic example of a process of transferring knowledge from one human being (the teacher, the caregiver, the analyst) to another (the learner, the young person, the patient) and argue that the mechanisms of psychic change in analysis cannot be fully understood without appreciating this aspect of the biological/evolutionary origins of our essential humanity. Finally, we discuss the clinical implications of the model we are proposing for the psychoanalytic process.

Keywords: Mentalization, culture, evolutionary theory, natural pedagogy, communication, mechanisms of therapeutic change.

INTRODUCTION

The aim of this paper is to create a link between modern anthropological and psychological understanding of culture and the nature of psychic change in psychoanalysis. This effort is based on a value-laden belief that psychoanalysis as Freud conceived it represents the essence of what it is to be human: the capacity to engage in meaningful communication that entails the transmission of knowledge. Williams (1983) described culture as “one of the two or three most complicated words in the English language” (p. 87). The American anthropologists Kroeber and Kluckhohn (1952) critically reviewed concepts and definitions of culture seventy years ago and compiled a list of 164 different definitions. Claiming no particular expertise, we consider culture as sustained by the transgenerational transfer of knowledge, a continuous cycle of learning and teaching that spins the thread of human history. With greater confidence, within this framework we consider mental disorder as a catastrophic disruption of this learning process. Both culture and normal individual development depend on an interpersonal facet of this process, honed by natural selection to achieve an efficient and reliable social mechanism for knowledge transfer. As it is ordinarily practiced, psychoanalysis has this process under a microscope, and insights from psychoanalytic writings shed light on this most important of human functions even when not intended to illuminate it.
In this paper, we will reframe our earlier speculations concerning the nature of consciousness as revealed by its disruption in a patient with severe personality pathology (Fonagy & Allison 2016) in the context of our subsequently developed theory of epistemic trust and its significance for the transmission of culture (Allison & Fonagy 2016; Fonagy & Allison 2014; Fonagy et al. 2021). We consider individual development alongside speculations concerning human evolution and elaborate a model of human communication that focuses on trust as a motivator of enduring knowledge acquisition. We suggest that the psychoanalytic process exemplifies an evolutionary biological model of learning. We suggest that the notion of lasting psychic change as embodied in the metaphor of structural change may be conceptualized as permanent modifications in the way an individual acquires social knowledge relevant to their current life circumstances. We will begin by recapping our previous communication about consciousness.

In our 2016 paper on consciousness, we argued that (reflective) consciousness evolved to enable us to share our experiences (Fonagy & Allison 2016). We suggested that consciousness creates units or unity to enable communication, and thus, self-awareness is intersubjective in its essence. Without the ability to create coherence within our world of experience, it is inconceivable that we would be able to communicate with fellow members of the species. We proposed that the capacity to experience ourselves as conscious, intentional agents in a coherent world of objects is not merely a modality of perception (Freud 1924; Solms 1997); rather, it is a maturational and developmental achievement that to some degree depends on adequate experiences of caregiving. Consciousness is socially created. By watching other people’s reactions, the infant learns which elements of his idiosyncratic and inchoate experience other people are also sensitive to, and these elements become the building blocks of his sense of self. This is the process of *marked contingent mirroring*, which we have described in many earlier papers (Fonagy, Gergely & Target 2007; Nichols, Gergely & Fonagy 2001; Watson, Futo, Fonagy & Gergely 2011).

Of course, the corollary of this is that only those internal states that are confirmed by contingent, marked, and congruent responses from the object world will achieve adequate representation. We argued that the infant’s unmirrored and unreflected self-states come to constitute
what Freud described as the seething cauldron of the id (Freud 1933, p. 73). We suggested that these unmirrored states are of two kinds. Sexuality and aggression are difficult for the caregiver to mirror but still produce a contingent response that allows them to be experienced as personal and directed, although they are not represented fully consciously (Fonagy 2008). Through this process, the psychoanalytic unconscious is constituted. We defined this as “the intrusive, disturbing, disruptive mental contents that we as psychoanalytic clinicians struggle to decode, interpret and work with” (Fonagy & Allison 2016, p. 97). But there is also a second kind of unmirrored state: that which cannot be mirrored because its fragmentary character precludes the possibility of creating a genuine phenomenological inner horizon (Husserl 1931; Merleau-Ponty 1945). It is non-experiential and is beyond the unseen or unattended dimensions of an experience, which, although not directly in focus, significantly shape the overall understanding and interpretation of that experience. These states constitute what we referred to as the primary unconscious. Strictly speaking, they cannot be described in words because they entail an undoing of the synthetic processes by means of which objects are constituted as such. We can approach them by saying that they are destructive of the coherence and intentionality constitutive of consciousness, but this language still implies an object relationship, and in our view the primary unconscious is not object related. For the sake of simplicity, we used the word negativity to talk about these states of mind. This idea is not entirely new to psychoanalysts (e.g., Eigen 1996; Grinberg 1964; Grotstein 1991; O’Shaughnessy 1999) but its couching in the context of infant and early childhood development may be more controversial. Normally, as Sandler and Sandler (1983) also pointed out, the content of the primary unconscious remains forever far from conscious subjectivity. However, we argue that

1 “[The id] is the dark, inaccessible part of our personality; what little we know of it we have learnt from our study of the dream-work and of the construction of neurotic symptoms, and most of that is of a negative character and can be described only as a contrast to the ego. We approach the id with analogies. We call it a chaos, a cauldron full of seething excitations.”

2 In Husserlian terms, every perceived mundane entity has by necessity an inner horizon (Innenhorizont), a horizon of not-yet-manifest and no-longer-manifest aspects of the same entity—and as we have seen, without such an inner horizon, we would not perceive any mundane entity at all (Overgaard 2004).
in cases of severe trauma and profound neglect, part-mirroring of the negativity of the primary unconscious does occur, and this can have a disastrous impact on mental functioning.

We illustrated these ideas with the case of Mr. K. who was a chronically suicidal, delusional, and at times a frankly psychotic patient. As a child, he was profoundly neglected by parents whose fundamentalist religious ethic involved self-deprivation and condemnation of emotion, considering it an unacceptable deviation from a righteous path. When he came to analysis, Mr. K’s central ambition was to end his life, and his only means of dealing with emotional perturbation was through callous acts of violence toward himself or others. His capacity for reality testing was tenuous, and his experience of both the physical and psychological worlds were delusional, persecutory, hallucinatory, and bizarre. He exhibited a notable lack of self-reflection and continuity in his thought processes. He repeatedly brought the same thoughts and feelings to each session, almost as if experiencing them afresh each time. He coped with internal turmoil through starkly narrated violent acts toward himself or others. His sessions were steeped in a sense of ennui.

Here is a brief illustration of working with Mr. K. Upon arriving at a Monday session, he was overwhelmed by unexpressed, intolerable emotions. He detailed his debilitating weekend marked by prolonged sleep, bouts of crying, overeating to the point of sickness, and an inability to function without his analyst. In response to the analyst’s observation of his solitary struggle with his weekend memories, he acknowledged, yet overlooked, his comment and dismissively recounted a dream. In his dream, Mr. K was searching for someone in a desolate building, opening doors to empty rooms. He insisted that the dream’s message was transparent—there was nothing in life for him and he should be dead. He imagined that the analyst’s other patients provided more fascinating insights. The analyst suggested that his dream emphasized his emptiness, making him feel unique. He expressed both a paradoxical need for the analyst and a belief in his utter inadequacy. Asserting that it was too late for him, he likened himself to a black hole consuming any amount of care. As he presented this image, the analyst was left feeling insignificant and powerless, engulfed by despair about their situation. His complaint about the analyst’s silence prompted him to voice his countertransference experience of Mr. K’s existential dread. The analyst suggested that he feared pulling
them both into this black hole and worried that it might be too late. He grudgingly acknowledged the analyst’s understanding, but the ephemeral relief soon gave way to the overpowering weight of Mr. K’s despair.

