Mentalizing, Epistemic Trust, and The Active Ingredients of Psychotherapy

Chloe Campbell[a,c], Güler Beril Kumpasoğlu[a,b,c], and Peter Fonagy[a,c]

[a] Research Department of Clinical, Educational and Health Psychology, University College London, UK

[b] Department of Psychology, Ankara University, Ankara,

[c] Anna Freud Centre, London

Corresponding author: Dr Chloe Campbell, c.campbell@ucl.ac.uk

Abstract

This article explores the implications of epistemic trust within the mentalizing model of psychopathology and psychotherapy, emphasizing the role of the restoration of epistemic trust in therapeutic settings. At the core of this exploration is the developmental theory of mentalizing, which posits that an individual's ability to understand mental states—both their own and others'—is cultivated through early caregiver interactions. The paper expands on this concept by reviewing and integrating evolutionary theories suggesting that humans have evolved a unique sensitivity to teaching and learning through ostensive cues, enhancing our capacity for cultural transmission and cooperation. However, adversities such as trauma or neglect can disrupt this developmental trajectory, leading to epistemic disruption, where individuals struggle to engage with or learn from social experiences effectively. This disruption can manifest in psychological disorders, where mentalizing failures associated with difficulties in social functioning and maintaining relationships. The article proposes that psychotherapeutic approaches can effectively address these disruptions and outlines three key communication aspect that unfold in psychotherapeutic interventions. We discuss that

effectiveness of these interventions may hinge on the reestablishment of epistemic trust, enabling patients to re-engage with their social environments constructively and adaptively.

Introduction

A phrase we have used several times to encapsulate recent developments in thinking in the application of epistemic trust to the mentalizing model of psychopathology and psychological treatment is the following: "if I feel that I am understood, I will be disposed to learn from the person who understood me, who I feel is a trustworthy potential collaborator. This will include learning about myself but also learning about others and about the world I live in" (Bo et al., 2022', p. 7). The purpose of this article will be to unpick this potentially platitudinous-sounding statement to get to the heart (or, more accurately, the shared minds) of psychotherapy. The theory of epistemic trust is informed by developmental research, particularly recent empirical work using self-report measures, but to a powerful extent it is shaped by and motivated to inform clinical practice. We aim to describe this dimension of the work here.

A Developmental Theory

Mentalizing is the capacity to understand one's own and others behavior with reference to mental functions such as beliefs, knowledge, feelings, desires and intentions. It is both an everyday and a highly complex social cognitive process. Mentalizing theory originally emerged in the context of developing clinical treatments for individuals with borderline personality disorder, and mentalization-based treatment now has an extensive evidence base (Bateman et al., 2023). The theory is a developmental one: it suggests that in the context of early relationships, the infant's mind—in terms of its mentalizing capacity—is incubated by the minds around it. Well before an infant has learned to recognize or understand their own thought processes, or appreciate that other people have their own

entirely distinct minds, a caregiver tends to interact with their baby as if they already have a separate, independent mind. This experience of interacting in a way that attributes valid and separate mental states to the infant supports their emerging sense of self as an agentive being whose perceptions of the world are valid. In addressing the physical, social, and emotional needs of a child, caregivers—parents, grandparents, siblings, or others—engage in a dynamic process that transcends simple provision of care. They interpret and reflect the child's internal psychological world, effectively bridging the gap between the child's experiences and the caregiver's understanding. An example of this is the universal way in which human caregivers mimic their infants, responding to the infant with exaggerated facial displays of imitation and use of voice in a way that reflects how the infant may be feeling back to the infant, but in a slightly pretend or "play-acting" manner (Zettersten et al., 2024). One aspect of these exchanges has been described as *marked mirroring*. The use of mirroring that is marked in the sense of offering unmistakable cues (e.g., exaggerated facial movements or tone of voice) indicates to the infant that the caregiver is not displaying their own dispositional state, but that of the infant (Gergely, 2018).

Parental mentalizing has been found to influence the development of the child's capacity for mentalizing (Luyten et al., 2020). When the development of mentalizing is poorly supported, it is not that the individual will necessarily stop thinking about mental states, but such thinking will be inadequate for the task; it may be too concrete and real, so the thought will not be in an intermediary space between totally abstract and physically concrete, but will tend to be excessively located at either end—a thought of doing something will feel the same as actually doing it, or the thought feels so unreal it has no implications for reality. Furthermore, if thoughts become meaningless, physical action can be experienced as the only currency of importance.

