Affect Regulation, Mentalization, and the Development of the Self

Peter Fonagy
György Gergely
Elliot L. Jurist
Mary Target

OTHER

Other Press
New York

actual emotional experience adequately, but also to the construction of an experience of self around this false internalization (Winnicott 1965).

While the conscious experience of "putting on an act" may be a fairly general experience, particularly in adolescence (Harter, Marold, Whitesell, and Cobbs 1996), in the current context we are referring to the highly distressing experience of severely personality-disordered children who experience a sense of alienation from their core self (Bleiberg 1984, 1994). A strategy adopted by many such children somewhat later in development is to attempt to externalize this false part of their self-representation and manipulate the behaviors of others around them so that these match the incongruent self-representation. We would argue that this model explains the strangely coercive behavior with the caregiver of preschool children whose attachment at the stage of infancy was classified as disorganized (Cassidy, Marvin, and The MacArthur Working Group on Attachment 1989; Crittenden 1992; Main and Cassidy 1988). These children—and adults—become quite skilled and sensitive in manipulating others to get them to behave consistently with their internal representations. These are not their representations of self–other relationships, which we all try to actualize. In a far more desperate way these children and adults try to provoke behavior consistent with a part of the self-representation experienced as "alien," which they feel forced to externalize in order to achieve a more coherent perception of the residual self (Fonagy and Target 1995).

ATTACHMENT SECURITY IN THE CHILD AND REFLECTIVE FUNCTION

There is general agreement that the "harmoniousness of the mother–child relationship contributes to the emergence of symbolic thought" (Bretherton, Bates, Benigni, Camaioni, and Volterra 1979, p. 224), and the idea has a long and distinguished history (Mahler, Pine, and Bergman 1975; Vygotsky 1978; Werner and Kaplan 1963). Bowlby (1969) recognized the significance of the developmental step entailed in the emergence of "the child's capacity both to conceive of his mother as having her own goals and interests separate from his own and to take them into account" (1969, p. 368). Moss, Parent, and Gosselin (1995) reported that attachment security with mother was a good concurrent predictor of metacognitive capacity in the child in the domains of memory, comprehension, and communication. The Separation Anxiety Test, a projective test of attachment security, has been shown to be a good predictor of belief–desire reasoning capacity in 3½- to 6-year-old children when age, verbal mental age, and social maturity were all controlled for (Fonagy, Redfern, and Charman 1997).

In a prospective study of the relationship of attachment security to mother (12 months) and to father (18 months) and performance on three tests of theory of mind at 5½ years (Fonagy 1997), 92 of 96 children tested in the Strange Situation at 12 and 18 months were seen. Of those classified as secure at 12 months with mother, 82% passed the belief–desire reasoning task, in which the child is challenged to predict what a character would feel, based on his or her knowledge of the character's belief. (If Ellie thinks the can has coke in it, and likes coke, will she want the drink—even though it is really milk?) In contrast, 46% of those who had been classified as insecure failed. Infant–father attachment (at 18 months) also predicted the child's performance, with 77% of infants classified as secure passing the test, compared to 55% of children classified as insecure. There was some indication of an additive relationship, in that 87% of children with two secure relationships passed the belief–desire task, and 63% of those with only one secure relationship and only 50% of those insecure with both did so. A similar but somewhat weaker pattern could be observed with the second-order false-belief task, which requires the child to reason on the basis of what one character knows about another character's wrong belief. Of those secure with both parents, 36% passed, compared with 23% who were secure with one and 9% who were insecure with both.

In a somewhat smaller but nevertheless careful longitudinal study of mother–infant dyads, Meins and colleagues (Meins, Fernyhough, Russel, and Clark Carter 1998) reported that 83%

of children who were securely attached in infancy passed a false-belief task at the age of 4, in comparison with 33% of insecurely attached peers. At age 5, 85% of securely attached children and 50% of insecurely attached ones passed a task requiring an understanding of information access. Although the study was not able to replicate our results on the false-belief and emotion task (probably because of its small sample), the general trend of the findings confirms that security of attachment is significantly linked to symbolic abilities in general and to precocious mentalizing in particular.

