Mentalization and intersubjectivity towards a theoretical integration

Article *in* Psychoanalytic Psychology · January 2015

DOI: 10.1037/a0037129

CITATIONS
6

READS
218

2 authors:

Rikard Liljenfors
Linnaeus University
6 PUBLICATIONS **37** CITATIONS

SEE PROFILE

Lars-Gunnar Lundh
Lund University
155 PUBLICATIONS **4,152** CITATIONS

SEE PROFILE

Psychoanalytic Psychology
2015, Vol. 32, No. 1, 36–60

0736-9735/15/$12.00 http://dx.doi.org/10.1037/a0037129

MENTALIZATION AND INTERSUBJECTIVITY TOWARDS A THEORETICAL INTEGRATION

Rikard Liljenfors, PhD
Lund University and Kristianstad University College

Lars-Gunnar Lundh, PhD
Lund University

The introduction of the concept of mentalization in psychological science by Fonagy and colleagues has opened up new perspectives for the understanding of psychopathology, psychotherapy, and child development. The present study reviews the theory of mentalization, with a focus on its 4 dimensions (cognitive/affective, implicit/explicit, self/other, and external/internal), and some unclear points and unresolved issues are identified. Mentalization theory is then contrasted with the theory of primary intersubjectivity, which is often seen as an incompatible approach to the development of social understanding. It is argued that this theory, at least in 1 of its interpretations, is not only compatible with mentalization theory, but may also possibly contribute to the resolution of some problems in mentalization theory. More specifically, it is argued that mentalization originally develops in the context of primary intersubjectivity, and that primary intersubjectivity is a basic prerequisite for the development of mentalization; but also that there is a considerable overlap between the concepts of primary intersubjectivity and those of implicit and externally focused mentalization.

Keywords: mentalization, primary intersubjectivity, affect regulation, metacognition, social cognition

During recent years, Fonagy and colleagues (e.g., Fonagy, Bateman, & Bateman,

This article was published Online First June 16, 2014.

Rikard Liljenfors, PhD, Department of Psychology, Lund University and School of Health and Society, Kristianstad University College, and Lars-Gunnar Lundh, PhD, Department of Psychology, Lund University.

We thank professor Ingar Brinck and two anonymous reviewers for helpful comments on earlier drafts of the article.

Correspondence concerning this article should be addressed to Rikard Liljenfors, School of Health and Society, Kristianstad University College, Elmetorpsvägen 15, SE-291 88, Kristianstad, Sweden. E-mail: Rikard.Liljenfors@hkr.se

2011; Fonagy, Bateman, & Luyten, 2012; Fonagy, Gergely, Jurist, & Target, 2002; Fonagy, Gergely, & Target, 2007) have generated a theoretical framework around the concept of mentalization, with far-reaching implications both for developmental psychology—in terms of a theory for the development of mentalizing ability and social understanding—and for psychotherapy and psychoanalysis more generally (e.g., Allen & Fonagy, 2006; Bateman & Fonagy, 2004; Fonagy & Bateman, 2012). *Mentalizing* is defined as an "imaginative mental activity, namely, perceiving and interpreting human behavior in terms of intentional mental states (e.g., needs, desires, feelings, beliefs, goals, and reasons)" (Fonagy et al., 2007, p. 288). This includes both the interpretation of *others'* behavior in terms of mental states, and the understanding of *one's own* mental states, as well as the ability to differentiate one's own and others' mental states, and to differentiate mental states from external reality. The full capacity for mentalization is described as a developmental achievement that depends on the early interactions between infants and caregivers, and is assumed to depend on the development of "a symbolic representational system for mental states" (Fonagy et al., 2007, p. 289).

This theoretical framework, which we will refer to as the *theory of mentalization*, is probably not yet to be seen as a full-fledged theory. Rather it is as a framework which is still developing, and consists of a number of conceptual innovations, distinctions, hypotheses, and suggestions. Therefore, the theory of mentalization, at the present stage, should not be expected to be complete and free from inconsistencies. The purpose of the present study is to explore some unclear points in the theory and to contrast it with an alternative perspective, which is sometimes seen as incompatible with Fonagy et al.'s approach: the theory of primary intersubjectivity as developed by Trevarthen (1979) and others. It is argued that this theory, at least in one of its interpretations, is not only compatible with mentalization theory, but may also be used to explain unresolved issues in mentalization theory.

The article has two parts. In the first part, the main aspects of the theory of mentalization are described and analyzed, with a focus on the four dimensions of mentalization, and their role in the description and explanation of the development of mentalization in the child. This leads to an identification of some unresolved problems in the understanding of how mentalization develops. In the second part of the article, the theory of mentalization is compared and contrasted with the theory of primary intersubjectivity, and it is argued that an integration of these theoretical frameworks may serve to solve some of the problems that were identified in the first part of the article.

The Development of Mentalization and Its Dimensions

The theory of mentalization in part leans on research carried out since the 1980s on children's theory of mind, which demonstrated that children generally become able to reason in terms of "false beliefs" when they are around 4-years-old (e.g., Wellman, 1990). This accomplishment indicates that by this age children are able to explain other people's behavior in terms of their having mental representations—beliefs—that may or may not correspond with physical reality. As pointed out by Fonagy, Gergely, and Target (2007), however, the classical construct of theory of mind with its associated false beliefs paradigm is too narrow, "as it fails to encapsulate the relational and affect regulative aspects of interpreting behavior in mental state terms"

(p. 288). Hence, the theory of mentalization has more wide-ranging ambitions than the classical theory of mind paradigm.

Attachment theory and developmental research on attachment is a second main influence on the theory of mentalization. Main (1991) held that much of children's early experiences are important for their subsequent metacognitive skills such as monitoring of attachment experiences. Specifically, Main suggested that early interaction experiences alter the contents of the child's mind as well as the ability to operate upon these contents (Main, 1991; Sharp & Fonagy, 2008). The relation between attachment security and the development of mentalization at large has been extensively investigated, but the results remain unclear. It seems that there are several factors moderating this relation (e.g., Fonagy et al., 2012). Gergely and Unoka (2008) conclude that the correlations between secure attachment during the infant's first year and the development of the child's theory of mind are not very strong, and that the evidence is more compatible with the alternative view that the latter is due to "an innate social-evolutionary adaptation implemented by a specialized and prewired mindreading mechanism that seems active and functional at least as early as 12 month of age in humans" (p. 59). At the same time, however, they argue that early attachment experiences may be important for the child's capacity to *make use of* this inborn mechanism.

In their later work Fonagy and colleagues (e.g., Fonagy et al., 2012) have substantially broadened their view of mentalization by elaborating it in terms of four dimensions, or polarities: (a) cognitive versus affective mentalization, (b) automatic (implicit) versus controlled (explicit) mentalization, (c) self-oriented versus other-oriented mentalization, and (d) internally focused versus externally focused mentalization. These opposing polarities are said to "represent balanced systems where a dysfunction at one pole may manifest as the unwarranted dominance of the opposite polarity" (Fonagy et al., 2011, p. 106). An evaluation of a person's mentalizing capacity can thus be made by detailing their mentalizing profile, that is, their overall functioning in terms of appreciating their current balance in the respective polarities. Mentalizing is thus not "a static and unitary skill or trait" (Fonagy et al., 2012, p. 19). It is more properly viewed as a "dynamic capacity that is influenced by stress and arousal, particularly in the context of specific attachment relationships" (p. 19). The model also entails a certain "dynamic" in itself as the dimensions are considered conceptually and theoretically distinct, whereas they are expected to empirically involve considerable overlap.

The purpose of this first section of the article is to summarize the theory of mentalization as it appears in the literature today, in terms of these four dimensions, with a particular focus on the suggested timetable for their development, as well as some shortcomings and possible inconsistencies in this account. Because the affective-cognitive differentiation was the first polarity to appear in Fonagy and colleagues' writings, it is treated first. Then the focus shifts to the implicit/explicit and the self/other differentiations, and in particular on how these combine with each other and with the affective-cognitive one. Finally, the discussion turns to the internally/externally focused polarity, which is the most recent addition to the theory.

Cognitive Versus Affective Mentalization

Fonagy et al. (2002) summarized evidence that the understanding of desires may antedate the capacity to understand beliefs by as much as 18 months, and concluded that "[t]his discrepancy in the developmental timetable suggest that separate mechanisms for interpersonal understanding concerning emotions and belief states should be

considered" (p. 137). They referred to the hypothesized underlying neuropsychological structures in terms of an Interpersonal Interpretive Mechanism (IIM), and suggest that "the IIM subdivides anatomically into two substructures: the IIM-a (for affect) and the IIM-c (for cognition)" (p. 137). The IIM-a is exemplified by emotional resonance (empathy), whereas the IIM-c is illustrated by the ability to reason about epistemic states (beliefs).

In their most recent account, Fonagy, Bateman, and Luyten (2012) refer to Baron-Cohen et al.'s (2008) differentiation between a *Theory of Mind Mechanism* and the *Empathising System* to further describe the affective and cognitive dimensions of mentalization. They suggest that "different forms of psychopathology may be distinguished in terms of the inhibition, deactivation, or simply dysfunction of one or both systems involved, leading to potential *dissociations* between both systems or difficulties in *integrating* cognitive and affective aspects of mentalization" (Fonagy et al., 2012, p. 30). For example, some individuals may show substantial understanding of mental states without being in contact with the "affective core" of those experiences, whereas others tend to be overwhelmed by "automatic, affect-driven mentalizing" and therefore lack the ability to integrate the affective experiences with "more reflective and cognitive knowledge" (p. 30). Integration of these is described as "full mentalization" (Fonagy et al., 2012, p. 29).