An Evolutionary Theory

Recent developments in this thinking have expanded the implications of the recognition of agency that is experienced in a mentalizing relationship. This expansion has its origins in the work of Gergely Csibra and György Gergely, who developed a revolutionary theory about how humans evolved to teach and learn new and relevant cultural information from one another (Csibra & Gergely, 2009, 2011). They suggest that the human capacity for knowledge transfer is based on a particular sensitivity to signals that learners look out for when trying to decide whether to trust the person communicating the information (the "teacher"), and that teachers transmit when trying to encourage a learner to listen to them. We are using the terms "teacher" and "learner" broadly here; in evolutionary terms, the "teacher" often is a more experienced individual, typically a genetically related elder, imparting valuable information to a novice (the "learner"). However, this dynamic applies broadly to any situation where knowledge transfer occurs, such as between a schoolteacher and pupil, psychotherapist and client, a politician appealing to voters or, of course, between a parent/caregiver and a child. The teacher must indicate to the learner that what they are conveying is relevant and should be stored in the learner's mind as useful and valid cultural knowledge. They do this by giving what are termed *ostensive cues*—particular behavioral signals. Human infants are attuned to respond with particular attention to ostensive cues, which include eye contact, contingent or sensitive attuned responding, turn-taking, and the use of a special vocal tone, as well as language as the infant develops, all of which appear to trigger a special mode of listening generating a mind-state of learning in the infant (Egyed et al., 2013).

Put simply, ostensive cues trigger *trust in communication*: they signal that what the caregiver is trying to convey is of interest and significance to the child. This trust in the communication of social knowledge is termed *epistemic trust*. Epistemic trust is defined as

trust in the social value, personal relevance, and generalizability of information that a helping individual may have to offer. This particular form of trust, or openness to social learning, is one of the mechanisms that powers the human inclination to teach and learn new and relevant cultural information from one another, which has been termed *natural pedagogy* (Csibra & Gergely, 2011), which is thought to be a key adaptation in enabling social complexity and cooperation (Gergely & Csibra, 2005). Triggering epistemic trust in order to support natural pedagogy is necessary because indiscriminate openness to all forms of cultural communication would be risky. In contrast to the widespread assumption that young children are prone to uncritical credulity, recent evidence shows that they extend trust with appropriate selectivity (Harris & Corriveau, 2011). Even preschoolers monitor the reliability of a particular individual who is providing them with information, differentiate between informants on the basis of the validity of their past claims, and are guided by their interpretation of informants' minds when evaluating new information from these people (Koenig & Harris, 2005). It is possible that a communicator, whether motivated by a deliberate attempt to mislead and cause harm, or simply as a result of being misinformed themselves, might misinform the listener. Some form of epistemic vigilance is required as protection against such situations (Sperber et al., 2010).

Epistemic Trust and Epistemic Disruption

As mentioned at the outset, individuals whose social experiences have left them feeling understood are more likely to be open to social communication and present high levels of epistemic trust (Campbell et al., 2021; Fonagy et al., 2017a, 2017b). By contrast, those with a history of childhood adversity may be more likely to show *epistemic disruption*, which can impede learning from social experience (Campbell et al., 2021; Fonagy et al., 2015). We have described two expressions of epistemic disruption. The first manifests as high levels of

epistemic mistrust, characterized by excessive vigilance – a tendency to reject or avoid any communication; this may impede learning from social experience (Fonagy et al., 2015) Epistemic disruption may also involve excessive *epistemic credulity*, in which insufficient discrimination is used when receiving information, leaving the recipient vulnerable to misinformation and/or exploitation (Campbell et al., 2021).

The theory of epistemic trust was first elaborated clinically in relation to the effects of epistemic mistrust, the notion being that the closing down of epistemic trust may act as a general factor that creates a broad vulnerability for psychopathology, particularly in its more complex and persistent forms, for example, borderline personality disorder (Fonagy et al., 2015). Epistemic trust may be undermined through the early-life experience of surrounding caregiving minds not being reliable, benign, and able to recognize what is meaningful and relevant to the infant. In the presence of trauma (which essentially disavows both subjectivity and agency, i.e., the victim is not being mentalized), epistemic mistrust tends to crystallize, rendering the individual unable to learn from their social environment, and this perpetuates maladaptive relationship patterns. The presentation of fresh information is of little value to the individual because although it may be understood at a superficial level, it cannot be internalized as being personally relevant or meaningful, or encoded as being generalizable social currency.