Both trivial and substantive explanations could be offered to account for these findings. They would be trivial if the association of secure attachment and false-belief understanding were due to an as yet unknown and unmeasured third factor, such as temperament. More plausibly, it could be argued that the facilitative effect of secure attachment is due to a more relaxed, task-oriented attitude, to a general facility to engage in a cognitively demanding task, or to an ability to relate to an adult experimenter in a playful, exploratory way. All these explanations reflect child performance rather than competence. This suggestion could be tested using a false-belief task where implicit and explicit knowledge of false belief is separately assessed (Clements and Perner 1994). If attachment security relates to performance, then securely attached children would be expected to do better only on the explicit (verbal/pointing) task. Implicit, procedural false-belief reasoning would be expected to be facilitated by secure attachment only if this was associated with superior reflective capacity. This remains to be studied. In what follows, we shall cautiously assume that the relationship between false-belief reasoning and security of attachment is nontrivial.

There are two alternative explanations for this relationship: (*a*) One class of models would suggest that security of attachment in infancy predisposes children to benefit from certain social processes that may be directly involved in the development of reflective abilities and social understanding. (*b*) The second class of models would suggest that security of attachment is an indicator of that quality of infant-caregiver relationship which generates psychological understanding. In this second model the social processes that accelerate the mentalizing quality of self-organization are the very same as those that ensure security of attachment.

Mediational models would require that two conditions be satisfied: (*a*) compelling evidence that a specific set of social processes is involved in this aspect of the development of self-organization, and (*b*) that such social processes are enhanced in securely attached individuals. At least three candidates meet these criteria.

1. The first is *pretense.* There is evidence that children in their third year who engage more readily in cooperative interaction (Dunn et al. 1991), and specifically in joint pretend play (Astington and Jenkins 1995; Taylor, Gerow, and Carlson 1993; Youngblade and Dunn 1995), show superior mind-reading and emotion understanding performance. A separate body of observations from longitudinal studies of attachment indicates that preschool children who were securely attached to their mother in infancy manifest stronger engagement in fantasy play than do avoidant children, whose level of engagement is low and whose pretend play is impoverished (Rosenberg, cited in Carlson and Sroufe 1995; Main, Kaplan, and Cassidy 1985). Children rated as securely attached to their mothers during infancy have been reported to engage in more frequent and sophisticated solo pretense (Belsky, Garduque, and Hrncir 1984; Bretherton et al. 1979; Matas, Arend, and Sroufe 1978). Slade (1987) found that maternal involvement in 3-year-olds' play served a facilitating function only for securely attached children. Meins and colleagues (Meins et al. 1998) demonstrated that children who were classified as securely attached in infancy were better able to incorporate the pretense suggestions of an experimenter in their play at 31 months.

It is probable that joint pretend play or playfulness fosters the understanding of mental states. Deliberate role-taking is seen as being integral to the "off-line simulation" model of the performance of mind-reading tasks (Currie 1995; Goldman 1989). Within other models, pretend play is an early manifestation of the theory of mind mechanisms (Leslie 1987). The question of why

3-year-olds can understand that someone is entertaining a pretend representation but not a false belief (Harris and Kavanaugh 1993; Harris, Kavanaugh, and Meredith 1994), a pretend/real distinction but not an appearance/reality one (Flavell, Flavell, and Green 1987) is an important puzzle. In the case of pretend, the representations, while they are different from reality, are shared by those engaged in the pretend game. As Astington put it: "they are *inter*mental, not *intra*mental" (1996, p. 193). The sharing of representations that are different from reality may help in understanding situations where representations are not only different from reality but are not shared in a social pretend domain. In joint pretend play or playfulness the adult adopts the child's mental stance and re-presents it to the child in relation to a third object, which is symbolically held in mind by both. Pretending requires a mental stance involving the symbolic transformation of reality in the presence of, and with a view to, the mind of the other. The adult, or perhaps the older sibling, holds the frame of external reality while accurately representing the child's mental state. The scaffolding provided by the child's playmate in pretend play (Vygotsky 1967) not only promotes earlier success but is the mechanism whereby the development of reflection comes about. Lillard (1993) argued that symbolic play may offer a "zone of proximal development" for the skills that subserve mind-reading ability. Children with a secure attachment history may be more likely to engage in an activity that presumes a degree of trust, insofar as the child relies on the other's version or perception of reality.