Although the cognitive/affective dimension was the first dimension to be introduced, the most fundamental polarity underlying mentalizing, according to Fonagy et al. (2012), is the automatic–implicit versus controlled–explicit dimension. This raises the question whether the above quoted discrepancy in the developmental timetable between the mentalizing of affect (desires and emotions) and cognition (belief states) refers to implicit or explicit mentalizing, or both.

Implicit (Automatic) Versus Explicit (Controlled) Mentalization

Controlled or explicit mentalizing reflects a "serial and relatively slow process, which is typically verbal and requires reflection, attention, awareness, and effort," whereas automatic or implicit mentalizing instead "involves parallel and therefore much faster processing; is typically reflexive; and requires little or no attention, intention, awareness, and effort" (Fonagy et al., 2012, p. 20). Fonagy et al. (2012) describe this distinction as one between reflective and reflexive systems and refer to evidence from neuroimaging studies, which suggest two different neural systems as underlying the respective polarities (Lieberman, 2007; Satpute & Lieberman, 2006). Fonagy et al. maintain that in our daily interactions, mentalizing is "predominantly implicit" because we tend to "rely on automatic and unreflective assumptions about ourselves, others, and ourselves in relation to others" in most interpersonal situations (Fonagy et al., 2012, p. 20). When things go smoothly, for instance as in secure attachment relationships, individuals often rely on automatic mentalization because more reflective processing is unnecessary. Yet, if necessary, the individual can flexibly shift to controlled mentalization.

From a developmental perspective, the explicit/implicit differentiation is relevant for the understanding of both cognitive and affective mentalization. With regard to cognitive processes, Fonagy et al. (2007) note that explicit mentalization is seen in people's conscious verbal reasoning, whereas implicit mentalization is indicated by an individual's performance on nonverbal tasks. For example, even though children do not normally pass the traditional false-belief task until around 4 years of age, research

with nonverbal tasks indicate that infants as young as 15 months of age have clear expectations about the behavior of a person with a false belief (Onishi & Baillargeon, 2005). Fonagy et al. (2007) conclude that infants in their first year of life "already possess implicit understanding of intentional action, although the exact nature of these abilities and the implications they have for our understanding of the development of early psychological reasoning remain controversial" (p. 296). They refer to several studies that "seem . . . strongly to suggest that human infants as young as 13 months of age have the mentalising capacity to attribute beliefs to others based on automatic monitoring of the other's perceptual access to the situation" (Fonagy et al., 2007, p. 292; cf. Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007). Moreover, they hypothesize that this kind of "early implicit intuitive mentalising . . . and its failure may have greater social impact than the acquisition in late preschool years of an explicitly representational concept of mind that would be revealed by performance on a false-belief task" (Fonagy et al., 2007, p. 296). Hence, they ascribe the implicit dimension a more fundamental role for the development of social understanding than the explicit one.

However, the findings that are supposed to indicate infants' appreciation of others' false beliefs are not uncontroversial. Other explanations have been presented that do not call for the conclusion that infants attribute false beliefs to the others (e.g., Apperly & Butterfill, 2009; de Bruin, Strijbos, & Slors, 2011; Perner & Roessler, 2012). Actually, the exact nature of the implicit/explicit differentiation remains to settle in detail. There is a risk of extrapolating a model of conceptual thinking to domains of nonconceptual thinking—as if implicit mentalization is just like explicit mentalization, except that it is nonverbal and nonconscious. What is needed is a detailed understanding of the exact characteristics of these different processes. In their discussion of these matters, Perner and Roessler (2012), for example, suggest that (a) the young infant is capable of keeping an *experiential record* of what another individual did or did not perceptually track, and that this experiential record is *activated automatically* whenever the infant attends to that individual, although this does not mean that the infant is able to *intentionally switch to* another individual's perspective; and that (b) although this experiential record *reflects* that agent's beliefs about the world, the infant's implicit representation need "not involve the concept of belief at all (though it may play an important role in facilitating the acquisition of the concept" (p. 523). Further, they suggest that (c) whereas the former involves a *level 1 perspective taking* (an ability to differentiate between what another individual can see and what that individual cannot see), the latter involves a higher *level 2 perspective taking* (an ability to understand that two people may see or interpret the same thing differently depending on their vantage point). This is not the place to discuss which theory may best explain the findings concerning implicit mentalization in infants; suffice it to say here that the issue is far from settled, and that there are good reasons to question the belief–desire model as a model for infants' performance in this area.

With regard to affective processes, Fonagy et al. (2002) speak of the developmental process from implicit to explicit mentalization in terms of a development from "nonconscious primary representations" to the building of "secondary representations" (pp. 185–186). To explain this process, they argue for a "social biofeedback model" that assigns a central role to "the contingent reflective externalizations provided by social partners as the informational basis for rerepresenting the amodal internal structure of nonconscious primary representations" (Fonagy et al., 2002, p.

186). According to the model, this "secondary representation-building process" is assumed to take its start in the young infant in the form of the caregiver's affect mirroring, which leads the infant to discover its own feelings at an explicit level. During the following years the process proceeds in the form of repeated experiences of interactions with a playful caregiver who reflects the child's ideas, fantasies, and emotions, in such a way that the child becomes aware of these.

Although this implies that the interactions with the caregiver is given a central role in the development from implicit to explicit mentalization of *affective* processes, the role of such interactions in the development of the mentalization of *cognitive* processes has been questioned by Gergely and Unoka (2008). According to them,

> the currently popular view endorsed by a number of attachment theorists and infant researchers that assumes a possibly evolutionarily based and human-specific direct causal and functional link between the ontogeny of early security of infant attachment on the one hand, and the acquisition of explicit mentalizing skills on the other, must be significantly revised. The relevant developmental and comparative evidence of recent years seems more compatible with the alternative view that an implicit and automatic capacity for mentalizing about others is not a developmental achievement, but an innate social–cognitive adaptation implemented by a specialized and prewired mindreading mechanism that seems active and functional at least as early as 12 months of age in humans (pp. 58–59).

The quality of early attachment, however, is still assumed to play a significant role for the child's later ability to *use* this capacity for mindreading "in coping with interpersonal interactions and relationships during childhood and adulthood" (Gergely & Unoka, 2008, p. 59). For example, "when maladaptive patterns of affective parental reactivity become the dominant features of the reoccurring interactive structure of the infant's primary attachment relationships, they can play a significant causal role in pathologically *undermining* the developing self's potential to rely on its innate mindreading capacity" (pp. 59–60). In other words, although the quality of early attachment is not so important for the development of the child's basic competence in mentalizing others' cognitive processes, either explicitly nor implicitly, it is probably of importance for the developing individual's ability to mentalize in emotionally stressful interactions with others.

It should also be noted that Gergely and Unoka (2008) refer specifically to implicit mentalizing about *others,* and only with regard to their *cognitive* processes (i.e., "mindreading"), in their discussion of what is innate about mentalization. That is, the distinction between explicit and implicit mentalization is only applied to the mentalizing of others' cognitive processes—not to the child's own cognitive processes, and not to affective processes, neither in self or others. For example, they do not discuss whether there is also an innate implicit mentalization of others' *affective* processes; nor do they discuss if the infant also engages in an implicit mentalization of *its own* psychological processes, and if so, whether this capacity is innate. This naturally takes us to the third polarity in the theory of mentalization, that is, the self/other polarity. The purpose of the following sections is, therefore, to summarize and discuss the theory of mentalization with regard to (a) mentalizing of cognition in oneself, explicit versus implicit; (b) mentalizing of affect in oneself, explicit versus implicit; and (c) mentalizing of affect in others, explicit versus implicit.

Mentalization of Cognition in Self, Explicit Versus Implicit

Although the literature on the development of mentalization includes many references to research on the mentalizing of *other's* cognitions (as seen in research with the false belief test, and other studies of children's mindreading), it contains little mention of

research on the development of the child's ability to mentalize *its own* cognitions. The latter topic is generally studied in research on metacognition, as defined in terms of the monitoring and control of cognitive processes (e.g., Flavell, 1979). Traditionally, metacognition has been conceived in terms that do not permit metacognition in preverbal infants. Lately, however, the field of metacognition has undergone a similar development to that of research on the child's theory of mind—that is, capacities that were originally studied only in explicit form, by methods that relied on verbal report, have been identified at much earlier ages when operationalized in terms of nonverbal performance (e.g., Agnetta & Rochat, 2004; Balcomb & Gerken, 2006). Fitting metacognition into the scheme of dimensions of mentalization requires a consideration of whether the implicit/explicit differentiation is applicable also to metacognition. The theory of mentalization, however, is silent here—we are not aware of any application of the distinction between implicit and explicit mentalization to metacognition. This, therefore, represents a first unclear point in the theory.

There are, however, other theoretical attempts to outline principles of implicit metacognition. For example, Proust (2007) developed a model of implicit metacognition considered as an adaptive control system consisting of regulated and regulating subsystems, which does not require the involvement of metarepresentation. Elaborating further along partly similar lines, Brinck and Liljenfors (2013a, b) differentiated between implicit metacognition, perceptual metacognition, and metarepresentational metacognition. Implicit and perceptual metacognition are assumed to rely on heuristics and environmental affordances, without any use for algorithms and metarepresentations. Although perceptual metacognition does not rely on verbal language or metarepresentations, however, it does involve consciousness and attention-based strategies; although it is not "explicit" in being verbally expressed, it is "controlled" in the sense that it requires attention, awareness, and effort. Metarepresentational metacognition, on the other hand, requires higher-order propositional or symbolic strategies which develop first with language acquisition. Once the child can use language, implicit and perceptual metacognition can be coupled to a representational system that enables metarepresentation and allows for a higher-order redescription of implicit and perceptual metacognition in terms of algorithmic strategies operating on propositional representations. This does not mean, however, that implicit and perceptual metacognition disappear and are *replaced* by metarepresentational forms of metacognition. "On the contrary, implicit, perceptual, and metarepresentational metacognition coexist in adult subjects, each type contributing in its particular way to the general metacognitive machinery" (Brinck & Liljenfors, 2013a, p. 86).