Difficulties in mentalizing, which are often a developmental outcome of the child's own mental states being misrepresented, obstruct the formation of accurate representations of others' minds that makes effective selective social learning possible. In a cyclical fashion, consistent errors in the application of selective social learning disrupt the interpersonal mentalizing processes that can act to correct distortions in the understanding of cultural content that is being conveyed. The outcome of these difficulties is that the individual becomes disadvantaged in trying to navigate their social environment. The individual

becomes unable to benefit from cultural learning that makes possible cooperation, social functioning, and the ability to tolerate one's own and others' minds. We suggest that this is one of the underpinning states of any mental disorder, and a continuing disadvantage for individuals with more complex and longstanding difficulties. The loss of epistemic trust associated with missing out on the straightforward transmission of cultural learning means that these individuals are less able to access the benefits, consolations, and sustenance of shared human activity that makes resilience possible. It is through the stimulation of epistemic trust that the considerable cultural and social landscape that any human infant must learn about in order to navigate it successfully becomes manageably knowable. Cultural learning is thus shaped by exposure to surrounding minds, which generates a kind of Bayesian, social interaction-driven form of epistemic trust that allows the efficient passing on of cultural content.

Mentalizing is a powerful cue for opening a channel for epistemic trust, that is, the attitude needed to benefit fully from opportunities to learn from others—to learn about the world in general and the social world, including oneself, in particular (Fonagy et al., 2015; Fonagy et al., 2016; Fonagy et al., 2017a, 2017b). Across development, the key to keeping the learning channel open is the experience of self-recognition that enables genuine learning from the "teacher", whoever serves that function at that point in our lives. Self-recognition occurs through identifying how one is seen by the other and appreciating that the other person's perception of oneself is congruent with one's own. This congruence creates what we have called an *epistemic match*, and the channel for efficient social communication is opened, making social learning, affect regulation, and flexible, adaptive social functioning more possible. Feeling a degree of certainty that our priorities and personal narratives have been accurately perceived engenders confidence that it is safe to learn from the person who has given this impression.

A child in a nonmentalizing social system will likely benefit from adopting a more automatic, less reflective mode of mentalizing, which will impinge on the possibility of recognizing an epistemic match. What developmentally emerges as mental disorder may begin as an adaptation in cultural learning style; epistemic trust may not be advisable in all social environments, and hence the closing down of epistemic trust may itself be a form of cultural learning. Only when there is a consistent misalignment of functioning with the environment will psychic distress or behavioral difficulties. Emerging empirical evidence in this area supports these ideas: individuals who score highly on a self-report measure of epistemic mistrust and excessive credulity also tend to suffer from increased psychological distress, are more likely to have experienced childhood abuse and/or neglect, and are more likely to have an insecure attachment style (Campbell et al., 2021; Kampling et al., 2022; Liotti et al., 2023).

Thinking Together and the Creation of Culture

Tomasello et al. (2005) described the process of knowledge accumulation as "the ratchet effect": adopting new elements, which are superior to and enable us to dispense with prior information. This process allows the accumulation of knowledge within a community to be transmitted through interpersonal learning. An idea is retained within a social group until a better solution to the same problem comes along (of course, what is "better" will be understood according to one's prior cultural learning). This capacity for cultural adaptation and refinement has developed through natural selection. Modern evolutionary theory recognizes that survival to reproduction is not the ultimate goal of evolution. Social inheritance makes sense in the context of recognizing natural selection having as its focus optimal adaptation to the environment to ensure the survival of our DNA (Hamilton, 1964). This explains why we prioritize not just our own survival but give such primacy to social

learning and adaptive fitness in our kin, as it is through them that our DNA actually survives us. Social learning in this way is dependent on levels of "thinking together"—joint attention, intellectual collaboration, empathic entanglements in relation to understandings and misunderstandings—that involves interpersonal engagement and functioning at quite a sophisticated level. In our terms, we would describe social learning as being substantially supported by robust, and at times highly reflective, mentalizing.

Relational mentalizing refers to the shared thinking and feeling within a family or other group and it involves intentional states that are assumed by individuals in the system to be joint or shared by everyone. Raimo Tuomela (2005, 2007) has named this category of states *jointly seeing to it*. It has been argued that this form of relational mentalizing constitutes a special "we-mode" (Gallotti & Frith, 2013) in which, it is assumed, the social context (i.e., the mere presence of others) improves a person's potential for mentalizing when others, whether the family or just any other collection of individuals, decide to *be and act together*. In this condition, there is a sense in which no member of the group can be assumed to be doing it "on their own" or can be appropriately considered as thinking or feeling in isolation from others in the "psychological collective". This sense of "we-ness" of shared minds has an irreducibility that means that it must be addressed separately from the individual mentalizing of self and others, as joint actions are experienced in a qualitatively different way and involve shared intentions. Relational mentalizing relies on underlying mutually accepted yet often implicit conceptual and situational presuppositions, and it does not necessarily involve making an agreement to generate joint intention (Tuomela, 2005).