Mr. K’s perception of the world was markedly skewed. He reported fragmented experiences, with objects seeming to alter when he looked away, suggesting a lack of basic object constancy. Minor changes in analyst’s room, like a tidied desk, could provoke panic. He often described experiences in terms of physical sensations rather than cohesive narratives. His world was perceived as devoid of appeal and filled with insecurity and hostility: barren, hostile, and indifferent. The analyst’s and patient’s emotional experiences of their shared physical environment were starkly different. Mundane office chairs in the analyst’s room took on terrifying forms in Mr. K’s mind, resembling electric chairs or even medieval torture devices. This distortion pervaded his entire consciousness, leading to a desire to escape his unbearable reality. He felt engulfed by death, nothingness, meaninglessness, and chaos. Experiencing Mr. K’s death-imbued consciousness in the countertransference seemed to infiltrate the therapeutic space. Recognizing his pervasive negativity not as a stance towards his analyst or the world but rather as a disruption of consciousness, greatly simplified his analyst’s understanding of his condition.

In Mr. K’s case, his analyst’s efforts to apply his general principle of enhancing mentalizing, the principle that the therapeutic engine of psychoanalysis is the patient’s recognition of aspects of his own mind of which he is unaware in the therapist’s mind, failed dismally. Mr. K got worse, not better. The analyst’s efforts to introduce a sense of intentionality and coherence to the contents of Mr. K’s unconscious mind actually further undermined Mr. K’s mental functioning. The overwhelming feeling of being surrounded by death, nothingness, meaninglessness, and chaos that Mr. K brought into analysis was a consequence of his traumatic early experiences—in which the negativity that normally remains far from conscious subjectivity in the primary unconscious had been mirrored—allowing it into consciousness where it could be manifest only as anguish and disruption. When the analyst sought to help Mr. K to mentalize these states of mind, this only further destabilized Mr. K’s capacity to achieve the coherence and intentionality required to produce sharable conscious experience.
In order to help Mr. K, the analyst first had to grasp that his negativity did not express an intentional attitude toward him or the world at large; rather, it signified how profoundly disrupted his capacity to experience coherence and intentionality was. Thus, insight into motives or mental states in self and other surrounding such distress and incoherence could be of no benefit to him. On the one hand, trying to mentalize this negativity exacerbated its impact, and on the other there was no way to repress, regulate, or mute the experience—no way to unmirror that which should never find reflection in external reality; all that could be done was to help Mr. K live with it in a slightly more comfortable way. Realizing that his negativity across all situations was more a function of an internal distortion within Mr. K’s mind rather than a transference communication, allowed the analyst to distance himself and cultivate empathy for Mr. K’s struggles. The most pivotal therapeutic advance, and a prerequisite for further self-understanding, was to establish a genuine alliance with Mr. K, recognizing that he and his analyst were jointly battling a force greater than either of them. His experience of his unconscious mind could not simply be unmirrored or made unexperienced. At the time of writing the earlier paper, we attributed progress to the analyst’s effort to help Mr. K make conscious objects out of his monstrous subjective universe. The negativity was variously represented as a magma, a monster, or a robot with which he needed to learn to coexist; it was and would always be part of him but he could live with it in a way that was less incapacitating. This involved slowly defining the nature of Mr. K’s pervasive negativity, identifying when it surfaced during his sessions and when it emerged outside of therapy. We metaphorically referred to this as Mr. K’s badness feelings, likening it to the earth’s mantle, ubiquitous in everyone but particularly active in him due to a unique vulnerability that allowed it to surface. Externalizing this badness provided Mr. K with a sense of liberation he had never before experienced. While some may perceive this approach as collusive, it was perhaps less so than a direct interpretive approach, which Mr. K tended to co-opt as part of the destructive forces against his survival. The next phase involved Mr. K befriending his negative state, a process that required having a label for these unwanted conscious experiences. This collective engagement with the unconscious differed from the typical interpretive approach used with neurotic patients. Rather than striving to understand or situate the negativity within a narrative, the patient and analyst acknowledged its
presence. The therapeutic objective was to transform Mr. K’s overwhelming subjective universe into conscious objects, adopting a broad, rather than a targeted, analytic narrative. Surprisingly, this approach proved both feasible and relatively harmonious. We argued that this effort to expand and contain Mr. K’s consciousness was a key (albeit not the sole) agent of change. Acceptance and some insight into his passive-aggressive, masochistic posture also played their part. However, our firm belief was that it was not an alteration in unconscious processes that signaled the change, but rather a shift in conscious function—an accommodation his consciousness made to his pervasive negativity, facilitated through the approach just described. We would now see the analytic process as involving more than this.

In the 2016 paper, we did not specify clearly enough the role played by consciousness in bringing about change. In the current paper we return to Mr. K and further develop the theoretical model in order to try to redress this omission. Our starting point is the argument previously advanced that consciousness is essential to the human capacity to communicate, collaborate, and generate transgenerational culture. Without consciousness, thoughts, beliefs, ideas, and knowledge essential for individual survival cannot be shared and passed on to the next generation. Our proposal that consciousness is socially created implies that our early relationships with our caregivers are the crucible of culture. In this paper, we will argue that the quality of these relationships plays a key role in establishing our capacity to learn and thus to benefit from our cultural inheritance, and to contribute to the cultural resources that will be available to future generations.

With a few distinguished exceptions (e.g., Bion 1961; Freud 1921; Hinshelwood 2008; Kernberg 1998; Shapiro & Carr 1991), psychoanalysts have tended to refrain from extensive discussions of culture, seeing this as entering the treacherous territory of applied psychoanalysis and preferring to remain within the sphere of individual psychology, firmly focused on inner psychic reality. Yet there has also always been some degree of acknowledgement that seeing the psychoanalytic situation solely as a setting for the exploration of individual psychology is a misrecognition. As Freud (1921) wrote in Group Psychology and the Analysis of the Ego.
In the individual’s mental life someone else is invariably involved, as a model, as an object, as a helper, as an opponent; and so from the very first individual psychology, in this extended but entirely justifiable sense of the words, is at the same time social psychology as well. [p. 69]

In “Group Dynamics: A Re-View,” which became the final chapter of *Experiences in Groups*, Bion (1952) concurred that it was misleading to think in terms of a difference in kind between individual and group psychology. He wrote:

In my view no new instinct is brought into play [in groups]—it is always in play. It is not necessary for a number of people to be brought together—the individual cannot help being a member of a group even if his membership of it consists in behaving in such a way as to give reality to the idea that he does not belong to a group at all…. The individual is a group animal at war, both with the group and with those aspects of his personality that constitute his ‘groupishness’. [p. 238]

Thus, Bion also recognized the implicit but inevitable presence of the social world and therefore culture within all individuals.

We do not tend to think of Winnicott as a theorist of culture, but in his 1967 paper “The Location of Cultural Experience” he develops his thinking about the transitional phenomena of infancy in order to consider the nature of cultural experience (Winnicott 1967). His wish to explore this domain derives from his sense that psychoanalysts “have not yet started to describe what life is like apart from illness or absence of illness. That is to say, we have yet to tackle the question of what life is about” (p. 370). Culture may have disappeared from psychoanalytic vision just as fish (as imagined by David Foster Wallace) fail to perceive the water they swim in.³

Winnicott’s (1967) definition of culture is characteristically idiosyncratic. He writes:

³ Wallace’s commencement address to the 2005 graduating class at Kenyon College opened with the following story: “There are these two young fish swimming along, and they happen to meet an older fish swimming the other way, who nods at them and says, ‘Morning, boys, how’s the water?’ And the two young fish swim on for a bit, and then eventually one of them looks over at the other and goes, ‘What the hell is water?’”
In using the word culture I am thinking of the *inherited tradition*. I am thinking of something that is in the common pool of humanity and into which individuals and groups of people may contribute, and a source from which we may all draw *if we have somewhere to put what we find.* [p. 370]

This “somewhere” is of course the self, which (if all goes well) is gradually established in Winnicott’s model through the caregiver’s finely tuned interactions with the baby. Thus, Winnicott sees the caregiver as the vehicle for the transmission of culture to the infant. A crucial stage in the process of carefully titrated separation and reunion is the infant’s employment of a transitional object, which “symbolizes the union of two, now separate things, baby and mother, at the point of the initiation of their state of separateness” (Winnicott 1967, p. 369). Winnicott was well aware that the acquisition of one’s cultural inheritance was inherently object related and could not be detached from the emotional context within which identification with one’s social niche occurs. Thus, when Winnicott speaks about culture, this is his attempt to address psychoanalysts’ neglect of the way in which meaningful or intimate interpersonal experience gives us access to social experience considered more broadly.