2. The second is *talking*. There is evidence that conversations about feelings and the reasons behind people's actions are linked to the relatively early achievement of reflective function (Brown, Donelan-McCall, and Dunn 1996; Dunn and Brown 1993). Mothers who spontaneously explained their emotions to 3½-year-olds during laboratory simulation were shown to have children with enhanced emotion understanding over the subsequent 15-month period (Denham et al. 1994). Conversational opportunities concerning mental states appear to improve children's mentalizing performance in experimental studies (Appleton and Reddy 1996).

Strage and Main (cited in Carlson and Sroufe 1995) reported that discourse patterns between mothers and children at age 6 could be predicted from early attachment classification. Secure dyads were more fluent and able to discuss a wider range of topics than those classified as avoidant, who showed little topic elaboration. Patterns of mother–child interaction characteristic of secure dyads—shared play, comforting, or joking—also define the contexts within which the mother's explanations of mental states are found to be particularly facilitative of reflective function (Dunn 1996). Secure children find it easier to deal with emotional issues in an open and free way (Bretherton, Ridgeway, and Cassidy 1990; Cassidy 1988). Mother–child verbal interactions associated with the early acquisition of reflective capacity predominantly concern emotionally charged matters (Dunn 1996). Secure attachment may then engender patterns of verbal interaction between child and caregiver, which in turn support thinking about feelings and intentions.

P. K. Smith (1996) forcefully advanced the central role of language in the acquisition of mentalizing capacity. Using primate evidence, he suggested that the availability of symbolic codes (words) for mental states was crucial for the developing individual to acquire mind-reading abilities, and the explicit use of such codes by caregivers was therefore likely to be important. Even more pertinent in this context is Harris's (1996) proposal that the experience of engaging in conversations *per se* alerts children to the fact that people are receivers and providers of information, irrespective of whether that conversation involves reference to mental states (knowing, thinking, desiring, etc.). The structure of information-bearing conversations (e.g., being told about a past event that one has not witnessed, challenging of information in dissent and denial, or filling in information gaps in questions or when information exchange misfires and repair is needed) strongly implies that partners in a conversation differ in what they know and believe about a shared topic. Effective conversation requires that gaps in shared knowledge and belief are acknowledged and addressed. The measurement of attachment in adults (Main and Goldwyn 1994) strongly endorses the suggestion that secure attachment is associated with greater sensitivity to the rules of conversation as defined by Grice (1975).

3. The third potential mediator is *peer group interaction.* We have already noted that the possibility of interaction with siblings is likely to enhance theory-of-mind performance (Jenkins and Astington 1996; Perner et al. 1994; Ruffman et al. 1998). Importantly, the child's use of mental-state terms with siblings or friends is a better predictor of performance on false-belief tasks than is mother–child conversation (Brown et al. 1996). Likewise, Lewis and colleagues (Lewis, Freeman, Kyriakidou, Maridaki-Kassotaki, and Berridge 1996) demonstrated that false-belief understanding was related to the amount of time that children spent with older siblings, older friends, and older kin, but not with younger persons. There is an independent body of evidence that supports a strong link between secure attachment in infancy and ratings of peer competence (Elicker, Englund, and Sroufe 1992). Children with secure attachment histories are consistently observed and rated more socially oriented, capable of reciprocity, popular, and empathic than those with insecure histories (Lieberman 1977; Pancake 1985; Park and Waters 1989; Sroufe 1983).

Both simulation theory and theory–theory explanations of the development of mind-reading offer good explanations of the facilitative effect of more intense peer-group interaction (Ruffman et al. 1998). Peer-group interaction should increase the opportunities that children have for simulation, imagining what they would see, think, feel, and so on if they were in another person's situation. Equally, interaction with peers or older siblings could be seen from a theory–theory perspective as a rich source of ideas about how the mind works. An alternative view may be that enculturation is itself the source of the child's mental-state concepts (Astington 1996). Bruner (1983) proposed that parents' tendency to treat the infant's spontaneous gestures *as if* they were intentional communications leads to infants seeing themselves as having intentions and starting to communicate intentionally. The social world—in the first instance, the parent—fosters the child's sense of his mental self through complex linguistic and interactional processes, behaving toward the infant in a way that leads him eventually to share the assumption that his own behavior and (by simulation or the observation of similar interactions between the caregiver and others) that of others may be best understood in terms of mental states. Through participation in activities of their culture they come to share their culture's way of regarding others' and their own actions. If children's entry into the folk psychology is viewed as a process of "apprenticeship" in which senior peers and caregivers are seen as encouraging the child's adoption of mentalizing concepts (Astington 1996; Lewis et al. 1996), then secure attachment may be considered as a kind of catalyst to this learning process. The greater readiness with which secure children are willing to explore and engage with the social world could then account for their relative competence in mentalizing abilities.