We have recently applied the thinking of Tomasello and Tuomela on the we-mode, joint attention and the transmission of culture to developmental psychopathology. Specifically, what happens when an individual's capacity for social communication and joint attention is disrupted? We have suggested that the interpersonal mentalizing context that

provides the framework for social functioning, which has enabled human culture to sustain itself and accumulate in unique ways, is also the basis for the equally species-specific capacity to develop psychological disorder (Fonagy et al., 2022).

Social Exclusion, Inequality and Psychic Distress

Our interest in the we-mode highlights how, at least in western culture, we tend to underestimate the importance of dependence on our community relative to the significance of our individual actions. For example, when people are asked to rank the importance for their health of various influences, such as exercise, diet, smoking, social support, medication, and social connectedness/integration, they place the influences that are under individual control at the top of the list and rank social integration and social support last (Haslam et al., 2018). Yet, the odds ratios for reduced mortality are greater than 60% for both social support and integration, whereas those for exercise and diet are about 20% (about the same as influenza vaccination). One study revealed that the 10-year mortality of single (separated or bereaved) fathers is 100% greater than that of partnered fathers (Chiu et al., 2018). Social relationships are key for individual survival: the human mind is essentially social and interpersonal. Both psychoanalytic therapy and other mentalizing-focused treatments for mental disorder make sense only in terms of serving to reintegrate the patient into the large, complex, ever-moving stream of human social communication. We propose that psychological disorder— particularly in its more severe and debilitating forms, such as personality disorder—is the outcome of being unable to navigate effectively a role in the shared human project of culture, quite likely as an adaptation to their environment in interaction with genetic risks.

We speculate that the epistemic trust created by the experience of being mentalized might act to regulate the imaginative activity of mentalizing. When we experience being appropriately mentalized by others, we do not need to make endless and fruitless effort to

understand ourselves, and this experience is internalized to provide a secure platform that enables us to turn outward to others and to the world. If community is central to maintaining mentalizing, then it makes sense that disruptions of the links to one's community would undermine mentalizing and generate vulnerability to mental disorder. There is strong evidence that many of the international differences in the prevalence mental disorder are potentially interpretable in terms of the social cohesion of these different societies (Pickett & Wilkinson, 2015). The social environment—in particular economic deprivation or exposure to discrimination, social isolation, or a hostile community—has such a powerful effect on child development because it can signal to the child that they are not in a reliable, safe environment in which it is adaptive to be trusting and open to social learning.

Further, in interpersonal trust experiments, individuals who are less socioeconomically advantaged tend to behave in more community- and socially oriented ways than more affluent individuals (Dubois et al., 2015). Less affluent individuals are more engaged with and dependent on their community; wealthier and more socially protected individuals have a stronger perception of their self-agency and do not need to be as community focused (Keltner et al., 2003; Lee & Tiedens, 2001; Overbeck & Droutman, 2013). As a result, individuals who are in a lower socioeconomic status (SES) environment may be more sensitive to their social environment and attuned to assessing how reliable and benign/supportive it may be than those in higher-SES environments (Gruenfeld et al., 2008; van Kleef et al., 2008). The absence of personal power probably drives greater openness to the need for social networking and the prioritization of the surrounding community (Rucker & Galinsky, 2017). When the social environment is hostile or unsupportive, the individual may be more sensitive to the meaning and significance of the hostility and lack of support, increasing the risk of a breakdown in social learning. It is this effect that may contribute to the relationship between socioeconomic forces and relatively poor mental health outcomes in

lower-SES groups. Likewise, failure of epistemic trust may affect low-SES people more than high-SES people because low-SES individuals are more mutually dependent on their community. If resilience is related to we-ness, joint intentionality, openness to social learning and the epistemic trust embedded in links with the community, then an individual who is more dependent on those links because of their relative lack of power will be more likely to develop mental disorder than an individual whose position of power enables them to act as an independent agent.

We propose that experiencing a meaningful connection with the broader social community requires individuals to feel acknowledged as contributing agents and to perceive that their thoughts and feelings are understood and valued by the social system. This involves having rich experiences of joint attention with others in the course of everyday interactions. It may not necessarily involve explicitly reflective mentalizing, it can be provided through the engagement working together that acknowledges each partner's role and agency in the task at hand (Fancourt & Steptoe, 2018). A characteristic of nonmentalizing social systems, particularly those around children and young people, is that they tend to make high demands from individuals while simultaneously not providing the sense of agentive recognition that comes with meaningful support and thinking together. Mentalizing theory has always maintained that when someone's immediate social context does not provide this kind of experience, they are vulnerable to the development of psychopathology. Could it be that we can identify the mechanism underlying the observable relationship between psychopathology and socioeconomic alienation and inequality in disadvantaged individuals' experience of a broader social context that fails to take into account the subjective richness of the individual minds of those lacking economic or social capital, as is the case in highly unequal social systems? Experience with a social system, such as a school, that fails to understand and appreciate an individual's personal narrative may risk eliciting predominantly unmentalized

reactions to misunderstood social situations, resulting in the individual responding in decontextualized ways in which their intent is dramatically distorted. Tragically, this becomes a characteristic of the entire social system. Social alienation might be understood as a systemic breakdown in epistemic trust directed toward the wider environment, capturing all players within the social system.