In parallel to the discussion of the role of early interaction with the caregiver for the development of mindreading, Brinck and Liljenfors (2013a) discuss the role of early dyadic interaction for the development of metacognition. Here they make use of Trevarthen's (1992) concept of "primary intersubjectivity," defined as an innate drive in the infant to take part in dialogue, which involves an inborn receptivity to the subjective states of other persons. Is such a model of metacognition compatible with the theory of mentalization?

This model of metacognition also raises a wider question of relevance to mentalization theory: Is the distinction between explicit and implicit mentalization optimal, or would it be of help to divide explicit mentalization into perceptual and metarepresentational—the former relying on attention, awareness, and effort but not on verbal language, and the latter introducing language into the picture? Brinck and Liljenfors (2013a, p. 90) refers to perceptual metacognition as "the experiential dimension of

metacognition"—would it be relevant to speak also of an experiential dimension of mentalization more generally?

Mentalization of Affect in Self, Explicit Versus Implicit

A central tenet of Fonagy et al.'s (2012) theoretical approach is that the capacity for both self- and other-oriented mentalization develops in the context of attachment relationships; the child "observes, mirrors, and then internalizes his or her attachment figure's ability to represent and reflect on internal mental states" (p. 25). This is especially emphasized in the discussion of the child's mentalizing of its own affects, which is also in accordance with the empirical evidence. As pointed out by Gergely and Unoka (2008), although the facilitating effects of early secure attachment on theory of mind development are not very strong, "in several studies they only show up in theory of mind tasks that involve reasoning about emotions as well (such as belief-desire tasks) and not only purely epistemic states such as false beliefs" (p. 56).

The child's ability to mentalize its own affects, according to Fonagy et al. (2002), develops for the first time in the context of affect regulation in early, dyadic, intersubjective interactions, where the infant's affects are mirrored by the parent, or caregiver. The parents are described as using their facial and vocal expression to represent to the child the feelings they assume the child to have, and as doing this in such a way as to reassure and calm rather than intensify the child's emotions. For this process to occur in an optimal way, two things are required: (a) the parents have to mirror the child's feelings in a sufficiently *congruent* or "attuned" fashion, and (b) the affect mirroring must be *marked* as indicating an affect belonging to the infant, and not as an expression of the parents' own affect.

This kind of "marking" can be achieved "by producing *an exaggerated version* of the parent's realistic emotion expression, similarly to the marked "as-if" manner of emotion display that is characteristically produced in pretend play" (Fonagy et al., 2002, pp. 177–178). If it were not "marked," the child would not be able to distinguish the mirrored expression from the parent's own emotional expression—or, in other words, without the "marking," the affects that are expressed would not be perceived by the child as *representing* the child's own affects. Fonagy et al., 2002 conceive of markedness and its representational consequences as a "further and separate evolutionary function" of so-called infant-directed speech (or motherese), which consists in slightly exaggerated, prosodic patterns that caregivers across the world use in communicating with their infants (p. 178).

If affect mirroring is sufficiently congruent and marked, it makes the infant able to construct a representation of the parent's emotional display. This representation may thereby come to function "as a *secondary representational structure*, which will become activated through associative routes whenever the set of internal-state cues corresponding to the given dispositional emotion state are activated in the infant" (Fonagy et al., 2002, pp. 180–181). The underlying mechanism for this process, according to Fonagy et al. (2002), is the contingency detection module (CDM) originally described by Gergely and Watson (1999). This module is an innate mechanism for detecting contingency relations between types of responses and their environmental effects. During the first 2–3 months of life, it is genetically set to seek out and explore how physical actions of the self affects perceived stimuli—a process which is basic to the differentiation between self and external reality. At approximately 3 months, however, the infant's preference changes to an exploration of the social world. The infant focuses on the regularities in affect

mirroring, where parents' emotional displays come to be linked to the infant's own emotional states without being perfect reflections of these. By becoming familiar with the regularities of affect mirroring through the interaction with the caregiver, the infant is said to be *sensitized* to those "sets of internal-state cues" that are indicative of various emotional states, as reflected in the affect mirroring from the caregiver. Thereby the infant becomes able to "detect and group together" the regularities into "categorically distinct dispositional emotion states" (Fonagy et al., 2002, p. 201). To summarize, the hypothesis is that the infant discovers his or her own affects by means of the caregiver's emotional expressions in the affect mirroring process. The emotional expressions are said to be internalized as "'proto-symbolic' emotion representations in the baby′s awareness" (p. 181).

In Gergely and Unoka's (2008) description, this process is formulated in terms of a "social construction of the representational affective self." The infant's *subjective sense of differential emotional states* is said to be established as a result of repeated experiences of patterns of parental reactivity and social "mirroring" feedback evoked by the infant's automatic expressions of *initially nonconscious basic emotional arousal states*. This process, which means that the "basic emotion states of the primary constitutional self . . . become introspectively accessible" (Gergely & Unoka, 2008 p. 62), requires that two basic conditions are satisfied: (a) "the primary and procedural basic emotion programs of the constitutional self (which are prewired, initially nonconscious automatisms) need to become associated with *second-order representations* that, when activated, are cognitively accessible to introspectively oriented attentional self-monitoring processes" (Gergely & Unoka, 2008 p. 62); and (b) the primarily externally oriented attention systems of the infant "*need to become introspectively sensitized* and turned toward the self's internal states to a sufficient degree" (Gergely & Unoka, 2008 p. 63).

As described by these authors, there is a development from what seems to be a purely *nonmentalizing* state ("initially nonconscious basic emotional arousal states" that have the form of "procedural basic emotion programs," or "prewired, initially nonconscious automatisms") to a capacity for *explicit mentalization of* affective states ("second-order representations" that are "cognitively accessible to introspectively oriented attentional self-processes"). Clearly, although Gergely and Unoka (2008) admit that "the infant's innate biological or constitutional self . . . contains a basic set of prewired universal emotional categories that are primary biological adaptations" (p. 61), they do not include any implicit form of self-oriented mentalization of affect as part of this innate organization—having "nonconscious emotional arousal states" apparently does not require any form of "metaemotion" apart from the emotion itself. Neither do they seem to involve experiences in any interesting sense. Their description of mentalization here does not seem to leave any room for a level of implicit mentalizing either, as the process is described in terms of the development of "second-order representations" of affects that are "introspectively accessible"—which clearly imply processes of conscious explicit mentalization. Similarly, the use of formulations like "awareness," "a subjective sense of differential emotional states," "discovering one's own affects," and so forth, in the description of the process do suggest that what occurs is a development of explicit mentalization. This raises the question if the differentiation between implicit and explicit mentalization is at all applicable in the development of child's ability to mentalize its own affects. In other words, is there any role for implicit mentalization in what Gergely and Unoka (2008) refer to as "the social construction of the representational affective self"? And if not,

what are we then to make of Fonagy et al.'s (2012, p. 20) assumption that the implicit/explicit distinction represents the "most fundamental polarity underlying mentalizing." This represents a second unresolved issue in the theory of mentalization.

Mentalization of Affect in Others, Explicit Versus Implicit

What about the development of other-oriented mentalization of affect? As part of their argument, Fonagy et al. (2002) deny that the infant perceives the caregiver's affects in the affect mirroring process, because this would imply that the young infant was capable of some form of primary intersubjectivity—an assumption they deny. On the contrary, they claim that "the infant is unaware that he is seeing the other's subjective state. It is likely that the infant does not yet know that others have internal feelings" (p. 127). However, this particular position creates a problem for the explanation of how mentalization develops. Given that the infant is incapable of perceiving emotional qualities in the caregiver's voice and face expressions, how then is it able to perceive whether the caregiver's affect mirroring is actually *congruent* with the child's own affective state? Further, how can the child perceive it as *marked* in a way that indicates that the caregiver refers to the *infant's* own affective state and are not expressions of *the caregiver's* own affect, if it is not capable of perceiving emotional qualities in the caregiver's voice and face expressions?

The problem for mentalization theory in this regard is that the experience of congruence seems to require the ability to compare the emotional qualities perceived in oneself with those perceived in the caregiver. In the same way, the experience of "markedness" would require the ability to first compare and then differentiate between one's own feelings and those of the caregiver. However, according to Fonagy et al. (2002), the process of affect mirroring leads the infant to discover its own affects. Yet, would this be possible without the infant already being able to perceive affects in the parent's face and voice? If not, this seems to imply that the child is *first* able to perceive affect in *others*, and then becomes able to perceive its own affects. However, this is not how Fonagy and his colleagues reason. In fact, they claim not only that the infant is "unaware that he is seeing the other's subjective state," but also that "at this level of human proximity the other's subjective state is automatically referred to the self" (Fonagy et al., 2002, p. 127). Now, this claim creates a further problem concerning the role of markedness in the theory: If it is the case that the other's state is automatically referred to the self, then what function does markedness serve? Markedness, we are told, is what helps the infant "decide" whether the expressed emotion belongs to the parent or refers to the infant's own affect. But if there already is a mechanism that settles the issue, why do we need markedness?