Winnicott concludes the paper with the suggestive observation that the outcome of good enough early experience is that it builds a capacity to trust, which in his view is essential to enable the individual to participate in cultural life. In this context, he quotes the Jungian analyst Fred Plaut’s (1966) statement that “The capacity to form images and to use these constructively by recombination into new patterns is— unlike dreams or fantasies—dependent on the individual’s ability to trust” (quoted by Winnicott 1967, p. 372). In the US, Erik Erikson had argued as early as 1950 that the development of basic trust as an outcome of early psychosocial experience laid the foundations of a healthy personality (Erikson 1950). It is also important to acknowledge the US interpersonal psychoanalytic tradition’s recognition of the role played by significant others in the constitution of a sense of self (e.g., Sullivan 1953; Bromberg 2009; Stern 2015; Frie & Sauvayre 2022). We are perhaps most deeply indebted to the brilliant Jessica Benjamin (1990), who—in a paper written over 30 years ago—pointed out that object relations theories, even those interested in intersubjectivity, have overlooked
the core element of intersubjectivity, which is mutual recognition. She showed that the capacity for mutual recognition can be conceived as a separate trajectory from the internationalization of object relations. The subject gradually becomes able to recognize the other person’s subjectivity, developing the capacity for attunement and tolerance of difference. As we now suggest, the differentiation of self and other is a process that consists not merely of separation but of the continual breakdown and repair of mutuality in the psyche’s stance toward the interpersonal world beyond the self. The sense of the other comes and goes; the breakdown and reconstruction of the interpersonal world is an endless experience of aggression and the creation of an intrapsychic fantasy of mutuality. In general, we owe much to the (now classical) psychoanalytic intersubjectivity approach that shares key elements with our developmental theory of mentalizing (Benjamin 1995, 1998; Ogden 1994; Renik 1993).

We will begin by drawing on archeology, anthropology, developmental psychology, and neuroscience to propose a model of cultural transmission that, as Winnicott suggested, depends on a dyadic interpersonal process whose vicissitudes can have a lifelong impact. We will argue that culture emerges as a necessary consequence of our helplessness in infancy, which in turn requires a greater degree of collaboration and social organization than is necessary for other mammals. We think that mentalizing and the capacity to experience ourselves as conscious, intentional agents in a coherent world of objects play an important role in the process of sharing of experience which cultural transmission entails. We now think that we can be more precise about why mentalizing matters. We will argue that mentalizing is important because it allows the establishment of what we call epistemic trust, that is, an attitude of openness that enables the individual to absorb and make use of the knowledge they are being offered. We would now regard Mr. K’s agony as part of a state of devastating epistemic isolation and separation from meaningful social connection in a state of epistemic hypervigilance. We now think that the psychoanalytic process achieved much more than we initially assumed. Within the framework we are proposing, we would suggest that psychoanalysis reduced Mr. K’s chronic lack of epistemic trust and enabled his relative return to normal functioning by restoring a degree of reliance on his social world. More broadly, we contend that an aspect of the entire psychoanalytic and psychotherapeutic enterprise can be
seen as one, albeit perhaps paradigmatic, example of a process of transferring knowledge from one human being (the teacher, the caregiver, the analyst) to another (the learner, the young person, the patient). It is easy to overlook the fact that this is not an automatic cognitive process but one that is powerfully conditioned by biologically primed ways of evoking feelings of trust. We argue that the mechanisms of psychic change in analysis cannot be fully understood without appreciating this aspect of the biological/evolutionary origins of our essential humanity. Finally, we will discuss the clinical implications of this model for the psychoanalytic process.

AN ANTHROPOLOGICAL PERSPECTIVE

Our working definition of culture is derived from Daniel Sperber (1990) who encapsulated it in the phrase “culture is the precipitate of cognition and communication in a human population” (p. 42). Animal societies do not show steady adaptive change or cultural progress. In the animal kingdom, the oldest individual accumulates knowledge. The 1993 drought in Tanzania variably affected elephant families. The ones that survived best were those led by matriarchs who remembered the severe drought of 1958. But there is no durable transgenerational transmission of this knowledge for elephants. It goes to the graveyard with them. Humans possess an evolutionarily unparalleled capacity to transfer information between agents. This enables remarkable levels of cooperation between partners across infinitely variable environments and knowledge domains. The human mind is not just a product of the social, rather it is the social (Gallotti & Frith 2013). A comprehensive model of cultural evolution needs to take into account the development of social as well as individual aspects of cultural behavior, considering a range of cognitive capacities including, but not limited to, human cooperation (Tomasello, Melis, Tennie, Wyman, & Herrmann 2012), group conformity (Claidiere, Bowler, Brookes, Brown & Whiten 2014), memory (Causey & Bjorklund 2014), and the social transmission of information (Csibra & Gergely 2011).

The primatologist, linguist, and developmentalist Michael Tomasello described the process of knowledge accumulation as the ratchet effect. Tomasello et al. (2005) argue that the ratchet effect (adopting new
elements that are superior to and enable us to dispense with prior solutions) allows the accumulation of knowledge within a social group to be transmitted through interpersonal learning. An idea is retained within a social group until a better solution to the same problem comes along. This is what natural selection prepared us for, the better to ensure the survival of our DNA. Modern evolutionary theory recognizes a mutual interaction between genetic and cultural evolution (dual inheritance theory [Richerson & Boyd 2005]) such that genes and culture co-evolve. We prioritize not only our own survival but also social learning and adaptive fitness in our children, because our DNA survives through them. Each new human depends on their caregivers to help them to acquire the skills necessary to enable them to navigate the physical world safely. But, as our survival depends on collaborative effort, we also need to learn about our world to enable working in partnership, which in turn demands that we share our experiences. We can then form significantly larger and more complex communities than either our ancestors or our nearest animal relatives.

The question of how and when the ratchet effect of human culture got started has been extensively debated (e.g., Lind 2013; Klein 2003; Haidle 2014; Baron-Cohen 2020) and is beyond the scope of this paper to answer. However, whatever the precise timeline, the human capacity to create complex but beneficial cultural objects seems likely to be linked to the emergence of mentalizing in evolutionary history, since the ramping up of the demand characteristics of “cognitively opaque cultural forms” (Gergely & Csibra 2006, p. 8) compels a parallel process, ensuring the transmission of knowledge (the replication of memes). Mentalizing provided the necessary teaching function that increasingly complex technology demanded and was an essential requirement to ensure the preservation of human culture. At the core of the concept of culture is the common underlying theme of transmission of knowledge across generations: the inherited tradition, to use Winnicott’s phrase. In order to enable earlier achievements to be retained and built upon, a robust process for education is required.

A helpful set of distinctions is offered by Granqvist (2021) in describing the interface of attachment and the evolution of human culture. Modern biology emphasizes the importance of culture. In contrast to the classical biology that traditionally informed psychoanalysis
(Sulloway 1979), contemporary biologists no longer consider that brain and behavior are the consequences of a past environment of evolutionary adaptation, as depicted for example in Freud’s (1921) speculations about the phylogenetic origins of the Oedipus complex. From the classical perspective, substantial aspects of human behavior, such as the vicissitudes of the instincts, are seen as having roots in our prehistoric past, and Freud’s biology searches for elemental components of humanity in its evolutionary origins. By contrast, modern biology suggests that we have evolved to retain the potential to adapt to current local conditions (Davies, Krebs & West 2012; Laland & Brown 2011). Human beings are genetically underspecified. We are shaped by our ecology and the biological perspective may be as relevant to social adaptation as to genetic natural selection. Many biologists have therefore become interested in the biological principles governing culture, which reflects an adaptation to ecological conditions (Boyd 2017).