What Does Psychotherapy Do with Epistemic Trust?

The proposition that significant deficits in epistemic trust are at the heart of enduring mental health issues implies that cultivating or reinstating epistemic trust could be essential for the efficacy of psychotherapeutic interventions. The idea that epistemic trust may be a critical common factor across different therapeutic modalities suggests that a patient's inherent capacity for epistemic trust could significantly affect the therapeutic process (Fisher et al., 2023). The psychotherapeutic journey essentially entails the transfer of beneficial knowledge from therapist to patient, requiring the patient's trust to reinterpret emotional experiences and social perceptions (Knapen et al., 2020). Thus, epistemic trust becomes a key factor influencing how patients internalize and apply therapeutic insights, with those exhibiting higher levels of epistemic trust being more likely to engage actively and experience significant transformations during therapy (Fisher et al., 2021; Springham & Huet, 2018). Conversely, patients with lower epistemic trust might find it challenging to form a therapeutic alliance and benefit from treatment unless this trust is (re)built during the therapeutic process (Fonagy & Allison, 2014). Therefore, the initial capacity for epistemic trust, reflecting an individual's openness to interpersonal assistance and information communicated by an other, could serve as an indicator to identify those who might show poorer involvement and responsiveness in therapy and who are thus at risk of treatment nonresponse (Knapen et al., 2020).

We will return to our first idea: that of *when an individual ly feels understood*. What does that mean in a clinical context, and what complications surround such a seemingly simple concept? One way into this process is to refer back to the three communication systems of change in psychotherapy that we have described in previous publications (Fonagy et al., 2022), in particular communication system 1, *the teaching and learning of content*. The process at work here is one that is common to all therapeutic modalities, although the nature of the content concerned may vary quite dramatically; it begins with the task of the therapist of conveying to the patient their framework for understanding the mind. For a cognitive-behavioral therapist working with a patient with depression, for example, the work of communication is very clearly presented and signposted as such. For a classically trained psychoanalyst, the model being communicated might be understood according to the classic advice given to novelists: "show, don't tell." The process is one by which, through their interpretations, and through the prism of reference to possible transference processes, the analyst's emerging model of the patient's mind and the way in which past experiences shape current relationships, both with the self and others is better understood. Of course, this process of communicating a therapeutic model is not as simple as it sounds—and the patient may not feel themselves genuinely understood in the therapist's communication. There are various ways in which this can be imagined, all of which help us think about the process of psychotherapy and its active ingredients. The first of these is when a patient's difficulties in mentalizing leaving them deeply vulnerable in social situations (Ensink et al., 2015; Levy et al., 2019; Macintosh, 2013; Taubner & Curth, 2013). Even in the absence of a pervasive failure of imagination, a compromised capacity to mentalize may lead the individual to be biased in their perception of social reality (Cicchetti & Curtis, 2005; Germine et al., 2015; Kay & Green, 2016) and misrepresent how others represent them, leading them to feel persistently misunderstood and to experience an intense and consistent sense of injustice.

Secondly, the long-term outcome of epistemic isolation secondary to the failure of imagination we have described here may create problems for individuals who have distorted personal narratives that generate inaccurate views of the self, so that even an accurate perception of one's personal narrative by others is not experienced as a match, and a painful experience of interpersonal alienation persists. For example, an individual whose personal narrative entails a sense of failure or badness may not initially benefit from a supportive, positive reflection of their self-narrative, and will be left with a sense of not having been understood. Conversely, deprivation and trauma may generate unwarranted trust. We understand such excessive epistemic credulity as being triggered by a hyperactive or unmoored social imagination, generating a personal narrative that is too diffuse to provide an accurate sense of differential awareness of others' capacity to perceive oneself. Excessive credulity results in all personal narratives feeling as if they "fit" sufficiently for trust to be generated, making the person vulnerable to exploitation by others. Of course, limited imagination may cause profound misperceptions of the other's representations of one's personal narrative, and an illusory fit is created where none exists in reality.