One possible solution would be to assume that, although the child does not yet have the capacity for an explicit mentalization of others' affects, it does have the capacity for an *implicit* mentalization of others' affects. The latter would not require the infant to be aware of the other's subjective state, or knowing that the other has internal feelings; it would merely require an ability to *differentiate between others' emotional states at a nonverbal level of performance.* Although Gergely and Unoka (2008) assume that the infant has both an innate "prewired mindreading mechanism" (pp. 58–59) for mentalizing others' cognitive processes, and "a basic set of prewired universal emotional categories that are primary biological adaptations" (p. 61), they do not state that the ability to differentiate between emotional qualities in others behavior is also prewired. Moreover, the possibility that the ability to mentalize

others' affective states would be a prewired characteristic of the infant's constitutional self was rejected by Fonagy et al. (2002) as inviting notions of primary intersubjectivity. This, therefore, represents a third unresolved issue in the theory of mentalization.

Internally Versus Externally Focused Mentalization

The dimension of internally and externally focused mentalization has been added rather recently to the theory of mentalization. The distinction is formulated as follows:

> Internally focused mentalization refers to mental processes that focus on one's own or another's mental interior (e.g., thoughts, feelings, experiences), whereas externally focused mentalizing refers to mental processes that rely on physical and visible features and one's own or another's actions (Fonagy et al., 2012, p. 22).

In applying this distinction to the development of mentalization in children, Fonagy et al. (2012) state that, in the affective mirroring process, the infant has to work out what a "marked" emotion display refers to, that is, what internal state underlies the emotion. To accomplish this he or she must rely on *external* cues such as the caregiver's eye-gaze direction that accompanies the communicative display. Because the caregiver faces and looks at the infant while making these marked emotion mirroring displays, the effect will be that the infant will attend to "his or her own face and body—that is, his or her own external physical self as the referent that the caregiver's cues indicate and to which the marked (and decoupled) affect display should be referentially anchored" (p. 24). For this process to work, the infant must be capable not only of externally focused mentalization (i.e., respond to the caregiver's emotional expression) but also of internally focused mentalization of the caregiver's intentions. That is, the child must be able to use the external cues that the caregiver communicates (e.g., eye-gaze direction) to "work out" that the caregiver has the infant's emotional expressions "in mind." The affect mirroring process thus involves "a continuous back-and-forth between *external* and *internal* features of self and others" (Fonagy et al., 2012, p. 24). For this process to succeed the caregiver's affective expressions have to be contingent on the infant's affect—if not, the caregiver instead risks undermining the adequate labeling of the internal states. In that case, the infant will not establish the introspectively accessible second-order representations with the effect that the internal states will remain "confusing or frightening and experienced as unsymbolized and difficult to regulate" (p. 24).

Whereas the other three dimensions (cognitive/affective, implicit/explicit, and self/other) seem to be orthogonal to each other, this does not seem to be the case with the suggested external/internal dimension. For example, whereas this differentiation seems easily applicable to emotions, which have both external expressions and internal feeling qualities, this does not seem to be the case with cognitions, which do not have bodily expressions to any similar degree. At the same time, an interesting thing about this dimension is that it implies a view of mental processes which transcends the external/internal duality, in the sense that the concept of mental processes is not restricted to internal events, if "internal" is defined as "within the body." On the contrary, an important aspect of at least some mental processes (primarily emotional processes) is that they include outward bodily processes that are publicly observable; externally focused mentalization refers to the ability to grasp the meaning of these bodily processes.

With regard to the self/other polarity, it would seem that externally focused mentalization is primarily relevant with regard to others, whereas it is more of an open question what role the mentalization of one's own bodily expressions may have for self-understanding. On the other hand, it seems clear that externally focused mentalization may

be either implicit or explicit. For example, when Fonagy et al. (2012) state that "patients with BPD [that is, borderline personality disorder] are often hypersensitive to facial expressions . . ." (p. 22), this may be seen as an implicit form of externally focused mentalization. On the other hand, when they state that patients with antisocial personality disorder may lack the ability "to read fearful emotions from facial expressions" (p. 23), they are obviously referring to a more explicit form of externally focused mentalization.

To conclude, whereas the external/internal dimension seems easily applicable to emotions, which have both external expressions and internal feeling qualities, it is perhaps not prima facie as evident with cognitions, which do not necessarily involve any similar bodily expressions. And similarly, whereas externally focused mentalization seems more applicable to the mental processes of others than to those of oneself, internally focused mentalization seems more applicable to mental processes in oneself than in others. This raises the question whether the external/internal differentiation should really be rendered the status as a basic dimension of mentalization. Maybe the external/internal polarity is rather an aspect of mentalization that is relevant primarily for emotional processes in others? This represents a fourth unclear point in the theory. Essential here is also that the notion of an externally focused mentalization transcends the external/internal duality in its implications with regard to the ontological status of psychological processes; in this regard, it bears a certain degree of resemblance to the notion that primary intersubjectivity functions through embodied expressions of psychological states.

Summary: Some Unresolved Problems

To summarize, this means that we have identified at least four unresolved problems, or unclear points, in the theory of mentalization:

1. What is the role of implicit and explicit processes in the development of the infant's mentalization of its own cognitive processes (i.e., the development of metacognition)? Although mentalization theory is silent on this topic, there are other theoretical models that may possibly help to fill this gap. Brinck and Liljenfors' (2013a) differentiation between implicit, perceptual and metarepresentational metacognition may possibly be relevant here; further, their theory that metacognition has its developmental origin in primary intersubjectivity suggests that it may be of interest to explore the relation between mentalization and intersubjectivity in more detail.
2. What is the role of implicit and explicit processes in the infant's developing ability to mentalize its own affective processes? The descriptions provided by Fonagy et al. (2002) and Gergely and Unoka (2008) do not make any use of the implicit/explicit differentiation in the development of this process. Does this mean that the implicit/explicit distinction is not applicable here—but, in that case, what are we then to make of Fonagy et al.'s (2012) statement that the implicit/explicit distinction represents the most fundamental polarity of mentalizing? Could it be that the same applies here as was suggested by Brinck and Liljenfors (2013a) in the case of metacognition—that there is room for implicit processes and primary intersubjectivity in the understanding of the development of the mentalization of affect in the self?
3. Fonagy et al.'s (2002) description of the affect mirroring process gives a central role to the infant's capacity to perceive whether the caregiver's affect mirroring is (a) congruent with the child's own affective state, and (b) "marked" in a way that indicates that the caregiver refers to the *infant's* own affective state and are

not expressions of *the caregiver's* own affect. This seems to imply that the infant must be able to perceive emotional qualities in the caregiver's voice and face expressions already from the start—otherwise the affect mirroring process cannot get started. One possibility would be that the infant has an inborn ability to implicitly mentalize others' emotions, in the form of a prewired capacity to differentiate between emotional qualities in others' behavior This possibility, however, was rejected by Fonagy et al. (2002) as inviting notions of primary intersubjectivity. Gergely and Unoka (2008, p. 60) similarly reject explanations in terms of primary intersubjectivity in favor of their "social constructivist view of affective self development." The problem, however, is that the latter theory does not suggest any solution to this unresolved problem in the theory of mentalization.

4. Does the external/internal differentiation really refer to a separate dimension of mentalization, or is it more correct to see it as an important aspect in the child's developing ability to mentalize emotional processes in others?

As seen above, the notion of primary intersubjectivity has been raised as a possible solution to three of these four unresolved issues in the theory of mentalization, although we have not as yet explained how. Both Fonagy et al. (2002) and Gergely and Unoka (2008), however, have vehemently opposed notions of primary intersubjectivity. This raises the question about what is meant by "primary intersubjectivity," and whether this construct can be of any real help here or has to be discarded from further discussion. In the following section, it will be argued that there are different interpretations of the notion of "primary intersubjectivity" in the literature—one "mentalistic" which has to be discarded, and one nonmentalistic which may in fact be integrated with the theory of mentalization in a way that helps to resolve the unsettled issued that have been described above.

Intersubjectivity and Mentalization

The relation between attachment and mentalization is a complex matter that has been touched on briefly in the first part of the article. Another important question, however, refers to the relation between *intersubjectivity* and mentalization. There are several reasons why it is important to distinguish between intersubjectivity and attachment. Cortina and Liotti (2010), for example, claim that from an evolutionary perspective, attachment and intersubjectivity most likely serve different functions. Whereas the attachment system primarily concerns the function of seeking security and protection, including the forming of mental representations, that is, working models, that eventually warrant a sense of security even in the absence of the "secure base" (e.g., Bretherton, 2005; Sroufe & Waters, 1977), the function of intersubjectivity is to communicate "at intuitive and automatic levels" (p. 410) in order to facilitate cooperation and social understanding. Cortina and Liotti suggest that intersubjectivity is a separate motivational system, with motives (intimate sharing and altruistically cooperating with others) that have a less intense quality than those associated with the attachment system. Further, they argue that the conditions under which intersubjective motives become active are more general than the more specific conditions that activate attachment motives. Lyons-Ruth (2007) makes a similar point, noting that intersubjective communication has a ubiquitous role that permeates all other mental functions and makes us uniquely human. According to Lyons-Ruth, the human attachment system is "filtered through and mediated by the

increasingly complex intersubjective processes that emerge from birth onward." (p. 599). Thus, she argues that research should be more focused on intersubjective processes, such as exchange of affective cues, not only seen as their function in soothing and modulating negative affects like fear, but also as including positive affect and what it means in terms of facilitating communication and social understanding.