There is nothing mutually exclusive about these three mediational models. Pretense often involves the use of mental-state language. Interaction with peers often involves both language and pretense. In general, these mediational models suggest that propensity for social engagement across a number of contexts enhances the development of social understanding and that such social engagement is more readily accessible in the families of securely attached young children. However, there is a major problem with such a singular model. Evidence from Dunn's work suggests that these different contexts correlate poorly with one another (Dunn 1996). For example, observational data indicate that individual differences found in pretend play, management of conflict, and discourse about mental states are not correlated among social situations (mothers, siblings, close friend), although each correlates with socio-cognitive assessments (Brown et al. 1996; Slomkowski and Dunn 1992; Youngblade and Dunn 1995). The fact that children's behavior correlates poorly across social partners and situations, although each of those situations relates to test performance, could suggest that there are several independent, simultaneously operating pathways between attachment and social situations.

Alternatively, it is possible that the variables that *prima facie* may be considered to be mediating the attachment/mind-reading relationship are not on the causal path at all, that their correlation with the rate of acquisition of mind-reading is spuri-

ous, and that the causal sequence of the relationship does not involve these social experiences but is directly related to the child's attachment status. The early experience with the caregivers in the first year of life may create the bedrock of theory-of-mind competence, ensuring the child's move from a teleological to a mentalizing model of behavior. What evidence do we have to support such a contention? First, it is important to note that in the London data, a mother's attachment classification before the birth of the child was a powerful predictor of the child's theory-of-mind competence at 5 years; 75% of children of secure, autonomous mothers passed the cognitive-emotion task, whereas only 16% of children of preoccupied mothers and 25% of those of unresolved mothers did so (Fonagy 1997). Although on the face of it this could be accounted for by the models already discussed, we believe that there is now evidence that the caregiver brings something to the parent–child relationship, evident even before the birth of the child, which may be critical in the child's establishment of both secure attachment and mind-reading.

What might this be? It is well established that, in infancy, mothers of securely attached children are more sensitive to their children's needs (Ainsworth, Bell, and Stayton 1971; Isabella 1993). We have already touched on the fact that the caregiver's capacity to envision the mental states of her own parents is predictive of the infant's security of attachment to each of his caregivers (Fonagy, Steele, Moran, et al. 1991). In the London project, mothers who were more likely to invoke mental states in their accounts of their childhood attachment experiences had children with superior mind-reading abilities (controlling for verbal fluency in the child). Ratings on the RF scale were found to predict the child's performance on cognitive-emotion tasks also for fathers. Even more important, a path analysis revealed that a mother's capacity to reflect on her own childhood in the AAI shared that portion of the variance with the child's theory-of-mind performance that was predicted by the quality of mother–infant attachment. In a more complex path analysis, we found that the mother's mentalizing ability had a direct as well as indirect relationship with the child's theory of mind. Thus, the child's attachment security was not the only predictor. The

mother's capacity to envision the child as a mental entity also seemed to be important.

Such data suggest that common mechanisms underpin attachment organization in the caregiver and the infant and the precocious emergence of mentalizing. It should be remembered that no unequivocal causal paths were identified among mediational models. The relative importance of various potential mediational mechanisms for the attachment–theory-of-mind relationship varies according to context, but intergenerational data may be consistent with at least two of the models (pretense, language). Further experimental research that manipulates parental behavior and explores attachment and theory-of-mind task performance (van IJzendoorn, Juffer, and Duyvesteyn 1995) will be necessary to determine whether specific behaviors that engender secure attachment simultaneously enhance mentalizing. In order for such a study to be feasible, we need a model of how attachment may directly relate to theory-of-mind performance. Next, therefore, we offer one model of how such a mechanism may operate.