Explaining ecology-specific cultural differences requires conjecture in relation to a mechanism of cultural transmission wherein ways of doing things can be transmitted from one generation to another. There are at least two schools of thought in relation to cultural inheritance in evolutionary science. The first, known as memetics, focuses on culture and how ideas nested within particular cultures—memes, for example, beliefs in particular deities—are disseminated. However, within this approach, cultural transmission and possible evolutionary mechanisms to ensure the perpetuation of traditions are underemphasized in favor of an extension of natural selection principles to memes—in other words, the ideas best fitted to a particular environment survive (Dawkins 1976, 2004). Thus, ideas struggle for survival to reproduction (inclusive fitness) just as physical organisms do. Clearly, this approach is consistent with the rapid emergence of ideas, which effectively, mainly pertains to the 12,000 years since the agricultural revolution. While this perspective from natural selection explains the desperation with which we can find ourselves holding on to ideas—fighting and even dying for them—it assumes an inherently improbable construct of a tabula rasa mind, since the individual has to be ready to absorb whatever ideas survive to be dominant in their culture; and this would be incompatible with, for example, a psychoanalytic model rooted in instinct theory.
Our contribution could be seen as part of dual inheritance theory, also known as gene-culture coevolution, which posits that human evolution is driven by a dynamic interplay of both biological (genetic) and cultural (learned) factors (Durham 1991). Unlike other evolutionary theories that primarily focus on biological transmission of traits, dual inheritance theory acknowledges that humans are shaped not only by their genetic makeup but also by cultural influences, with both avenues of inheritance interacting and influencing each other. For instance, certain cultural practices may drive genetic selection over time and, conversely, genetic predispositions can shape cultural norms and behaviors. By considering both of these transmission pathways, dual inheritance theory offers a more comprehensive understanding of human evolution and adaptation, encompassing the complexities of human nature and societies (Laland & Brown 2011).

As Granqvist (2021) emphasizes, more recent biology does not posit that ideas evolve or that the notion of inclusive fitness can be extended to thoughts and beliefs. Rather, the increased complexity of ideas is thought to be explained by a selection pressure for ever-increasing flexibility to allow for adaptation to ever more complex environments (Laland 2017). The flexibility that we have evolved enabled the process of industrialization in the West and the move from small villages to large cities as the prototypical environment. The selection pressure for flexibility generates increasingly complex cultural frameworks of belief, which in turn create new kinds of selection pressures that favor even higher levels of flexibility because of an underlying selection pressure for ever-greater capacity to adaptation to an ever more rapidly changing ecology. The question then becomes what is the biological mechanism, what form of bias to learning (Richerson & Boyd 2005), was selected by evolution that can ensure an ever-increasing flexibility of adaptation to an ever-accelerating pace of change and complexity of environments? Our answer to

4 The positively accelerating rate of change in our environment motivated by increased behavioral capacity for flexibility, which in turn enables the creation of even more complex environments with yet greater demands is ultimately a catastrophic scenario that entails its own destruction when the environments generated will be too complex for the evolved human capacity for adaptation. Indeed, climate change may be just this nemesis.
this question lies in the infinite resources offered by language and human communication.

**THE FUNDAMENTAL CHALLENGE TO COMMUNICATION: TO TRUST OR REMAIN VIGILANT**

*In order to communicate effectively, we need to know when to trust and when to remain vigilant.* We have always stressed the importance of mentalizing both for mental health and the successful navigation of the social world. We have even gone so far as to agree with those who claim that mentalizing can be seen as that which defines humanity and separates us from other higher primates (hominids such as the orangutans of Sumatra and Borneo, the gorillas or the chimpanzees and bonobos of Central and West Africa) (Heyes & Frith 2014).

The notion of *mind* enables us to interpret events that we observe in the external world in terms of the experience and intentions of psychological agents, rather than attributing them to physical laws or chance. Limitations of mentalizing can make an individual overly dependent on appearance or concrete reality, leading to enactments such as self-harm and suicide as well as misunderstandings of others’ intentions that generate needless interpersonal conflict. Moreover, inadequate understanding of one’s own states of mind could not only lead to a profound absence of empathy in relation to others but also to oneself, creating an extreme sense of guilt and shame and ultimately a sense of emptiness, the pain of which is almost impossible for most of us to fathom.

We have adopted a developmental perspective to conceptualize these difficulties that owes something to Winnicott (1956) and Bion (1962), both of whom expanded the capacity of psychoanalytic discourse to take into account the interpersonal matrix out of which the psychological self arises. The mentalizing model suggests that our understanding of ourselves emerges from an internalization of the understanding that others (in the first instance the primary caregiver) have of us. In

---

5 Mentalizing is the capacity to envision that the actions of others are determined by their mental states, their beliefs and feelings and, indeed, that one’s own actions are driven by one’s cognitions and the emotions linked to them.
order to understand them, however, we rely on self-understanding. If our self-understanding is good enough, this creates the potential to establish a virtuous cycle. The better I understand myself, the better I will be able to detect my caregiver’s attempt at understanding me; perceiving myself in the caregiver will enable me to elaborate my understanding of myself, in turn improving my capacity to grasp what she is teaching me about myself, and so on. The knowledge we acquire about who we are is absorbed from others, kept current, updated in response to changing social contexts, enabling the social adaptation crucial to human survival. We build and constantly unconsciously update representations of ourselves through our social interactions, which are imbued with learning opportunities. As Joseph Sandler (1962) might have put it, the social context influences the shape of our self-representation, that is, the particular form and character assumed by that representation or image in the representational world at any one moment. Sandler later modified this formulation in order to clarify that what is represented is the self in interaction with an object (Sandler 1993a).

All this remains true, and we would still maintain that mindreading is pivotal to self-regulation and the emergence of a fully functioning mental self. However, in the past we may well have given the impression that good mentalizing was an end in itself. In earlier papers, we have suggested that the aim of analytic treatment with severely personality disordered patients is (or ought to be) to develop and consolidate the patient’s capacity to maintain and, if necessary, recover a mentalizing or reflective mode of experience of psychic reality even in the face of traumatic emotional upheaval. Yet a few moments’ reflection will show us that it is possible to think deeply and insightfully about our own internal world and enter imaginatively into the experiences of others and yet remain completely stuck in profoundly unsatisfactory life situations; we all know individuals who might be excellent mentalizers but could still do far better both in love and at work. We have had them in treatment and perhaps occasionally even as colleagues. While mentalizing may be part of the answer, demonstrably it is not the whole answer to enabling psychic change.

More recently, we have sought to clarify our account of the significance of mentalizing for psychic development and entry into the social world and to articulate its importance in the therapeutic process more precisely
We have argued that mentalizing is important because it confers *epistemic trust*: the attitude we need in order to benefit fully from opportunities to learn from others—to learn about the world in general and the social world, including ourselves, in particular. The human world, outside and inside, is so complex that it is inefficient, impossible, and dangerous to try to figure it out all by ourselves, and yet taking everything we are told at face value is a high-risk strategy. What Sperber et al. (2010) called *epistemic vigilance* is an appropriate attitude to guard us against the risk of being misled. We are more likely to survive and thrive if we do not forget that not everyone will teach us things that are true or indeed in our interest to learn. Yet we do need to be able to learn from others, and trust is required to open up our naturally selected channel for knowledge transmission. Both the capability for vigilance and a mechanism for selectively circumventing it must be deeply etched into our human origins. Trust in knowledge (which, following Sperber, we call *epistemic trust*) is at the heart of what it means to be human. Once trust has been established, we are open to internalize and assimilate the knowledge offered by the trusted or deferentially regarded source in order to optimize our future interactions with our social world (Recanati 1997; Sperber et al. 2010; Wilson & Sperber 2012). This is a more complex process than it may appear.

As we have seen, in order to safely depend on others to learn about reality, we need to be able to identify those who are reliable sources of information. The young human needs to be able to distinguish trustworthy, benevolent, and reliable sources of knowledge from communicators who are either poorly informed or malevolent. In either case, the latter are the purveyors of useless or deceptive information. Thus, in order to ensure effective cultural knowledge transfer via teaching, humans needed to evolve a reliable way of distinguishing trustworthy sources of knowledge.

**THE DEVELOPMENTAL VIEW OF ACHIEVING AND MAINTAINING EPISTEMIC TRUST**

As with most puzzles of acquired human function, the developmental perspective yields important insights. How can a young human infant,
completely naive to the complex and potentially dangerous world he or she was brought into, know who to trust? Does learning invariably imply a social process? Could a child acquire knowledge of their culture by passively observing the actions of others or is an adult paying specific attention to the child an essential component of the process?

The assumption that children turn to adults as key sources of knowledge about the cultural tools they need makes human cognition a unique dyadic interpersonal process, as Winnicott, Vygotsky, and many developmentalists since have recognized (Bronfenbrenner 1979; Fonagy & Target 1996, 2000; Target & Fonagy 1996; Vygotsky 1978; Winnicott 1971). There are two complementary accounts of how children learn from others, both relevant to our understanding trust in knowledge. The first, which involves intentional understanding, assumes that in child-directed interactions a joint attentional focus is created that generates a sense of a shared goal and a joint plan of action and in the course of this the infant registers important aspects of what the adult is experiencing (Moll & Tomasello 2007). It is argued that a mutuality is created through this joint engagement as children acquire a conceptual understanding of the other’s actions through sharing an attentional focus and their end goal (Barresi & Moore 1996). I catch the child’s eye, look over to an object, and say “That is a cup.” Children filter their social learning through the analysis of the intentions of others (Baldwin, Baird, Saylor & Clark 2001; Tomasello et al. 2005). Mutuality supports learning from others by ensuring that the child understands the communicative intentions of the adult (Moore 2010).