In developmental psychology, intersubjectivity is usually defined as the sharing of experiences (e.g., Stern, 1985; Trevarthen & Hubley, 1978), which requires the complementary abilities of (a) recognizing the experiences of another individual (a second person); and (b) making available one's own (first person) experiences to another individual (Brinck, 2008). To account for the emergence of intersubjectivity, Trevarthen (1992) construed primary intersubjectivity as an innate drive that accounts for infants' readiness to partake in dialogue, which is akin to Cortina and Liotti's (2010) view of intersubjectivity as a separate motivational system.

How then does the concept of primary intersubjectivity relate to the dimensions of mentalization? An analysis of the literature actually reveals at least two different interpretations of the construct, with widely different implications. According to one interpretation, the "nonmentalistic" one, primary intersubjectivity overlaps largely with *implicit* and *externally focused* mentalization, although all forms of primary intersubjectivity need not necessarily involve any mentalization at all. In other words, "nonmentalistic" here is equivalent to an absence of explicit mentalization, and internally focused mentalization. According to the other interpretation, the "mentalistic" one, on the other hand, the theory of primary intersubjectivity implies attributing a capacity for *explicit* mentalization and *internally focused* mentalization to infants. Although Fonagy et al. (2002) as well as Gergely and Unoka (2008) have good reasons to reject this mentalistic interpretation of primary intersubjectivity, they thereby also unfortunately have thrown out the possibility of a nonmentalistic interpretation of primary intersubjectivity.

Primary Intersubjectivity—The Nonmentalistic Interpretation

A major representative for a nonmentalistic interpretation of primary intersubjectivity is Gallagher (2004), who defines primary intersubjectivity as "the innate or early developing capacity to interact with others manifested at the level of perceptual experience" (p. 204), and characterizes it as a "nonmentalistic" mode of interaction (p. 205). Among other things, this means that

> pretheoretical (nonconceptual) sensory-motor capabilities for understanding others already exist in very young children. Infants already have a sense from their own proprioception and movement of what it means to be an experiencing subject-agent. They can sense that certain kinds of entities (but not others) in the environment are indeed subject-agents like themselves; and that in some way these entities are similar to and in other ways different from themselves. This sense is implicit, at least in a primitive way, in the behavior of the newborn. Infants from birth are capable of perceiving and imitating facial gestures presented by another (Meltzoff & Moore, 1997, 1994). This interaction depends not only on a distinction between self and nonself, and a proprioceptive sense of one's own body, but on the recognition that the other is in fact of the same sort as oneself (Bermúdez, 1996; Gallagher & Meltzoff, 1996). Infants are able to distinguish between inanimate objects and people (agents). They can respond in a distinctive way to human faces, that is, in a way that they do not respond to other objects (Legerstee, 1991; Johnson, 2000; Johnson et al., 1998) (Gallagher, 2004, p. 205; all references cited in Gallagher, 2004).

In this context, Gallagher (2004) also refers to what Baron-Cohen (1995) calls an

"intentionality detector" (ID) and an "eye-direction detector" (EDD). The ID is seen as an innate capability that allows the infant to interpret, "perceptually and nonmentalistically, rather than theoretically, the bodily movement of others as goal-directed intentional movement" (Gallagher, 2004, p. 205). The EDD, on the other hand, allows the infant to follow the gaze of another person, and thus to "sense what the other person sees (which is sometimes the infant herself), in a way that throws the intention of the other person into relief" (p. 206). What these capacities seem to involve are processes that, although they need not involve conceptual thinking or conscious awareness in the infant, still require an ability to perceive external expressions of mental processes.

In terms of mentalization theory, the capacities from the quote above seem to overlap primarily with *implicit* mentalization and *externally focused* mentalization. The overlap with implicit mentalization is seen in Gallagher and Hutto's (2008) way of characterizing primary intersubjectivity as "fast, automatic, irresistible and highly stimulus-driven," and the overlap with externally focused mentalization is seen in their description of it as involving a capacity "to see bodily movement as goal-directed intentional movement." But there is no complete overlap, because all manifestations of primary intersubjectivity need not involve even an implicit and externally focused mentalization.

An example of primary intersubjectivity which does involve implicit mentalization is seen in research evidence that infants at 10–11 months are able to parse ongoing behavior along boundaries correlated with the initiation and completion of intentions (Baldwin, Baird, Saylor, & Clark, 2001). On the other hand, an example of primary intersubjectivity which does *not* seem to require a capacity for implicit mentalizaton is neonatal imitation. Neonatal imitation is evidence of a "primary, perceptual sense of others" and of "a responsiveness to the fact that the other is of the same sort as oneself" (Gallagher & Hutto, 2008, p. 21). Further, because infants imitate only *human* faces, it may be concluded that they are "able to parse the surrounding environment into those entities that perform human actions (people) and those that do not (things)" (p. 21). As concluded by Gallagher and Hutto, this means that there seems to be a functioning intermodal tie between a proprioceptive sense of one's body and a visually perceived face already at birth. Does this mean, then, that neonatal imitation is a probable start for intersubjectivity? To serve even as a possibility, we must explain what imitation might mean.

If imitation is taken to mean *copying*, then it seems unconvincing to ground any form of intersubjectivity in it. But copying may not be the best way of understanding (neonatal) imitation. For example, Rochat, Passos-Ferreira, and Salem (2009) argue that imitation involves *reciprocation* in the sense of constituting a mutual interaction that (potentially) engages the other person. Along the same lines, Meltzoff and Moore (1997) argue that imitation is representationally mediated. Importantly, they develop their argument based on studies in which target imitative acts can be corrected to achieve "a more veridical match." This means, Meltzoff and Moore say, that the evidence speak in favor of the view that "information about the infant's own acts is available to comparison to a representation of the adult's act" (Meltzoff & Moore, 1997, p. 188). They also review two lines of evidence that suggest a more differentiated process than involving and requiring simple "transduction." A first line of evidence shows that the infant's response" need not be temporally coupled to the stimulus," and a second line of evidence shows that "imitation is not compulsory; infants need not produce what is given to perception" in a current situation (Meltzoff & Moore, 1997, p. 181). The studies referred to involved, first, infants differentially imitating gestures they saw 24 hr earlier, and second, infants, who met a first person showing them new gestures, repeated these gestures in a meeting shortly after with a second person instead of imitating the gestures Person 2 expressed on that occasion.

Thus, (neonatal) imitation seems at the very least equipped for serving important functions for developing intersubjective functions thus constituting primary intersubjectivity. But still, this does not require any involvement of implicit mentalization.

A basic requirement for a psychological process to be categorized as a form of mentalization is that it refers to (or implies) a mental process, either in oneself or another individual—this may be called *metamentality*. Although neonatal imitation seems to imply a *mental ability to differentiate* between people and things, and to see other people as in some sense similar to oneself, this does not imply the existence of any *meta*-mentality. This is in clear contrast to the evidence that 10–11 month old infants are able to parse human behavior according to intentional boundaries of the kind studied in Baldwin, Baird, Saylor, and Clark's (2001). Even though this does not require any conceptual understanding of intentions, it does require an ability to detect intentions, and to differentiate intentional actions from other events; intentions in that sense represent a meaningful category in infant's performance, and this would thereby count as an example of metamentality.

Another interesting implication of this is that the first manifestations of primary intersubjectivity, as seen for example in neonatal imitation, obviously precede the development of implicit mentalization. At the same time, because neonatal imitation shows evidence that the infant is able to differentiate between people and things, and to see other people as similar to oneself, these forms of primary intersubjectivity may also possibly represent important causal conditions for the development of implicit mentalization. First, because primary intersubjectivity as defined above is conceptually independent of implicit mentalization, it opens up for the possibility that primary intersubjectivity might be causally involved in the development of implicit mentalization. Second, by involving both a differentiation between people and things, and between self and other, while at the same time categorizing others as similar to oneself, primary intersubjectivity provides a "setting" for the development of implicit mentalization. In other words, it is possible that the neuropsychological systems underlying primary intersubjectivity are a basic prerequisite for mentalization to start to develop, and that the nature of interaction that takes place between infant and adult in the context of early primary intersubjectivity have an influence on the development of mentalization. Jaffee, Beebe, Feldstein, Crown, and Jasnow show in their study that vocal rhythm coordination at 4 months predicts attachment at 12 months. This result is interesting in relation to the arguments from Gergely and Unoka (2008) concerning the connection between attachment and mentalization. Whereas it is hard to find any clear evidence of attachment predicting results in mentalization—whether implicit or explicit—the relation may instead by the opposite, that is, the interactional patterns revealed in Jaffe et al.'s study, consist of the abilities termed primary intersubjectivity, that is, sharing the experience of coordinating joint expressions. Interestingly, in relation to the notion that whereas attachment is typically formed with a special person—the attachment figure—intersubjectivity is available to other people as well that are willing to engage in play, in Jaffee et al.'s study the interaction with a stranger more strongly predicted the attachment result at 12 months than did the interaction with the attachment figure. Stern (2001) explains that this probably is due to a certain vigilance in the relation with the stranger that makes the interaction a bit more "tight," whereas the interaction with the attachment figure (the parent) entails a more loving and therefore relaxed interaction.