Our stipulation is based on the assumption that the acquisition of the theory of mind is part of an intersubjective process between the infant and caregiver (see Gopnik 1993, for a highly elegant elaboration of such a model). In our view, the caregiver facilitates the creation of mentalizing models through complex linguistic and quasi-linguistic processes, primarily by behaving toward the child in such a way that the child is eventually led to postulate that his own behavior may be best understood if he assumes that he has ideas and beliefs, feelings and wishes, that determine his actions, and the reactions of others to him can then be generalized to other similar beings. The caregiver approaches the crying infant with a question in her mind: "Do you want your nappy changed?" "Do you need a cuddle?" The sensitive caregiver is unlikely to address the situation without having the person in mind, so is unlikely to say to herself, "Are you wet around your bottom?" or "Have you been standing alone too long?" The sensitive caregiver can bridge the focus on physical reality and internally directed attention sufficiently for the child to identify contingencies between internal and external experi

ence. Ultimately, the child arrives at the conclusion that the caregiver's reaction to him may be understood as rational, given the assumption of an internal state of belief or desire within himself. Unconsciously and pervasively, the caregiver ascribes a mental state to the child with her behavior, treating the child as a mental agent. This is ultimately perceived by the child and used in the elaboration of teleological models and permits the development of a core sense of selfhood organized along mentalistic lines. We assume that this, by and large, is a mundane process, in the sense of happening every day throughout early life, and that it is a process that is preconscious to both infant and caregiver, inaccessible to reflection or modification. Caregivers, however, execute this most natural of human functions in different ways. Some may be particularly alert to the earliest indications of intentionality, while others may need stronger indications before they can perceive the child's mental state and modify their behavior accordingly.

The child's development and perception of mental states in himself and others thus depends on his observation of the mental world of his caregiver. He develops a concept and is thus able to perceive mental states, to the extent that his caregiver's behavior implies such states. He does this when the caregiver is in a shared pretend mode of "playing" with the child (hence the association between pretend and early mentalization), and many more mundane interactions (such as conversations and peer interaction) will also involve such shared mentation. This is what makes mental-state concepts such as thinking inherently intersubjective; shared experience is part of the very logic of mental-state concepts.

We believe that the caregiver's capacity to observe the moment-to-moment changes in the child's mental state is critical in the development of mentalizing capacity. The caregiver's perception of the child as an intentional being lies at the root of sensitive caregiving, which attachment theorists view as the cornerstone of secure attachment (Ainsworth et al. 1978; Bates, Maslin, and Frankel 1985; Belsky and Isabella 1988; Egeland and Farber 1984; Grossmann, Grossmann, Spangler, Suess, and Unzner 1985; Isabella 1993; Isabella and Belsky 1991).

Secure attachment, in its turn, provides the psychosocial basis for acquiring an understanding of mind. The secure infant feels safe in making attributions of mental states to account for the behavior of the caregiver. In contrast the avoidant child shuns to some degree the mental state of the other, while the resistant child focuses on its own state of distress, to the exclusion of close intersubjective exchanges. Disorganized infants may represent a special category: hypervigilant of the caregiver's behavior, they use all cues available for prediction; they may be acutely sensitized to intentional states and thus may be more ready to construct a mentalized account of the caregiver's behavior. We would argue (see below) that in such children mentalization may be evident, but it does not have the central role in self-organization that characterizes securely attached children. We believe that what is most important for the development of mentalizing self-organization is the exploration of the mental state of the sensitive caregiver, which enables the child to find in the caregiver's mind (that is, in the hypothetical representation of her mind that he constructs to explain her behavior toward him) an image of himself as motivated by beliefs, feelings, and intentions. In contrast, what the disorganized child is scanning for so intently is not the representation of his own mental states in the mind of the other, but the mental states of that other that threaten to undermine his own self. They can constitute within the child's self-representation an alien presence that is so unbearable that his attachment behavior becomes organized around reexternalizing these parts of the self onto attachment figures, rather than around the internalization of a capacity for containment of affects and other intentional states. There is considerable evidence to support the view that secure attachment enhances the development of the self, inner security, feeling of self-worth, self-reliance, and personal power of the emerging self as well as the development of autonomy (Bates et al. 1985; Gove 1983; Londerville and Main 1981; Matas et al. 1978). Disorganized infants, even if they acquire the skill of mind-reading, fail to integrate this with their self-organization.

There may be a number of linked reasons for this: (*a*) The child needs to use disproportionate resources to understand

the parent's behavior, at the expense of reflecting on self-states. (*b*) The caregiver of the disorganized infant is less likely to be reliably contingent in responding to the infant's self-state and, further, to show systematic biases in her perception and reflection of the child's state. (*c*) The mental state of the caregiver of the disorganized infant evokes intense anxiety either through frightening behavior suggesting malevolence toward the child or through behavior suggesting fear, including inexplicable fear of the child himself.