A slightly different model was proposed by Gergely Csibra and George Gergely (Csibra & Gergely 2006, 2009, 2011). They demonstrated that when the communicators made a special effort to show that they recognized the infant as an individual, calling them by their name, smiling at them, responding sensitively to their actions in an attuned manner, the infant took the knowledge they communicated as relevant to them, to be remembered and generalizable to other similar situations. The model of natural pedagogy suggests that ostensive signals (Russell 1967) preceding or accompanying such interactions trigger an innate learning system, the biological mechanism selected by evolution to ensure flexible adaptation (Laland 2017), which enables the child to acquire knowledge from a single act of demonstration. This is because the ostensive signals
(attunement, smiling) have alerted the infant that the content of the demonstration is relevant for them to take on board, it is part of cultural knowledge, and is generalizable to other contexts (similar objects, similar occasions, different people). Ostensive signals open the way for the transmission of important generalizable knowledge and indicate that it is safe for the child to relax their natural and adaptive epistemic vigilance. A channel for learning is opened through the creation of epistemic trust. Experimental evidence suggests that very young children are more likely to imitate another person’s actions accurately when they have been directly addressed than when they just incidentally observe (Sage & Baldwin 2011). This difference is not explained by the amount of attention that the child pays to directed versus just observed events (Shneidman, Todd & Woodward 2014). There is experimental evidence to suggest that generalizing to contexts beyond the demonstration depends on the presence of ostensive or child-directed cues that lead children to infer that the information is generalizable (Egyed, Király & Gergely 2013; Trauble & Batz 2014).

So much for infants. But what is the equivalent of ostensive signaling beyond childhood? How do we open our adult mind to influence? How is epistemic vigilance lowered and epistemic trust created in older children and adults? With Patrick Luyten and Chloe Campbell, we (Allison & Fonagy 2016; Fonagy & Allison 2014) have advanced the view that all relationships to a varying extent entail the flexible activation of epistemic trust (Fonagy, Luyten, et al. 2017b). Our openness to long-term influence is a function of the extent to which our default of epistemic vigilance has been overcome. We suggest that a particular form of ostensive cueing remains relevant throughout life. The experience of mutual recognition of intent triggers the learning system envisioned by Gergely and Csibra, indicating to the individual that a special opportunity has arisen to acquire durable and significant cultural knowledge. Mutuality of the kind envisioned by Tomasello and Moore can serve as part of ostensive signaling, establishing a shared intentional structure in learner-directed actions and triggering the learning system that Gergely and Csibra describe.

Children and adults do not identify reliable sources of knowledge through logical deduction but through an act of imagination of a particular kind. Epistemic trust is established though the learner identifying the instructor’s intentions as benign in that they show awareness of the
learner’s personal narrative. As described above, the knowledge we acquire about who we are is absorbed from others, kept current, and updated in response to changing social contexts, enabling the social adaptation crucial to human survival. In Sandler’s language, it is the shape that is formed and re-formed as a product of the self’s interaction with an object (Sandler 1993b). We prefer to use the term personal narrative to describe what Sandler pointed to as the experientially accessible characterization of the uniqueness of one’s being. The learner identifies the instructor’s image of the learner’s self-narrative by mentalizing. This, as we have always stressed, is an imaginative process. Thus, mentalizing is not an end in itself; it is the critical tool that enables us to recognize others as worthy of trust and safe to learn from. One is recognized and therefore one can trust. Of course, prior to this, it is also critical to the instructor’s capacity to perceive the learner’s narrative. This process of mutuality, embodied in the experience of epistemic trust, enables us to inherit and transmit the treasure trove of accumulated knowledge that is human culture.

But how does mentalizing generate epistemic trust if the development of mentalizing itself depends on the opportunity to learn about oneself from the other?

OSTENSIVE CUES AND THE RECOGNITION OF THE OTHER AS A HUMAN AGENT

To understand this process better we must turn to development once more. The caregiver’s use of ostensive cues communicates to the infant that the caregiver recognizes them as an agent—that she recognizes that they have a mind that is capable of motivating action. This is the biological cue alerting the infant that the communicator’s message needs to be understood and taken into account. The caregiver’s communication to the infant about the infant forms the basis of the infant’s first self-representations. This is the process of marked mirroring, a process described originally more than twenty five years ago by Gergely and Watson (1996), that enables the interpersonal acquisition of self-representation, which becomes increasingly complex as the child’s self-other representations develop. An example would be the caregiver’s expression of congruent emotion in response to the infant’s distress, which serves as a template around which
the infant’s understanding of emotion begins to be organized. Ostension is the biological cue alerting the infant that the communicator’s message needs to be understood and taken into account.

In relationships between adults, self-other representations underpin our mentalizing of the other. As described above, the more nuanced our representation of ourselves, the more likely we are to be able to accurately identify the mental state of the other as they hold us in mind in their feelings and thoughts. This in turn creates trust and the opportunity for us to learn about ourselves by finding something that corresponds to our experience in their reaction to us, thus gradually elaborating our personal narrative. However, we vary in our capacity to imagine the minds of others. If we were not accurately imagined ourselves, we are likely to struggle, as individuals with personality disorder do (American Psychiatric Association 2013, section III), with putting ourselves into others’ shoes. If we are unable to interpret the other’s actions in terms of their mental states, we will not be able to find ourselves in their thoughts, and we lose the opportunity to learn about our own minds through this kind of dialogic exchange. We are then unable to modify our self-representation in line with the other’s perception of it. To put it more simply, our personal narrative remains too concrete, rigid, and insufficiently imbued with nuanced mental states necessary for true empathy. We believe that this rigidity lies at the core of severe personality pathology (Bateman, Campbell, Luyten & Fonagy 2018; Dagan, Groh, Madigan & Bernard 2021; Fonagy, Luyten & Bateman 2017; Luyten, Campbell & Fonagy 2021) and explains why it is so difficult for fresh knowledge to penetrate the shield created by imperviousness to new relevant information. This inaccessibility to change has led to the inappropriate designation of personality disorder. Clinicians blame character, with its biological overtones, for the persistence of distress. This may be a misattribution. It is simply the incapacity to benefit from dialogue in modifying personal narrative, which presents an insurmountable barrier to change.

What is essential for the establishment of epistemic trust, for the efficient and rapid transfer of information provided by natural pedagogy, is the learner’s ability to recognize that the instructor perceives them as the human mentalizing agent that the learner experiences themselves to be. This gives the instructor a special deferential status and allows them access to modify the learner’s ideas and beliefs, which are normally safeguarded by our
naturally conservative vigilant stance to novelty. *Feeling recognized as an agent serves as an ostensive cue.* The term agent in this context comes from the German sociologist Max Weber (1923), who identified a sense of control or agency as essential for an individual to experience full humanity.\(^6\)

In order to establish epistemic trust, we need to be able to mentalize the other well enough to be able to see ourselves accurately mentalized. *This imaginative experience is the key that unlocks our epistemic barrier.* Throughout development, the key to keeping the learning channel open is the experience of self-recognition that enables genuine learning from the object. That experience is based on detecting how one is seen—what we could term the *epistemic match.* As represented in the figure below, the learner’s imagined sense of self (their personal narrative) (1) is imagined by the informer (2), and this image is perceived (3) and compared with the learner’s personal narrative (4), and in case of a match the channel for rapid and efficient knowledge transfer is opened (5). Putting it at its simplest, if I feel that I am understood, I will be disposed to learn from the person who understood me. This will include learning about myself but also about others and about the entire world I live in.