Another aspect of primary intersubjectivity, which is stressed by Gallagher and Hutto (2008), is that it is primary not only in the developmental sense, but also remains primary in an experiential sense throughout adulthood and imbue the interactions of adult human beings:

> it remains primary across all face-to-face intersubjective experiences, and it underpins those developmentally later, and occasional, practices that may involve explaining or predicting mental states in others (see, e.g., Stern's (1985) idea of a "layered model" in which developmentally primary understandings are not "superseded" but remain and operate in parallel to more advanced ones) (Gallagher and Hutto, 2008, pp. 22–23).

Taking primary intersubjectivity as primary also in this second sense means that it continues to be a significant part of social understanding also in adult life. Pursuing the analogy with implicit mentalization, this may be compared with Allen's (2006) formulation that when we mentalize implicitly, we do this "intuitively, procedurally, automatically, and nonconsciously." We mentalize others implicitly for instance in conversations as "we take turns and consider the other person's point of view" (p. 10). Also, when we perceive and respond to others' emotional states "we automatically mirror them to some degree, adjusting our posture, facial expression, and vocal tone in the process (p. 10). He adds that if we were to attempt all of this explicitly, we would probably "come across as stiff and wooden rather than naturally empathic" (p. 10). Fonagy et al. (2012) confirm that in our daily interactions, mentalizing is "predominantly implicit" and "more reflective processing is unnecessary . . . Indeed, given the speed with which most interpersonal encounters unfold, controlled mentalizing may actually hamper interactions rather than facilitate them" (Fonagy et al., 2012, p. 20). Thus, the view that implicit mentalizing is more fundamental than is explicit mentalizing is clearly compatible with Gallagher's and Hutto's claim.

To conclude, what is referred to as "primary intersubjectivity" by writers like Gallagher (2004; Gallagher & Hutto, 2008) in part represents conditions that exist before the development of mentalization, and in part processes of implicit mentalization mentalization in Fonagy et al.'s (2012) sense. As argued above, this also opens for the possibility that primary intersubjectivity is causally involved in the development of mentalization.

Primary Intersubjectivity—The Mentalistic Interpretation

In contrast to this nonmentalistic interpretation of primary intersubjectivity, there is also a mentalistic reading of this construct, and it is this mentalistic version of the construct that is criticized by Fonagy et al. (2002) and Gergely and Unoka (2008). Fonagy et al. (2002) make it entirely clear that what they are discussing is "a truly mentalistic intersubjective stance" (p. 222), and that what they are objecting to is the assumption that this "mentalistic intersubjective stance" is available to the child before the second year of life. The mentalistic version of primary intersubjectivity that they have foremost in mind is derived from Trevarthen (1979) and is referred to as "the strong intersubjectivist position. Basically, it

> assumes (a) that human infants are born with innate mechanisms to identify and attribute mental states such as intentions and feelings to the other's mind during early contingent social interactions, (b) that from the beginning of life there is a relatively rich set of differentiated mental states of the self such as emotions, intentions, motives, and goals that are introspectively accessible to the infant, and (c) that such subjective mental states of the self can be recognized as being similar to corresponding mental states of the other and, as such, are experienced as "being shared" with her (Fonagy et al., 2002, p. 210).

On the basis of a review of research in the area, Fonagy et al. (2002) argue that there is no evidence for attributing such capacities to the infant, and that the child's performance can be explained much more parsimoniously. Overall, they argue for a developmental process where "subjectivity in the infant cannot be assumed but, rather, must be considered as acquired in the process of interaction" (Fonagy et al., 2002, p. 218). This also

means that the emergence of what they call "a truly mentalistic intersubjective stance" (p. 222) sometime during the second year results from the ripening of those representational abilities that make possible a "causal mentalistic interpretation of actions in terms of intentional mind states both for the other and for the self" (p. 222). Fonagy et al. (2007) contend that the thesis of primary intersubjectivity "leaves little room for developmental changes to the subjective sense of self induced by social environmental factors, producing individual variability in the quality and content of subjective affective and mental states across different persons" (pp. 290–291). The reason why intersubjectivity before the second year is not possible, they argue, is that it "involves a 'rich' mentalistic interpretation of the nature of the young baby's subjective experience of her own as well as of the caregiver's mind states during the organized patterns of mother–infant interactions from birth" (p. 292). That is, the notion of intersubjectivity is here by definition assumed to involve explicit and internally focused mentalization.

This mentalistic interpretation of primary intersubjectivity is obviously very different from the nonmentalistic one given by Gallagher and Hutto (2008) when they state that the early capabilities involved in primary intersubjectivity constitute "an immediate, nonmentalizing mode of interaction" (p. 21). Infants are "able to see bodily movement as goal-directed intentional movement, and to perceive other persons as agents. This does not require advanced cognitive abilities; rather, it is a perceptual capacity that is 'fast, automatic, irresistible and highly stimulus-driven'" (p. 21). Whereas the nonmentalistic version of the construct at most implies *implicit* mentalization and *externally focused* mentalization, the mentalistic version starts from the assumption that primary intersubjectivity requires the involvement of *explicit* mentalization and *internally focused* mentalization.

On the other hand, it seems that the critique of primary intersubjectivity as interpreted mentalistically was formulated by Fonagy et al. (2002, 2007) and repeated by Gergely and Unoka (2008) well before Fonagy et al. (2012) outlined the dimensions of mentalization as a conceptual framework for understanding the varieties of mentalizing, and well before Fonagy et al.'s (2012) statement that mentalizing is predominantly implicit. It may be the case that, with the latter developments in mentalization theory, the notion of primary intersubjectivity may be worth a second look (at least in its nonmentalistic version) to see if it can contribute to the development of mentalization theory.

For example, Gergely and Unoka's (2008) list of what characterizes early human caregiver-infant interactions includes "a particular species-unique pattern of 'proto-conversational' *turn-taking contingency structure*" (p. 52) involving "an early sensitivity to, preference for, and spontaneous motivation to engage in *highly response-contingent stimulus events* characteristic of the patterns of contingent interactive reactivity produced by infant-attuned social partners" (p. 52). How does this differ from Trevarthen's (1979) notion of primary intersubjectivity as an innate drive that accounts for infants' readiness to partake in dialogue? Is the difference that the infant's relational competence, in Gergely and Unoka's account can be described in terms of an entirely "mechanistic" interaction, without implying any implicit mentalization of the caretaker's expressions? And without ascribing the infant's experience any interesting role in the process? For example, can "proto-conversational turn-taking" be described mechanistically? Can the "innate sensitivity to and preference for ostensive-communicative cues . . . such as eye contact" (pp. 52–53), and the "spontaneous tendency to attend to and gaze-follow specific behavioral referential cues (such as gaze shift or head movement), but only if these are presented in an ostensive-communicative cuing context" (p. 53) be described in mechanistic terms—or do they require the assumption of an implicit and externally focused mentalization? And

is it possible to give a purely mechanistic description of the infant's part in those early mother-infant interactions that "are characterized by frequent *exchanges of a relatively rich and differentiated* (and ontogenetically quickly increasing) *repertoire of facial-vocal emotion displays expressing specific basic emotions* (including anger, joy, fear, sadness, disgust, and interest" (p. 53)? It is notable that Gergely and Unoka nowhere use the term "implicit mentalization" in their discussion of these processes; but does this mean that they think that the infant's part in these interactions is free from implicit and externally focused mentalization, and can be described in mechanistic terms alone—in that case they would also reject primary intersubjectivity in its nonmentalistic interpretation.

Can Primary Intersubjectivity Help Resolve Problems in the Theory of Mentalization?

We now turn to the four unresolved problems, or unclear points, in mentalization theory that were identified in an earlier section of the article. These four problems referred to (a) the role of implicit and explicit processes in the development of the infant's capacity for metacognition; (b) the role of implicit and explicit processes in the development of the infant's capacity for mentalizing its own affects; (c) the infant's ability to perceive emotional qualities in the other's behavior as a necessary condition for the affect mirroring process to take place; and (d) the question whether the distinction between externally and internally focused mentalization really represents a separate dimension of mentalization.

Primary Intersubjectivity and Implicit Metacognition

The mentalization literature contains no application of the implicit/explicit distinction to metacognition. Models of implicit metacognition have, however, been outlined by Proust (2007) and Brinck and Liljenfors (2013a, b). Further, the latter writers have suggested that primary intersubjectivity plays a role in the development of metacognition. An important aspect of their approach is that what makes a given process or action metacognitive is its operative *function*, and not its form (i.e., whether it is internal or external). That is, overt behaviors may have a metacognitive function if they serve to monitor or control one's cognitive processes, or if they make use of an adult as a sociocognitive resource. According to Brinck and Liljenfors, three features make intersubjectivity apt for initiating early metacognitive development: First, shared monitoring and control of cognition are integral to it; second, it enables learning and training of actions that realize monitoring and control functions; and third, feedback is immediate.

We picture reciprocal interaction as a monitoring-and-control game in which infants are motivated to participate by the urge for engaging with another subject. The fundamental role of monitoring and control of cognition for the proper functioning of protoconversation, is—we claim—initiating, maintaining, and achieving turns. These functions constitute the underlying mechanism that makes early intersubjectivity a productive platform for developing and learning sociocultural norms and metacognitive skills. In taking turns, infants begin learning how to manage cognitive states by influencing affect, motivation, and attention, with a direct impact on the interaction. Eventually, the infant becomes more skilled in taking turns and comes to share the responsibility for the interaction with the adult. The adult then will be of double epistemic use to the infant: On the one hand, as a teacher that comments on and corrects the infant's efforts, and on the other, as the infant's sociocognitive resource in its own right during the interaction . . . The way in which this kind of reciprocal interaction facilitates infants' metacognition is rooted in emotion and attention. Besides verbal expressions and vocalization, gesture, eye gaze

and bodily and facial expression of emotion are strong communicative signals. They enable joint monitoring and control of the interaction between infant and adult and make it possible for them to communicate about what they are doing. (Brinck & Liljenfors, 2013a, p. 94)

Primary Intersubjectivity and Metaemotion

As to the second problem, the same kind of processes as were discussed in the case of metacognition may be assumed to be involved also in the infant's developing awareness and regulation of its own affective processes. Let us use the term "metaemotion," or "metaemotional processing" (Lundh, Johnsson, Sundqvist, & Olsson, 2002) to refer to processes that serve to monitor and regulate one's emotions; applying Brinck and Liljenfors' (2013a) reasoning would then mean that processes will have a metaemotional function if they serve to monitor or regulate one's emotional processes, or if they make use of an adult for this purpose. This would also suggest that the development of metaemotion starts in the context of primary intersubjectivity, in the form of an implicit mentalizing of affect in the dyadic interaction with the caregiver.