These factors may combine to make disorganized infants become keen readers of the caregiver's mind under certain circumstances but, we suggest, poor readers of their own mental states. Thus, in terms of the rival models of theory-of-mind development, such children may acquire a theory–theory of mind but be unable to use simulation of mentalizing with the same confidence as can children whose attachment (albeit insecure) is coherent and organized. The alternative models may be more usefully thought of as alternative routes to mentalization, the first (theory–theory) accessible to all, the second (simulation) more readily available to children whose early attachment relationships made such a strategy more attractive and desirable.

In chapter 6, we attempt to describe the normal development of reflective function in the child aged 2 to 5 years. We suggest that there is a transition from a dual mode of experience to mentalization. Primarily from a clinical perspective, we advance a number of propositions concerning the development of the psychological part of the self. These are:

1. In early childhood, reflective function is characterized by two modes of relating internal experiences to the external situation: (*a*) In a serious frame of mind, the child expects the internal world in himself and others to correspond to external reality, and subjective experience will often be distorted to match information coming from outside—*psychic equivalence mode* (e.g., Gopnik and Astington 1988; Perner, Leekam, and Wimmer 1987). (*b*) While involved in play, the child knows that internal experience may not reflect external reality (e.g., Bartsch and Wellman 1989; Dias and Harris 1990), but then the internal state is thought to have no relationship to the outside world and to have no implications for it (*pretend mode*).

2. In normal development, the child integrates these two modes to arrive at the stage of *mentalization*—or *reflective mode*—in which mental states can be experienced as representations. Inner and outer reality can then be seen as linked, yet they are accepted as differing in important ways and no longer have to be either equated or dissociated from each other (Baron-Cohen 1995; Gopnik 1993).

3. We have hypothesized that mentalization normally comes about through the child's experience of his mental states being reflected on, prototypically through experience of secure play with a parent or older child, which facilitates integration of the pretend and psychic equivalence modes, through an interpersonal process that is perhaps an elaboration of the complex mirroring of the infant by the caregiver. In playfulness, the caregiver gives the child's ideas and feelings (when he is "only pretending") a link with reality by indicating the existence of an alternative perspective, which exists outside the child's mind. The parent or older child also shows that reality may be distorted by acting upon it in playful ways, and through this playfulness a pretend but real mental experience may be introduced.

4. In traumatized children, intense emotion and associated conflict can be thought of as having led to a partial failure of this integration, so that aspects of the pretend mode of functioning become part of a psychic equivalence manner of experiencing reality. This may be because where maltreatment or trauma has occurred within the family, the atmosphere tends to be incompatible with the caregiver "playing with" the most pressing aspects of the child's thoughts; these are often disturbing and unacceptable to the adult, just as they are to the child. The rigid and controlling behavior of the preschool child with a history of disorganized attachment is thus seen as arising out of a partial failure on the part of the child to move beyond the mode of psychic equivalence in relation to specific ideas or feelings, so

that he experiences them with the intensity that might be expected had they been current external events.

THE IMPLICATIONS OF REFLECTIVE FUNCTION FOR SELF-DEVELOPMENT

"Mind-reading" may not be an unequivocally positive experience. Judy Dunn's work, however, gives us an indication that at least the understanding of emotion at the age of 3½ predicts a positive perception of social relations, mature moral sensibility, and the understanding of complex emotions (Herrera and Dunn 1997). Stern (1985) pointed out that a sense of ownership of one's actions—whether derived from the experience of forming plans, proprioceptive feedback, or the objective consequences of physical actions on the environment—contributes significantly to the sense of self-agency. In our view, such agency also crucially depends on the quality and reliability of reflective function, as ownership of action is intimately tied to the mental state (belief or desire) that initiated it. It is impossible to conceive of self-agency as fully established by the actual actions of the child, as such a large proportion of these will fail to achieve their intended objective because of the child's immature physical and cognitive capacities. In fact, it could be argued that if the sense of self-agency were uniquely based on feedback from immature action systems, deficiency in this sphere would be universal. The recognition of the child's intentional stance by (older) others must, then, be critical in making the thought "real" for the child. Interpersonal interaction that permits the registration of perceptions, thoughts, and emotions as causes and consequences of action and the contemplation of these mental states without fear must constitute an important part of the foundation of self-agency. The earliest foundation is presumably the baby's sense that he brings about the caregiver's mirroring behavior (Gergely and Watson 1996). This idea is at the core of chapter 4.