At any given moment, for most of us, there is a predominant narrative—the most obvious straightforward way of describing ourselves—but we also all have a range of subdominant narratives, understandings of ourselves that are more nuanced or complex and are hidden from the normal shorthand we might use to describe ourselves. The dominant narrative is in the foreground, for example, the need to fulfill obligations to a valued colleague, but behind it may be a range of other narratives, such as anxiety about other tasks that are being neglected in order to meet the particular commitment. A dominant narrative a patient might have is that “I need to be admired, and to achieve that, I need to meet all the expectations you have of me,” but the subdominant narrative might be that “I’m tired of working so hard trying to please people

\(^6\) One of us (PF) has a recollection that illustrates the power of the illusion of feeling recognized: “I vividly remember a primary school teacher in 6th and 7th grade. Everyone in the class agreed that she was our best teacher by far. She did something that no other teacher did. She gave all of us a book at the end of the semester, individually chosen to fit with what she had learned about our personality. While I will not disclose the book that I received, I will tell you that I spent close to two years on the analytic couch trying to figure out why that book was chosen for me. In the end, I realized she gave them out randomly! But it was one of the all-time great ostensive cues!”
We believe the recognition of these subdominant narratives is a particularly potent means of ostensive cueing, leading to the re-establishment of epistemic trust. “I have noticed just how hard you work to make sure you meet all the needs people around you have; you know, in your shoes I would just get exhausted trying to meet every expectation anyone might have of me.”

Let us take a momentary step back into the consulting room. We are suggesting that to enable a patient to genuinely learn from us—that is, to modify their enduring representations of the world—as analysts we have to initially generate a joining of minds, a coming together of perspectives that produces a powerful sense of self-recognition when we open our minds to our patients sufficiently for them to have an experience of our experience of them. The content of that particular moment may not be relevant to our subsequent analytic work but the moment creates the recognition of the possibility of learning in that context.

FIGURE 1
When Daniel Stern et al. (1998) helped us recognize that there was something more than interpretation to psychoanalysis, we believe they were pointing to a similar insight, although they called it *moments of meeting*. They wrote:

A “moment of meeting” is the event that rearranges implicit relational knowing for patient and analyst alike. It is in this sense that the “moment” takes on cardinal importance as the basic unit of subjective change in the domain of “implicit relational knowing.” [p. 906]

The patient and analyst are experiencing each other as persons relatively unseen by their usual therapeutic roles. In analysis, such moments entail an enforced alteration of the frame that forces an intersubjective moment of meeting. Perhaps something unexpected happens, like a large wasp flies into the room, scaring both analyst and patient; something the patient says makes both analyst and patient laugh uncontrollably; or the analyst steps out of role and shows sadness in relation to a loss suffered by the patient. Stern et al. recognize that such moments have immense significance for subsequent analytic work. “When a change occurs in the intersubjective environment, a ‘moment of meeting’ will have precipitated it” (p. 906). In our view, the significance of moments of meeting lies in the accompanying change in the experience of epistemic trust in the patient. The mind of the patient has been opened to rapid, efficient knowledge transfer. The patient is ready to learn from experience (Bion 1962).

As we have suggested elsewhere, this is highly significant for two reasons. First, it enables the patient to use the analytic relationship in a new way. The patient’s identification of the therapist as a deferential figure in combination with the therapist’s mentalizing of the patient increases the patient’s confidence to identify with and internalize what they see of themself in the therapist’s mind. The virtuous cycle is reinforced. The patient’s self-image becomes increasingly nuanced and differentiated as the therapist acknowledges the patient’s emotional state. As the patient uses their more finely tuned sense of their experience of their own mind, they also become better at mentalizing others, in the first instance, their analyst. They are then better able to use the experience of insight and understanding that the analytic relationship provides.
But the benefit does not end in the consulting room. The improved reflective experience of subjectivity brings the patient closer to other minds—not just the analyst’s. The experience of being understood—being mentalized—unlocks our capacity to receive and learn from the social knowledge that has the potential to change our perception of ourselves and our social world. The gift of a mentalizing process in psychotherapy is to open up or restore the patient’s openness to broader social influence, which is a precondition for social learning and healthy development at any age (Allen 2014; Fonagy & Allison 2014). The greatest benefit from a therapeutic relationship comes from generalizing epistemic trust beyond therapy such that the patient can continue to learn, grow, and benefit from other relationships.

THE IMPACT OF TRAUMA AND NEGLECT ON THE ESTABLISHMENT OF EPISTEMIC TRUST

How can the process we have been describing go wrong? Let us return to the multiple experiences that undermined Mr. K’s capacity to generate epistemic trust, leaving him with a feeling of deep isolation even in the company of others and in almost indescribable mental pain.

Mr. K’s history of deprivation and trauma was such that he feared mentalizing others, was initially almost autistic in his stance to social relating, and did not even want to know how they represented his personal narrative of his mental state. His mentalizing capacity was so poor that not only did he consistently misrepresent how others saw him, but he even misinterpreted sympathy as attempts to shame and mock him. He felt persistently misunderstood. He had such an inaccurate view of himself (his personal narrative was dominated by an alien self of indescribable badness—see Fonagy & Target 2000) that even if he was at times able to interpret others’ perception of his mental state accurately, it never felt like a match.

All this left him unable to access new knowledge about the world and his place within it. To others, including his analyst, he appeared rigid and inaccessible to alternative views. He was stuck in a groove, seeing his life as a crushing failure and feeling that he was unworthy of the slightest affection. He was able to ignore all evidence to the contrary. At
times, he appeared to be responsive to therapy; he would discuss his feelings openly and could even show insight about the unconscious and conscious feelings and thoughts that motivated his actions. Yet he remained unchanged, apparently determined to kill himself. When briefly admitted as an inpatient, he convinced everyone that if his 1:1 nursing care was relaxed, he would run from the hospital to the nearby railway track, hoping to throw himself under the express train from London as it sped through the local station. He conveyed that any attempt by the ward staff to rescue him from this fate would be pathetic and misguided.

Obviously, our focus in this paper is only one aspect of a highly complex case. Mr. K had a history of neglect and severe physical abuse, multiple suicide attempts, and chronic self-harm (cutting, head-banging, vicious scratching, and even biting). He had a profound belief that he was a monster who should be destroyed. He manifested all the indications of being host to an alien self colonized by a brutal parental figure who he internalized in a desperate attempt to wrest back a subjective experience of control by identifying with (one of) his aggressor(s) and internalizing his abuser into an already enfeebled self-representation. But this left his torturer inside his experienced sense of himself, a personal narrative focused on determination to annihilate him from within—perhaps the experience of an ego-destructive superego (as Edna O'Shaughnessy [1999] beautifully described it). Other than our minor disagreement about his wish to end his life, he had a kind of treatment alliance with the analytic process, although there was also undeniable frustration over very many months, with little evidence of change.

There is another aspect of learning from others that we have noticed in other patients with histories of neglect and trauma. Mr. A, a patient we have described elsewhere (Fonagy & Target 2005), was a young man with a background of terrible physical abuse. “Fortunately” for him, he also developed childhood leukemia, which was successfully treated. He quickly learned that by feigning poor health he could to a large extent avoid maltreatment. Ultimately, he became quite uncertain about when he was genuinely unwell and when he was feigning poor health. In contrast to Mr. K, this patient manifested an excessive epistemic credulity. He was far too readily influenced by others, which left him vulnerable to terrible interpersonal judgments and exploitation by individuals he described as “his friends,” whose behavior indicated that they were
anything but. In other (less affluent) life circumstances, he might have been given a diagnosis of antisocial personality disorder. His history was a pathetic sequence of finding himself with sometimes violent criminals who used him in a heartless way, but despite repeated experiences of betrayal he was still strongly influenced by charismatic leaders. He was unable to reverse this pattern despite insight into the process. He started major projects with conviction only to abandon them weeks later. He joined opposing political parties and social movements, appearing to identify with their goals despite lacking the conviction and commitment that would normally be required; he found himself in various religious groups and churches, planned on becoming a counselor or a coach, and then applied to lead an international conglomerate. Of course, he was also remarkably arrogant and highly narcissistic, and frequently very dismissive of his analyst’s ministrations. Without going into too much detail about Mr. A, what seems relevant in this context is the lack of stability of his personal narrative. He presented a disjointed sequence of rather grandiose pictures of himself, which alternated with a remarkable complete absence of self-description—almost as if he simply did not know who he was. Naturally, sometimes he saw himself as a profound failure but this narrative did not persist (as it did with Mr. K). From our current point of view we would see Mr. A’s vulnerability to social influence and his readiness to identify with whatever context offered immediate affirmation as indicating the extreme facility with which a match between his labile self-representation and the other’s perceived representation of his self could be made.