Again, it may be assumed that primary intersubjectivity exists even before this development of implicit mentalization, at least in the form of emotional contagion, or emotional resonance (i.e., a primitive sharing of affect), presumably also in the form of reciprocity (Rochat & Passos-Ferreira, 2008), and serves as a facilitating context for this development. With regard to the self, it may be assumed that the infant has an inborn capacity to experience, or "feel," its own bodily state and changes in this state at an implicit, nonconceptual level (cf. Gendlin, 1962). Combining this with Gergely and Unoka's (2008) terminology and description of the process, this would suggest that what they refer to as "the primary and procedural basic emotion programs of the constitutional self (which are prewired, initially nonconscious automatisms)" (p. 63) is phenomenologically linked to an "implicit experiencing" of the bodily self. It would then be *the infant's nonconceptual "implicit experiencing" of its bodily state and changes in that state* that becomes associated with second-order representations as part of the affect mirroring process, and thereby leads to the development of an explicit mentalization in the form of an increasing emotional awareness. In other words, to speak of infants as "nonconscious automatisms" seems to present a flawed picture, as one necessary aspect of emotion is a felt experience (e.g., Izard, 2009; Prinz, 2004). That infants fail to show second-order *awareness of* their affect states does not mean that they do not experience them.

It may also, however, be argued that this kind of implicit experiencing of one's bodily state continues to be an important aspect of mentalizing throughout life.

Although it may sound like a contradiction to speak of the implicit experiencing of one's *bodily state* as a form of *mentalizing* (at least to a Cartesian thinker), it may be noted that such an implicit nonconceptual experiencing of one's body is trained both as part of mindfulness meditation (and mindfulness-based psychotherapies), and as part of focusing-oriented psychotherapy (Gendlin, 1996) where the person is trained to focus on an implicitly "felt sense" in the body and to search for words that "fit" this bodily felt sense. The latter might actually be seen as an example of an interaction between implicit and explicit mentalization. Maybe such an interaction between implicit and explicit mentalizing of affect is also characteristic of higher levels of mentalization, as seen in what Jurist (2005) has referred to as "mentalized affectivity."

Primary Intersubjectivity and the Mentalization of Others' Affect

As to the third problem which was formulated above, the affect mirroring process as described by Fonagy et al. (2012) seems to require an ability of the child to "compare" its own affective state with that as expressed by the caregiver. Otherwise the child would not be able to experience whether the caregiver's emotional expression is congruent or incongruent, and marked or not. Such a "comparison" need not, however, be an explicit conscious comparison of two experiences, but may take place in the form of an implicit experiencing of the "resonance" between the felt qualities of the child's bodily states and the perceived qualities of the caregiver's emotional expressions. Again, there is no need for the latter to be consciously apprehended; the entire process may probably start at the level of primary intersubjectivity, with an appreciation of both one's own affective bodily state and the caregiver's emotional expressions that may well qualify as implicit mentalization.

Primary intersubjectivity here may be said to take form primarily of interaffectivity, or sharing of affects (Brinck, 2008). One might conclude that primary intersubjectivity simply is the way our species is equipped. Hutto (2008) captures this thought in the statement that young infants are sensitive to

> a special class of natural signs—the expressions of intentional and affective attitudes as revealed in another's gaze, gesture, facial comportment . . . participants embody intentional attitudes that are directed toward the intentional attitudes of others . . . [i]ntentionality and affect are therefore expressed by the way organisms carry themselves (p. 117).

Again, it is worth noting that this account also corresponds to what Fonagy et al. (2012) describe as externally focused mentalizing, thus further strengthening the view that early mentalizing (external, implicit) overlaps with capacities involved in primary intersubjectivity. Also, the fact that contingency is part of affective communication does not entail that affective communication is reducible to contingency. And vice versa, to play down the role for the contingency detection mechanism is not to deny that it is important in intersubjective engagement. In fact, contingency processes constitute the foundation of social communication, because they reduce uncertainty about what is likely to happen next (Beebe et al., 2011). Interactive contingency involves consistently occurring, moment-to-moment adjustments that each individual makes to changes in partner behavior. Beebe et al. (2011, p. 176) define contingency as "the predictability of each partner's behavior from that of the other . . . translated into the metaphor of expectancies of 'how I affect you,' and 'how you affect me.'" To conclude, the occurrence of social contingency is fully compatible with claims for primary intersubjectivity.

Primary Intersubjectivity and Externally Focused Mentalization

Finally, concerning the fourth problem, it is relevant to ask if the differentiation between externally and internally focused mentalization has the same status as the other three polarities. The self/other polarity and the cognitive/affective polarity both apply to "the contents that are being mentalized," and are easy to crosstabulate—both affective and cognitive processes take place in both self and other. Although the implicit/explicit polarity differs from the self/other and cognitive/affective polarities by referring to "the nature of the mentalizing" rather than to "the contents that are mentalized," it is also easily cross-tabulated with the two other polarities. But the differentiation between externally and internally focused mentalization is not as easy to fit into this scheme. First, the notion of "externally focused mentalization" is applicable only to mental processes that have an external expression, like emotions, and is less applicable to cognitive processes, which

less often have such external expressions. Second, the notion of "externally focused mentalization" is not independent of the self/other polarity either, as it is primarily relevant to the mentalizing of *others'* emotions—although there may be situations where people infer their own emotions from an observation of their external expressions, rather than identifying them on the basis of internal experiences, this is hardly typical. Does a polarity which is relevant primarily to the mentalizing of emotion in others really deserve the status of an independent dimension of mentalization?

Again, it may be noted that "externally focused mentalization," at least when it refers to others' emotions, shows a clear overlap with primary intersubjectivity in its nonmentalistic interpretation. Could it be that an integration of primary intersubjectivity into the theory of mentalization would be a better way of integrating externally focused mentalization into the theory?

Conclusions

There is a certain tension in how the term mentalization is used in the writings of Fonagy and colleagues. On the one hand, mentalization is defined as a *developmental achievement* characterized by the ability to interpret others' behavior in terms of mental states and to understand one's own mental states, as well as to differentiate one's own and others' mental states and differentiate mental states from external reality. From this perspective, it would be absurd to attribute mentalizing capacities to infants. On the other hand, as described in the present article, the literature on mentalization also contains the ambition to understand the development of mentalization by analyzing it into a number of dimensions or polarities: cognitive versus affective, self-oriented versus other-oriented, implicit/automatic versus explicit/controlled, and internally focused versus externally focused, and to delineate its development as a many-faceted process with several different phases. From this perspective, mentalization serves as a label for a large variety of capabilities that develop from early childhood into adulthood. This project also allows a description of different individuals in terms of their profiles of capabilities and limitations with regard to mentalizing (cf. Luyten, Fonagy, Lowyck, & Vermote, 2012). Primitive forms of mentalizing can be attributed even to young infants, and it is an empirical question to determine *when* the different aspects of mentalizing typically develop, and the various ways in which these developments can go wrong with potential psychopathological implications.

We suggest that the concept of intersubjectivity is similar to that of mentalization in this sense: On the one hand it represents a developmental achievement, and on the other hand it represents a multifaceted phenomenon that may be analyzed into aspects and constituents, the development of which may be studied empirically. Its first manifestation is primary intersubjectivity, which was defined by Trevarthen (1979) as an innate drive to partake in dialogue with the caretaker, and which can be seen even before the development of implicit forms of mentalization. We have also argued that primary intersubjectivity, as defined by Trevarthen, Gallagher, and others, shows a partial overlap with implicit and externally oriented mentalization. Although Fonagy et al. (2002) and Gergely and Unoka (2008) have rejected the notion of primary intersubjectivity, we think this has been done on the assumption that it attributes a capacity for explicit and internally focused mentalization already to the small infant. We also think that this rejection of primary intersubjectivity was formulated before Fonagy et al. (2012) developed their dimensional conception of mentalization, where

implicit mentalizing is described as even more fundamental than explicit mentalizing not only developmentally but also for the adult's mental functioning. It is interesting to compare the latter with Gallagher and Hutto's (2008) statement that primary intersubjectivity remains primary in an experiential sense throughout adulthood. As we have argued in this article, the notion of primary intersubjectivity may possibly help to solve some unclear points and unresolved problems in mentalization theory. We therefore think that the notion of primary intersubjectivity may be worth a second look to consider whether it can be fruitfully integrated into the developing conceptual framework of mentalization theory.