Of course, the core of self-agency must lie in the body, where the infant's effort to be in control often succeeds after the earliest times. However, more complex actions, particularly those

that involve others in the child's world, frequently require the reflective caregiver to make sense of the young child's wishes and to express these, if the two-way connection between intentions and action is to be established. Hence, those who have experienced severe neglect or coercive, rigid, frightening, and, at an extreme, abusive parenting will frequently experience their sense of self-agency as massively curtailed and confined to the more firmly established bodily domain (see chapters 9 and 10 for a full exploration of this idea).

The model of the development of mentalizing capacity that we propose has considerable clinical implications. For example, in a study of attachment classification in patients with severe personality disorders, Fonagy et al. (1996) found that the AAI narratives of borderline-personality-disordered patients had lower reflective function, coupled with histories of severe trauma, which was apparently unresolved. The findings suggest that, given a sensitive attachment relationship, which provides the intersubjective basis for the development of mentalizing capacity, trauma (even if severe) is more likely to be resolved. Severe distortion of personality follows when abuse or neglect leads to a defensive inhibition of mentalization. Similarly, evidence is accumulating that among juvenile offenders, where histories of maltreatment are common, capacities for mentalization are severely restricted (Blair 1995; Levinson and Fonagy 2000). The evidence on disturbances of mentalization among clinical groups is elaborated more fully in later chapters (see chapters 10 and 11 in particular).

A DEVELOPMENTAL FRAMEWORK FOR ABNORMAL REFLECTIVE FUNCTION

It would be undoubtedly overly simplistic to make an absolute connection between developmental disorder and nonreflectiveness. There are variations across situations, or, perhaps more accurately, across relationships. The delinquent adolescent is well aware of the mental states of other gang members, and the borderline individual at times seems hypersensitive to the affec-

tive states of mental health professionals and family members. These "anomalies" may be clarified by more sophisticated developmental theory.

Our chosen framework is provided by "dynamic skills theory" (Fischer and Farrar 1987; Fischer, Kenny, and Pipp 1990), which depicts development as a person's elaboration of progressively more complex control systems (skills). Reflective function may be readily conceived of as one such control system, critical to the organization of the self. Within dynamic skills theory, reflective function would be seen as not simply a property of the person, but of the person and situation together, because all skills are composed of both the person's activities and the situations or context within which these occur. Particular tasks, specific events, other people, as well as culture are seen as part of the skill. Furthermore, the development of a skill is not seen as progression along any singular path, determined by maturation. Rather, reflective function, as a skill, evolves through varied pathways, molded by many dynamically interacting influences, such as the individual's emotions, social interaction, family relationships and environment, important social groups, the reactions of the wider social world, and so forth (Fischer, Knight, and Van Parys 1993).

Reflective function is a strand within the developmental web, one of the many distinct control systems that are neither strongly connected with each other nor coordinated or integrated (Fischer and Pipp 1984). The "fractionation" or splitting of all abilities as a function of tasks and domains is well demonstrated, and we might expect reflective function to be subject to the same kind of developmental *décalage* [unevenness] that characterizes the rest of cognitive development (Flavell 1982). "Fractionation" refers to the tendency for a person not to coordinate skills or experiences that are naturally separate but may be thought of as belonging together by some external criterion (Fischer and Ayoub 1994). Just as the understanding of conservation of liquid does not generalize to conservation of area, reflective capacity in one domain of interpersonal interaction should not be expected to generalize to others. Reflective function does not begin as a general capacity, but is a particular skill tied to the task and domain where it is learned: a specific

category of relationship. Reflective function as a skill may be more or less present in situations as a function of contextual support and emotional states that push an individual up or down a developmental strand. Differences in the meaning of an interaction as well as its physical context can lead to fractionation. For example, the concept of conservation of liquid may not generalize between the experimental task and one that involves helping a thirsty friend, even if both involve pouring a glass of orange juice (Rogoff 1990). We have noted above that the child's observed use and experience of mental-state language can differ markedly across social contexts (Dunn 1996). Fractionation does not disappear entirely with development, either in general or in the specific case of reflectiveness. It is clearly possible for task-based skills such as reflective function to come to be coordinated, but this should not be seen as automatic. Unevenness across situations is likely to remain prevalent even in adults, especially when they are emotional (Fischer and Ayoub 1994).