Mr. A felt understood and opened his mind and lowered his epistemic defenses to people who did not genuinely understand him because his self-representation was so diffuse that everything felt like a fit with his ever-changing personal narrative. At other times, he distorted his representations of himself to create an illusory fit with the person he was with when none actually existed. In both cases, he was excessively vulnerable to social influence.

In some patients, these two states, epistemic hypervigilance as in Mr. K and epistemic credulity as in Mr. A, can co-exist and can oscillate rapidly. This can be confusing for the clinician, whose understanding appears to be valued and idealized one day only to be scotomized, spurned, denigrated, and spat out the next.
Returning to consider Mr. K’s treatment, we can see that his traumatic early history severely limited his capacity to sustain an experience of either himself or his world of objects as having the properties of coherence and intentionality. Mr. K’s analyst’s early attempts to mentalize Mr. K failed to take this profound developmental deficit sufficiently into account. Either Mr. K would react to the representations he was offered with non-recognition, or his distress would deepen as the impact of his disorganizing negativity was reinforced by the analyst’s interpretations. No matter how accurate they were, he was not able to absorb them and learn from them. In order for this to begin to be possible, the analyst would have to show him that he could appreciate just how difficult it was to walk in his shoes, but this is not possible when ordinary empathy fails to provide the appropriate pointers, when conscious experience is beyond that which can be genuinely reflected, as we described in our previous paper. As he once said, “I walked in the street and I felt the soles of my feet on the ground.” His experience of walking was evidently radically different from that of his analyst; it was denuded of meaningfulness. It is difficult or perhaps even impossible to genuinely empathically reflect this type of experience, and yet Mr. K needed to be able to see that the analyst could see and empathize with this. This is the inherent technical challenge of finding a way through pervasive negativity. Our capacities for reflection are unequal to the task because of the coherence of our phenomenological experience.

The impasse for patient and analyst is caused by repeated failure to establish an epistemic match that would create epistemic trust. Attempts at empathy, which this analyst—like probably many others—frequently make, miss the mark by a considerable margin because they address ordinary conscious experience. When the patient’s consciousness is distorted, as was the case for Mr. K, supposedly empathic statements generate problems, revealing to the patient the analyst’s incapacity to join him within his world. The experience of alienation even in the absence of malevolence is iatrogenic and causes distress and sensitivity, however acute, causes anguish and pain. Paradoxically, only silence appears to feel real.

As we described in the original case report of Mr. K, the epistemic tipping point came with the recognition of patient and analyst’s shared helplessness in the face of the profoundly disruptive impact of the
negativity that Mr. K’s monstrous experiences had mirrored. They shared a sense of helplessness in the face of that which could not be shared yet was part of the patient’s consciousness, but not a part that could be communicated in a way that could be reflected. When the patient perceived the analyst’s recognition of the patient’s helplessness in relation to this experience, an epistemic match (a moment of meeting) could occur. While the analyst endeavored to reflect something that could not be reflected upon, his attempts could not but be experienced as laughably futile. The impasse began to be overcome when patient and analyst admitted that the malevolent internal state was too powerful for either patient or analyst to fight and that all the analyst could do was to try and help him to live with these experiences in a slightly more comfortable way. Mr. K’s apartment was frequently overwhelmed with black plastic bags of rubbish; there were times when he was unable to bring himself to remove the putrefying rubbish from his flat. In a dream, he tried but was defeated by the weight. Rather than focusing on the obvious depiction of Mr. K’s experience of feeling unable to clear his mind of its decomposing contents, the analyst focused on his own experience of wanting to be helpful but finding the accumulated problems they faced overwhelming. Mr. K acknowledged what was said and added “the most reassuring thing about coming to see you is that you do not pretend that things can change for me.” Both patient and analyst then fell silent, but both felt that something true had been said by each of them.

The work of differentiating and containing Mr. K’s negativity by talking in terms of the magma of his internal robot feeling of not being human or most commonly being a monster gave him the space to begin to glimpse in his analyst’s mind the possibility of an experience of coherent intentional selfhood apart from these unmentalizable experiences and very gradually to internalize this experience of alienation. We cannot stress enough that there are patients who lack the capacity to have a self-experience that includes the conscious personal narrative that most of us take for granted. Over time, this painstaking work—which, as we acknowledged in the earlier paper, is not analytic in the traditional sense—established a narrative, which in turn enabled epistemic trust to be created in the consulting room. The process ultimately allowed Mr. K to develop and consolidate sufficient capacity for reflective consciousness to begin to learn both in the analysis and beyond, and he began to
develop professionally. As reported in the previous paper, he managed to complete an extensive and extremely demanding professional training, which allowed him to take a job consistent with his age and ability. His interpersonal relationships benefited to only a limited extent.

**PSYCHOANALYSIS AS RE-EDUCATION**

Gergely and Csibra’s concept of *natural pedagogy* has played a vital role in our development of these ideas. Extending their idea, we suggest that it may be fruitful to think about the psychoanalytic process in terms of teaching. However, this may generate misgivings rather than opening the way for any creative re-imagining of our activity. Since Freud’s (1919) distinction between the *pure gold of analysis* and the *copper of direct suggestion*, the idea that psychoanalysis might involve any form of deferential knowledge transmission has been felt by many to go against the grain, and as a profession we have invested a great deal of effort to identify the proper therapeutic aim of psychoanalysis as concerned with truth and striving to bring about structural change, whereas other psychotherapies are focused on cure and the alleviation of symptoms.

Yet Freud (1917) himself, in his inimitably provocative style, stated in his *Introductory Lectures* that “in our technique we have abandoned hypnosis only to rediscover suggestion in the shape of transference” and on several occasions he described psychoanalysis as “a kind of *after-education* [Nacherziehung]” (p. 451). In *On Psychotherapy* (1905), he wrote that “Psycho-analytic treatment may in general be conceived of as such an *after-education* in overcoming internal resistances” (p. 267). He returned to this point in the posthumously published *Outline of Psycho-Analysis*, suggesting that the transference on to the analyst, which puts him in the position of the patient’s superego, creates an opportunity for “a sort of *after-education* of the neurotic” (Freud 1938, p. 175). However, Freud (1938) makes it very clear that this does not amount to an invitation to the patient to imitate the analyst: “However much the analyst may be tempted to become a teacher, model and ideal for other people and to create men in his own image, he should not forget that that is not his task in the analytic relationship, and indeed that he will be disloyal to his task if he allows himself to be led on by his inclinations.” He stresses that “In all his attempts at improving and educating the patient the analyst...
should respect his individuality” (p. 175). In other words, while the analytic work may furnish the patient with knowledge, the process is only genuine to the extent that the knowledge gained is authentic knowledge about the self.

Freud and analysts since would agree that psychoanalysis offers the patient a unique opportunity to learn from interpersonal experience. Accepting that analysis is a situation in which perhaps both patient and analyst learn in a social context foregrounds the importance of epistemic trust. Perhaps even before considering the interaction between patient and analyst, we should note that the unique features of the analytic setting, such as the patient’s position on the couch with the analyst behind him and the shared commitment to allow gross violations of Grice’s conversational maxims (Grice 1975; Grice 1989), in themselves act as a form of ostensive cueing, marking what transpires in the consulting room as having a special status that warrants reflection. The analytic frame is so tight that it cannot but help to facilitate the development of a shared interpersonal reality. The deliberate limiting of distracting extraneous elements facilitates the analyst’s capacity to build a clearly visible image of the patient’s personal narrative. The emerging transference gives both patient and analyst an opportunity to discern the shape of the patient’s internal world, and their joint attentional focus on the experience they are sharing disposes the patient to be open to learning from it. The process of imagining the patient’s mental experience provides a creative space for the patient to engage in learning about the functioning of both their own and other minds. In this dialectic process, what is imagined by the analyst about the patient’s self provides material with which the patient constructs their self-experience. These elements then gradually enable the patient to use their creative capacity to imagine more accurately the object’s sense of self, his or her wishes, beliefs, thoughts, and desires.

Our interpersonally generated experience of epistemic trust, which depends on our sense of being accurately perceived, provides a safe space within which we can think about both our own and other minds, and take seriously and learn from this experience. One measure of the success of analysis is there having been sufficient understanding to engender this attitude of trust in the importance of my own subjective experience, a confidence in self-perception. Another may be the extent
to which my trust extends beyond the specific therapeutic relationship to other close relationships, enabling me to establish and maintain an adaptive openness to learning from these. The epistemic trust that has been established prior to this process ensures that this work of imagination can serve as a source of learning and endure, permanently modifying the mental structures that underpin the patient’s experience.