This takes us to communication system 2, *the re-emergence of mentalizing*. An adaptive epistemic stance, in which vigilance and trust can be appropriately mobilized in response to judgments about the quality of information or the reliability of the informant, may underpin healthy functioning, which requires rapid, efficient checking and updating of social knowledge (Pennycook & Rand, 2021; Sperber et al., 2010). The reactivation of mentalizing by increasing epistemic trust opens up the patient to learning, leading to a virtuous cycle characterized by social salutogenesis. Again, fostering mentalizing may be an important route to opening up the patient to (social) learning. Mentalizing is the catalyst that activates the effective ingredients of therapy; mental flexibility is generated by the potential

to envision alternatives and to selectively adopt these (with the diminution of epistemic hypervigilance) as part of a modified working model of self–other relationships.

Finally, in communication system 3 (*applying social learning to the wider environment*), the patient responds to the experience of reduced social isolation, and stimulates their capacity to learn from others. This frees the patient to grow in the context of relationships outside therapy. This view implies that it is not just the content and techniques of the therapy, nor the insight acquired in the course of treatment, that are key to its success; perhaps primarily, it is the patient's capacity for social learning and thinking about mental states that improves their functioning as they become able to "use" their social environment in a different way. A further implication is, of course, that there may sometimes be a need to intervene at the level of the patient's social environment. Fisher et al. (2023; 2022) have sought to operationalize how this process unfolds in psychotherapy, describing three elements of psychotherapy—sharing, we-mode, and learning—as essential observable elements of an effective therapeutic process. As they argue: "When individuals put their overall experience into perspective and calibrate their mind to those of others (sharing) while establishing a mutual discourse for the processing of ideas (we-mode), new information pertinent to social adaptation (learning) can be acquired" (Fisher et al., 2022, p. 32). The process by which these experiences play out are, inevitably, iterative, reiterative, and non-linear; indeed, it is the unfolding of these relational experiences that constitute the work of therapy. For example, rupture and repair forms part of the work of thinking together (in making sense of the rupture) while holding on to the separateness of the minds of the therapist and patient (Bateman et al., 2021; Fonagy et al., 2023). Two people—or more, in a family or group setting —focusing their joint attention around a shared point of interest (whether that be a rupture or more generally, an issue or mental state or experience being considered at that moment in treatment) generate a micro-culture made up of the patient(s) and therapist in

which learning from one another is possible—a position of epistemic openness in which social learning and social connection become viable.

References

Bateman, A., Campbell, C., & Fonagy, P. (2021). Rupture and Repair in Mentalization-Based Group Psychotherapy. *International Journal of Group Psychotherapy*, *71*(2), 371-392. https://doi.org/10.1080/00207284.2020.1847655

Bateman, A., Fonagy, P., Campbell, C., Luyten, P., & Debbané, M. (2023). Cambridge Guide to Mentalization-Based Treatment (MBT). Cambridge University Press.

Bo, S., Sharp, C., Kongerslev, M. T., Luyten, P., & Fonagy, P. (2022). Improving treatment outcomes for adolescents with borderline personality disorder through a socioecological approach. *Borderline Personal Disord Emot Dysregul*, *9*(1), 16. https://doi.org/10.1186/s40479-022-00187-9

Campbell, C., Tanzer, M., Saunders, R., Booker, T., Allison, E., Li, E., O'Dowda, C., Luyten, P., & Fonagy, P. (2021). Development and validation of a self-report measure of epistemic trust. *PLoS One*, *16*(4), e0250264. https://doi.org/10.1371/journal.pone.0250264

Chiu, M., Rahman, F., Vigod, S., Lau, C., Cairney, J., & Kurdyak, P. (2018). Mortality in single fathers compared with single mothers and partnered parents: A population-based cohort study. *Lancet Public Health*, *3*(3), e115-e123. https://doi.org/10.1016/S2468-2667(18)30003-3

Cicchetti, D., & Curtis, W. J. (2005). An event-related potential study of the processing of affective facial expressions in young children who experienced maltreatment during the first year of life. *Dev Psychopathol*, *17*(3), 641-677. https://doi.org/10.1017/S0954579405050315

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends Cogn Sci*, *13*(4), 148-153.

   https://doi.org/10.1016/j.tics.2009.01.005

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical*

   Transactions of the Royal Society of London. Series B, Biological Sciences,

   366(1567), 1149-1157. https://doi.org/10.1098/rstb.2010.0319

Dubois, D., Rucker, D. D., & Galinsky, A. D. (2015). Social class, power, and selfishness:

   When and why upper and lower class individuals behave unethically. *Journal of*

   Personality and Social Psychology, *108*(3), 436-449.

   https://doi.org/10.1037/pspi0000008

Egyed, K., Király, I., & Gergely, G. (2013). Communicating shared knowledge in infancy.