References

Agnetta, B., & Rochat, P. (2004). Imitative games by 9-, 14-, and 18-month-old infants. *Infancy, 6,* 1–36. doi:10.1207/s15327078in0601_1

Allen, J. G. (2006). Mentalizing in practice. In J. G. Allen & P. Fonagy (Eds.), *Handbook of mentalization-based treatment* (pp. 3–30). West Sussex, UK: Jon Wiley Ltd. doi:10.1002/9780470712986.ch1

Allen, J. G., & Fonagy, P. (Eds.). (2006). *Handbook of mentalization-based treatment*. West Sussex, UK: Jon Wiley Ltd. doi:10.1002/9780470712986

Apperly, I., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116,* 953–970. doi:10.1037/a0016923

Balcomb, F. K., & Gerken, L. (2006). Does implicit metacognition provide a tool for self- guided learning in preschool children? In R. Sun (Ed.), *Cog Sci 2006: Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 1003–1008). Hillsdale, NJ: Erlbaum.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72,* 708–717. doi:10.1111/1467-8624.00310

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: Bradford, MIT-Press.

Baron-Cohen, S., Golan, O., Chakrabarti, B., & Belmonte, M. (2008). Social cognition and autism spectrum conditions. In C. Sharp, P. Fonagy, & I. Goodyer (Eds.), *Social cognition and developmental psychopathology* (pp. 29–56). Oxford, UK: Oxford University Press. doi:10.1093/med/9780198569183.003.0002

Bateman, A., & Fonagy, P. (2004). *Psychotherapy of borderline personality disorder: Mentalization-based treatment*. Oxford, UK: Oxford University Press.

Beebe, B., Steele, M., Jaffe, J., Buck, K., Chen, H., Cohen, P., . . . Feldstein, S. (2011). Maternal anxiety symptoms and mother-infant self- and interactive contingency. *Infant Mental Health Journal, 32,* 174–206. doi:10.1002/imhj.20274

Bretherton, I. (2005). In pursuit of the internal working model construct and its relevance to attachment relationships. In K. E. Grossman, K. Grossman, & E. Waters (Eds.), *Attachment from infancy to adulthood. The major longitudinal studies* (pp. 13–47). New York, NY: Guilford Press.

Brinck, I. (2008). The role of intersubjectivity in the development of intentional communication. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *The shared mind: Perspectives on intersubjectivity* (pp. 115–140). Amsterdam, The Netherlands: John Benjamins Publishing Co. doi:10.1075/celcr.12.08bri

Brinck, I., & Liljenfors, R. (2013a). The developmental origin of metacognition. *Infant and Child Development, 22,* 85–101. doi:10.1002/icd.1749

Brinck, I., & Liljenfors, R. (2013b). Reply to commentaries. *Infant and Child Development, 22,* 111–117. doi:10.1002/icd.1788

Cortina, M., & Liotti, G. (2010). Attachment is about safety and protection, intersubjectivity is about sharing and social understanding. *Psychoanalytic Psychology, 27,* 410–441. doi:10.1037/a0019510

de Bruin, L., Strijbos, D., & Slors, M. (2011). Early social cognition: Alternatives to implicit

mindreading. *Review of Philosophy and Psychology, 2,* 499–517. doi:10.1007/s13164-011-0072-1

Flavell, J. H. (1979). Metacognition and cognitive monitoring. A new area of cognitive–developmental inquiry. *American Psychologist, 34,* 906–911. doi:10.1037/0003-066X.34.10.906

Fonagy, P., & Bateman, A. (Eds.). (2012). *Handbook of Mentalizing in mental health practice*. Arlington, VA: American Psychiatric Publishing, Inc.

Fonagy, P., Bateman, A., & Bateman, A. (2011). The widening scope of mentalizing: A discussion. *Psychology and Psychotherapy: Theory, Research and Practice, 84,* 98–110.

Fonagy, P., Bateman, A., & Luyten, P. (2012). Introduction and overview. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 3–41). Arlington, VA: American Psychiatric Publishing, Inc.

Fonagy, P., Gergely, G., Jurist, E. L., & Target, M. (2002). *Affect regulation, mentalization, and the development of the self*. London, UK: Karnac.

Fonagy, P., Gergely, G., & Target, M. (2007). The parent-infant dyad and the construction of the subjective self. *Journal of Child Psychology and Psychiatry, 48,* 288–328. doi:10.1111/j.1469-7610.2007.01727.x

Gallagher, S. (2004). Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, & Psychology, 11,* 199–217. doi:10.1353/ppp.2004.0063

Gallagher, S., & Hutto, D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen (Eds.), *The shared mind: Perspectives on intersubjectivity* (pp. 17–38). Amsterdam, The Netherlands: John Benjamins. doi:10.1075/celcr.12.04gal

Gendlin, E. T. (1962). *Experiencing and the creation of meaning. A philosophical and psychological approach to the subjective*. New York, NY: Free Press.

Gendlin, E. T. (1996). *Focusing-oriented psychotherapy. A manual of the experiential method*. New York, NY: Guilford Press.

Gergely, G., & Unoka, Z. (2008). Attachment and mentalization in humans. The development of the affective self. In E. J. Jurist, A. Slade, & S. Bergner (Eds.), *Mind to mind. Infant research, neuroscience, and psychoanalysis* (pp. 50–87). New York, NY: Other Press.

Gergely, G., & Watson, J. (1999). Early social-emotional development: Contingency perception and the social biofeedback model. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life* (pp. 101–137). Hillsdale, NJ: Erlbaum.

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

Izard, C. (2009). Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology, 60,* 1–25. doi:10.1146/annurev.psych.60.110707.163539

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. D. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development, 66,* i–149. doi:10.1111/1540-5834.00137

Jurist, E. L. (2005). Mentalized affectivity. *Psychoanalytic Psychology, 22,* 426–444. doi:10.1037/0736-9735.22.3.426

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology, 58,* 259–289. doi:10.1146/annurev.psych.58.110405.085654

Lundh, L. G., Johnsson, A., Sundqvist, K., & Olsson, H. (2002). Alexithymia, memory for emotion, emotional awareness, and perfectionism. *Emotion, 2,* 361–379. doi:10.1037/1528-3542.2.4.361

Luyten, P., Fonagy, P., Lowyck, B., & Vermote, R. (2012). Assessment of mentalizing. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 3–41). Arlington, VA: American Psychiatric Publishing, Inc.

Lyons-Ruth, K. (2007). The interface between attachment and intersubjectivity: Perspective from the longitudinal study of disorganized attachment. *Psychoanalytic Inquiry, 26,* 595–616. doi:10.1080/07351690701310656

Main, M. (1991). Metacognitive knowledge, metacognitive monitoring, and singular (coherent) vs.

multiple (incoherent) models of attachment: Findings and directions for future research. In C. M. Parkes, J. Stevenson-Hinde, & P. Marris (Eds.), *Attachment across the life cycle* (pp. 127–159). London, UK: Tavistock-Routledge.

Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting, 6,* 179–192. doi:10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.0.CO;2-R

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308,* 255–258. doi:10.1126/science.1107621

Perner, J., & Roessler, J. (2012). From infants' to childrens' appreciation of belief. *Trends in Cognitive Sciences, 16,* 519–525. doi:10.1016/j.tics.2012.08.004

Prinz, J. (2004). *Gut reactions—A perceptual theory of emotions*. Oxford, UK: Oxford University Press.

Proust, J. (2007). Metacognition and metarepresentation: Is a self-directed theory of mind a precondition for metacognition? *Synthese, 159,* 271–295. doi:10.1007/s11229-007-9208-3

Rochat, P., & Passos-Ferreira, C. (2008). From imitation to reciprocation and mutual recognition. In J. A. Pineda (Ed.), *Mirror neuron systems: The role of mirroring processes in social cognition* (pp. 191–212). New York, NY: Humana Press.

Rochat, P., Passos-Ferreira, C., & Salem, P. (2009). Three levels of intersubjectivity in early development. In A. Carassa, F. Morganti, & G. Riva (Eds.), *Enacting intersubjectivity. Paving the way for a dialogue between cognitive science, social cognition and neuroscience* (pp. 173–190). Como, Italy: Da Larioprint.

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research, 1079,* 86–97. doi:10.1016/j.brainres.2006.01.005

Sharp, C., & Fonagy, P. (2008). The parent's capacity to treat the child as a psychological agent: Constructs, measures and implications for developmental psychopathology. *Social Development, 17,* 737–754. doi:10.1111/j.1467-9507.2007.00457.x

Sroufe, L. A., & Waters, E. (1977). Attachment as an organizational construct. *Child Development, 48,* 1184–1199. doi:10.2307/1128475

Stern, D. N. (1985). *The interpersonal world of the infant*. New York, NY: Basic Books.

Stern, D. N. (2001). Face-to-face play: Its temporal structure as predictor of socioaffective development. Commentary. *Monographs of the Society for Research in Child Development, 66,* 144–149. doi:10.1111/1540-5834.00147

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18,* 580–586. doi:10.1111/j.1467-9280.2007.01943.x

Trevarthen, C. (1979). Communication and cooperation in early infancy. In M. Bullowa (Ed.), *Before speech: The beginnings of human communication* (pp. 321–347). Cambridge, MA: Cambridge University Press.

Trevarthen, C. (1992). An infant's motive for speaking and thinking in the culture. In A. Heen Wold (Ed.), *The dialogical alternative, towards a theory of language and mind* (pp. 19–44). Oslo, Norway: Scandinavian University Press.

Trevarthen, C., & Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding, and acts of meaning in the first year. In A. Lock (Ed.), *Action, gesture, and symbol: The emergence of language* (pp. 183–229). London, UK: Academic Press.

Wellman, H. (1990). *The child's theory of mind*. Cambridge, MA: MIT-Press/A Bradford Book.