The patient’s capacity to mentalize both themselves and the other will vary across time, but it will also vary between individuals. However, even more striking is the variance in the extent to which these images endure and cause lasting change. It is the quality of epistemic trust, rather than the nature or depth of the insight or fresh understanding, that is likely to determine the extent to which new learning has led to structural change (Freud 1923, 1933). Stern et al.’s (1998) hypothesis about moments of meeting illustrates how the sudden generation of epistemic trust can consolidate the new learning and generate lasting change. In other words, we see the concept of structural change as consisting of the individual’s predisposition to be flexible and adaptive in response to changing social circumstances, which in turn relies on the experience of epistemic trust in relation to their social context.

What determines the level of epistemic trust? If the patient struggles with understanding the other person’s mind, this will restrict their capacity accurately to identify the other’s representation of their own subjectivity. The potential space that Winnicott (1971) refers to is the imagined common ground between self-reflection (the experienced self) and the experience of the other’s reflection on the self. The larger this common ground, the closer the shape of self and the other’s perceived view of the self, the higher the level of epistemic trust experienced by the patient, and the more open the patient’s mind is to imagining the analyst’s experience in relation to physical or mental phenomena to which the patient so far has no access. Patients unlock their imagination and accept a new reality when they feel that their reality has been adequately imagined and accepted. By imagining the patient’s mental experience, the analyst creates potential space for that experience in the patient’s mind.

With individuals whose mentalizing is prone to failure because of trauma or intense unregulated emotion, or a combination of the two, the patient will be unable or unwilling to recognize him- or herself in
the analyst’s mind even if the picture that the analyst offers is accurate and sensitively portrayed. In these circumstances, the analyst needs to be careful not to offer more than the patient can digest. The patient arrives in a state of epistemic vigilance or sometimes hypervigilance. They are reluctant to imagine reality as depicted by their analyst. The epistemic tipping point that begets trust in such cases may not be an interpretation at all, but rather an aspect of the analyst’s handling of the clinical situation that makes the patient feel that there is something that can be shared between the two of them; the analyst recognizes their humanity and they recognize the analyst’s. In fact, almost everybody has a story about an episode in their analysis that reminded them that they and their analyst were two human beings sharing an experience (albeit a rather unusual one). Such moments are powerful ostensive cues that tend to be remembered long after the content of specific interpretations, no matter how insightful, have been forgotten. The moment at which vigilance gives way to willingness to assimilate a new reality, not so far directly experienced, will occur when their two minds meet. The Boston group’s brilliant intuition about a moment of meeting that exists beyond interpretation could be called the *epistemic tipping point*. An event—often unrelated to the narrative course of therapy—occurs that is of immense significance to both patient and therapist.

PF recalls from his analysis as a teenager: At the age of seventeen for the first time I drove to my session in my first car. I lay down on the couch and of course started immediately blabbering about the car, my experience of driving to the clinic etc. Mrs. Hurry interrupted the session, got up, went to the window, came back, sat down and said: “Peter, that is a wonderful car!” Much that was said in the analysis I have trouble remembering but the power of personal recognition gave me courage to internalize and organize myself and my life in very many respects in her image of me.

The epistemic tipping point is the moment when the patient becomes aware of the proximity, or contiguity, or similarity of shape between their experience of themselves and the way their analyst sees them. This can dissolve their epistemic vigilance, allowing the analyst’s depiction of reality to become a felt and lived reality for the patient. What is important here is that the analyst’s description need not be *true* for the patient to nevertheless perceive it as matching their personal
narrative. This of course is a great danger, particularly for patients whose sense of their own reality is enfeebled.

But beyond the depiction of reality, a benign process has been set in motion. The epistemic tipping point ensures that the patient becomes concerned with the analyst’s perceptions of him and engages with the analyst in a process of learning how to imagine the subjective reality of others. By improving his capacity to do this, he will come to be able to identify genuine representations of his experience in order to continue to improve his self-understanding.

Genuine cure is never restricted to the analytic setting. Perhaps the most important achievement of the analytic process is to improve a patient’s capacity to accurately perceive themselves and others. These improvements are not necessarily significant in terms of the specific insights they represent. Taking the epistemic trust perspective combined with our suggestion about the value of mentalizing, we can understand why notwithstanding a great number of rival models of mind, psychoanalysts across Europe and the world achieve similar results.

The engagement of the analyst with the patient and the creation of epistemic trust gradually enhances the patient’s capacity to overcome what has blocked change and created rigidity in their mental world. A patient who has been unable either to see clearly enough how others thought or felt about him because of his inadequacy in mentalizing others or to have a coherent image of himself, gradually comes to be able to see how others see him and can trust others enough to open his mind to the social world around him. The analyst’s interpretation initiates the virtuous cycle of self-discovery in the other, a dialectic spiral of ever-increasing self-understanding, but simultaneously increases the potential for epistemic trust, thus opening the patient up to the possibility of learning from interpersonal experiences beyond the consulting room.

One more word about the systemic failure of epistemic trust. It is easy for us to feel that the world is a benign place and that given

---

7 There are far too many models that generate particular insights for us to take seriously any account that would prioritize a specific psychological understanding. The content of psychoanalysis, what we think and write about, has shifted with time and space; specific local analytic cultures appear equally potent in bringing about psychic change.
sufficient insight we can discharge our patients with a clear conscience into a social world that they will now be better adapted to navigate. But this is not reality. There is much in the social networks of our patients that is not benign, that is unjust in terms of deliberate and consistent social misunderstanding. Exposing patients to these experiences will compromise the progress that they have made. Being systematically misunderstood is less painful in a state of epistemic hypervigilance than when therapeutic progress has created a willingness to take on board knowledge from others about oneself and about the world. Hostile environments filled with epistemic injustice become more destructive for our patients following therapy, and this is a concern that our therapeutic optimism causes us to overlook.

None of this is new. More than sixty years ago Winnicott (1957) wrote: “Psychotherapy is not making clever and apt interpretations; by and large it is a long-term giving the patient back what the patient brings. It is a complex derivative of the face that reflects what is there to be seen” (p. 117). Developmentally, in the earliest stages this takes place via the caregiver’s holding and handling of the baby, which (if things go well) communicate her awareness of him as a new, separate person. Winnicott felt that while this kind of early experience has a fundamental importance, providing a ground of being that we usually take for granted, it is “apt to be forgotten by those who only look on.” Directly addressing mothers in a 1943 radio broadcast, he said: “I think the most important thing is that you easily feel that your baby is worth getting to know as a person, and worth getting to know from the earliest possible moment” (Winnicott 1945, p. 222). What takes place in the moments that enable a patient’s perhaps lifelong epistemic hypervigilance to begin to be relaxed is not an epistemic match at the level of cognitive reflection, but an experience of being treated in a way that mutually acknowledges personal narrative, the common ground between patient and therapist despite the asymmetry of their relationship. This may sow the seed of trust that enables the capacities for learning from personal experience to get underway across all life’s contexts. After all, this is what natural selection has defined for all of us as our primary task, to embrace and accommodate to our cultural context so that we can transmit these learnings to the generations to come.
No potential conflict of interest was reported by the author(s).

**DISCLAIMER**

Potentially personally identifying information presented in this article that relates directly or indirectly to an individual, or individuals, has been changed to disguise and safeguard the confidentiality, privacy and data protection rights of those concerned, in accordance with the journal’s anonymization policy https://www.tandfonline.com/journals/upaq20.

**REFERENCES**


——— (1923). The Ego and the Id. S. E., 19.


TRAUBLE, B. & BATZ, J. (2014). Shared function knowledge: infants’ attention to
function information in communicative contexts. J. Experimental Child
differences in sensitivity to internal and external cues at 12 months.
Bulletin Menninger Clinic, 75:64–93.
WILLIAMS, R. (1983). Keywords: A Vocabulary of Culture and Society. London:
Cambridge Univ. Press.
Tavistock, pp. 111–118.
London: Tavistock, pp. 95–103.

Peter Fonagy
Division of Psychology and Language Sciences,
University College London,
Gower Street,
London
WC1E 6BT

p.fonagy@ucl.ac.uk