   Psychological Science, *24*(7), 1348-1353. https://doi.org/10.1177/0956797612471952

Ensink, K., Normandin, L., Target, M., Fonagy, P., Sabourin, S., & Berthelot, N. (2015).

   Mentalization in children and mothers in the context of trauma: An initial study of the

   validity of the Child Reflective Functioning Scale. *British Journal of Developmental*

   Psychology, *33*(2), 203-217. https://doi.org/10.1111/bjdp.12074

Fancourt, D., & Steptoe, A. (2018). Community group membership and multidimensional

   subjective well-being in older age. *J Epidemiol Community Health*, *72*(5), 376-382.

   https://doi.org/10.1136/jech-2017-210260

Fisher, S., Fonagy, P., Wiseman, H., & Zilcha-Mano, S. (2023). I see you as recognizing me;

   therefore, I trust you: Operationalizing epistemic trust in psychotherapy.

   Psychotherapy (Chic). https://doi.org/10.1037/pst0000501

Fisher, S., Zilcha-Mano, S., & Fonagy, P. (2022). Conceptualizing Epistemic Trust in

   Psychotherapy: A Triadic Model. *Psychotherapy Bulletin*, *57*(2), 29-35.

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy*, *51*(3), 372-380. https://doi.org/10.1037/a0036505

Fonagy, P., Campbell, C., Constantinou, M., Higgitt, A., Allison, E., & Luyten, P. (2022). Culture and psychopathology: An attempt at reconsidering the role of social learning. *Dev Psychopathol*, *34*(4), 1205-1220. https://doi.org/10.1017/S0954579421000092

Fonagy, P., Campbell, C., & Luyten, P. (2023). Alliance rupture and repair in mentalization-based therapy. In C. F. Eubanks, L. W. Samstag, & J. C. Muran (Eds.), *Rupture and repair in psychotherapy: A critical process for change* (pp. p. 253–276). American Psychological Association. https://doi.org/10.1037/0000306-011

Fonagy, P., Luyten, P., & Allison, E. (2015). Epistemic petrification and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and its psychosocial treatment. *Journal of Personality Disorders*, *29*(5), 575-609. https://doi.org/10.1521/pedi.2015.29.5.575

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2016). Reconciling psychoanalytic ideas with attachment theory. In P. Shaver & J. Cassidy (Eds.), *Handbook of attachment* (3rd ed., pp. 780-804). Guilford Press.

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2017a). What we have changed our minds about: Part 1. Borderline personality disorder as a limitation of resilience. *Borderline Personality Disorder and Emotion Dysregulation*, *4*, 11. https://doi.org/10.1186/s40479-017-0061-9

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2017b). What we have changed our minds about: Part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication. *Borderline Personality Disorder and Emotion Dysregulation*, *4*, 9. https://doi.org/10.1186/s40479-017-0062-8

Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends Cogn Sci*, *17*(4), 160-165. https://doi.org/10.1016/j.tics.2013.02.002

Gergely, G. (2018). The social construction of the subjective self: The role of affect-mirroring, markedness, and ostensive communication in self-development. In L. C. Mayes, P. Fonagy, & M. Target (Eds.), *Developmental science and psychoanalysis: Integration and innovation* (pp. 45-88). Routledge.

Gergely, G., & Csibra, G. (2005). The social construction of the cultural mind: Imitative learning as a mechanism of human pedagogy. *Interaction Studies*, *6*, 463-481. https://doi.org/10.1075/is.6.3.10ger

Germine, L., Dunn, E. C., McLaughlin, K. A., & Smoller, J. W. (2015). Childhood adversity is associated with adult theory of mind and social affiliation, but not face processing. *PLoS One*, *10*(6), e0129612. https://doi.org/10.1371/journal.pone.0129612

Gruenfeld, D. H., Inesi, M. E., Magee, J. C., & Galinsky, A. D. (2008). Power and the objectification of social targets. *Journal of Personality and Social Psychology*, *95*(1), 111-127. https://doi.org/10.1037/0022-3514.95.1.111

Hamilton, W. D. (1964). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology*, *7*(1), 17-52. https://doi.org/10.1016/0022-5193(64)90039-6

Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *366*(1567), 1179-1187. https://doi.org/10.1098/rstb.2010.0321

Haslam, S. A., McMahon, C., Cruwys, T., Haslam, C., Jetten, J., & Steffens, N. K. (2018). Social cure, what social cure? The propensity to underestimate the importance of social factors for health. *Soc Sci Med*, *198*, 14-21. https://doi.org/10.1016/j.socscimed.2017.12.020

Kampling, H., Kruse, J., Lampe, A., Nolte, T., Hettich, N., Brahler, E., Sachser, C., Fegert, J. M., Gingelmaier, S., Fonagy, P., Krakau, L., Zara, S., & Riedl, D. (2022). Epistemic trust and personality functioning mediate the association between adverse childhood experiences and posttraumatic stress disorder and complex posttraumatic stress disorder in adulthood. *Front Psychiatry*, *13*, 919191. https://doi.org/10.3389/fpsyt.2022.919191

Kay, C. L., & Green, J. M. (2016). Social cognitive deficits and biases in maltreated adolescents in U.K. out-of-home care: Relation to disinhibited attachment disorder and psychopathology. *Dev Psychopathol*, *28*(1), 73-83. https://doi.org/10.1017/S0954579415000292

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, *110*(2), 265-284. https://www.ncbi.nlm.nih.gov/pubmed/12747524

Koenig, M. A., & Harris, P. L. (2005). The role of social cognition in early trust. *Trends Cogn Sci*, *9*(10), 457-459. https://doi.org/10.1016/j.tics.2005.08.006

Lee, F., & Tiedens, L. Z. (2001). Is it lonely at the top?: The independence and interdependence of power holders. *Research in Organizational Behavior*, *23*, 43-91. https://doi.org/10.1016/s0191-3085(01)23003-2

Levy, J., Goldstein, A., & Feldman, R. (2019). The neural development of empathy is sensitive to caregiving and early trauma. *Nat Commun*, *10*(1), 1905. https://doi.org/10.1038/s41467-019-09927-y

Liotti, M., Milesi, A., Spitoni, G. F., Tanzilli, A., Speranza, A. M., Parolin, L., Campbell, C., Fonagy, P., Lingiardi, V., & Giovanardi, G. (2023). Unpacking trust: The Italian validation of the Epistemic Trust, Mistrust, and Credulity Questionnaire (ETMCQ). *PLoS One*, *18*(1), e0280328. https://doi.org/10.1371/journal.pone.0280328

Luyten, P., Campbell, C., Allison, E., & Fonagy, P. (2020). The Mentalizing Approach to Psychopathology: State of the Art and Future Directions. *Annu Rev Clin Psychol*, *16*, 297–325. https://doi.org/10.1146/annurev-clinpsy-071919-015355

Macintosh, H. (2013). Mentalizing and its role as a mediator in the relationship between childhood experiences and adult functioning: Exploring the empirical evidence. *Psihologija*, *46*(2), 193-212. https://doi.org/10.2298/psi1302193m

Overbeck, J. R., & Droutman, V. (2013). One for all: Social power increases self-anchoring of traits, attitudes, and emotions. *Psychological Science*, *24*(8), 1466-1476. https://doi.org/10.1177/0956797612474671

Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. *Trends Cogn Sci*, *25*(5), 388-402. https://doi.org/10.1016/j.tics.2021.02.007

Pickett, K. E., & Wilkinson, R. G. (2015). Income inequality and health: A causal review. *Social Science and Medicine*, *128*, 316-326. https://doi.org/10.1016/j.socscimed.2014.12.031

Rucker, D. D., & Galinsky, A. D. (2017). Social power and social class: Conceptualization, consequences, and current challenges. *Curr Opin Psychol*, *18*, 26-30. https://doi.org/10.1016/j.copsyc.2017.07.028

Sperber, D., Clement, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind and Language*, *25*(4), 359-393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Taubner, S., & Curth, C. (2013). Mentalization mediates the relation between early traumatic experiences and aggressive behavior in adolescence. *Psihologija*, *46*(2), 177-192. https://doi.org/10.2298/Psi1302177t

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675-691. https://doi.org/10.1017/S0140525X05000129

Tuomela, R. (2005). We-intentions revisited. *Philosophical Studies*, *125*(3), 327-369. https://doi.org/10.1007/s11098-005-7781-1

Tuomela, R. (2007). *The philosophy of sociality: The shared point of view*. Oxford University Press.

van Kleef, G. A., Oveis, C., van der Lowe, I., LuoKogan, A., Goetz, J., & Keltner, D. (2008). Power, distress, and compassion: Turning a blind eye to the suffering of others. *Psychological Science*, *19*(12), 1315-1322. https://doi.org/10.1111/j.1467-9280.2008.02241.x

Zettersten, M., Cox, C., Bergmann, C., Tsui, A. S. M., Soderstrom, M., Mayor, J., Lundwall, R. A., Lewis, M., Kosie, J. E., & Kartushina, N. (2024). Evidence for infant-directed speech preference is consistent across large-scale, multi-site replication and meta-analysis. *Open Mind*, *8*, 439-461.