Normal development proceeds from fractionation toward integration, which involves the construction of specific coordinations among previously separate skills and provides the foundation for more complex, sophisticated control systems (Bidell and Fischer 1994). Abnormalities of reflective function, the continued use of a teleological rather than a mentalizing model for predictive behavior, should not, then, be seen as either a consequence of "arrest and fixation" at an early stage or a "regression" to that stage. Pathologies in the reflective function of the maltreated child may be expected to develop increased complexity with age and time, in a manner similar to other skills. The skill for limited reflectiveness developed by the child to anticipate and forestall maltreatment and its painful physical and psychological impact would be adaptive in their particular world but would be expected to produce sophisticated forms of difficulty rather than straightforward adaptations in other contexts (Noam 1990). The ability to be reflective in general, but to show only minimal reflectiveness in the context of one's own childhood with reference to the mental states of "caregiving" others or in specific relationships that reactivate the same schemata, could be a result of natural fractionation. Unevenness or splitting of reflective ability could also be the consequence of an active (purposeful, con

scious, or unconscious) attempt on the part of the individual not to coordinate or generalize reflective function to specific relationship domains. Here the unevenness is "a developmental achievement," in that the person must create a coordination in order actively to maintain the separation of contexts that would naturally move toward integration. The family might, of course, support such splits with sharp dissociations between their public, proper world and their private tyrannical one. The split is context- and affect-dependent, and skills developed within one context will not necessarily be matched by similar abilities in others. Within an attachment-theory framework we might say that the self is organized so that certain internal working models include considerable reflective components—expectations incorporating the mental states of self and other—while other working models of relationships appear impoverished, indicating only minimal mentalizing skills. In the latter contexts the subject will offer only stereotyped, simple, concrete, low-level descriptions. This does not imply developmental delay or regression; rather, it suggests a remarkably complex ability to coordinate two distinct levels of functioning. It is the function of the abusive or emotionally depriving world within which they developed that engendered in them the sophisticated skills required for such an adaptation. Thus to talk of deficit or absence of a capacity in such individuals would undoubtedly be an oversimplification. Measures of global abilities will not yield difference between these individuals and other groups. Efforts at going beyond mere clinical impression in terms of measurement and quantification have to take on board the unevenness of their abilities and the situational and interpersonal specificity of the failure of reflective function.

In our view, nonreflective internal working models come to dominate the behavior of personality-disordered adults and children when an element of conflict is present within an interpersonal relationship. Conflict—or, rather, its adaptive resolution—prototypically calls for the perception both of the self and of the other in relation to the self, requiring individuals to reconcile their own legitimate claims with concern for the other (Killen and Nucci 1995). Thus, for example, the abnormality of the early family environment of individuals with severe problems of con-

duct has been most evident in the context of normally expectable conflicts (Patterson 1982; Perry, Perry, and Kennedy 1992). In this context the child with a vulnerable capacity for mentalization experiences no affirmation of his intentional stance and fails to acquire the sense of ownership or inner endorsement of his actions that is essential for a sense of self-agency. Consequently, his sense of autonomy becomes vulnerable, and the importance of his original intention is exaggerated. The characteristics of oppositional defiant disorder (e.g., negativity, disobedience, aggression) may in part be seen as attempts at reasserting self-agency in a relationship where the connection between mental state and action within the self has been undermined by insensitive and coercive parenting.

Abnormalities of parenting represent but one route to limitations on reflective function. The child's biological vulnerabilities, such as hyperactivity, inadequate attentiveness, deficiencies in impulse control, are all likely to obstruct his opportunity for evolving a mentalized reflective model of conflictual interpersonal situations. Within a dialectic or transactional model there is a bidirectional causality inherent to such biological vulnerabilities: they both provoke situations of conflict and gravely limit the child's capacity to handle them flexibly (see chapter 9).

The separation of action from intention undermines the emotional reaction individuals may be expected to have in relation to the consequences of their action, since, as Hart and Killen pointed out, the acquisition of moral emotions requires that individuals are "active contributors to their own development, interpreting their world and making judgments that determine their actions in it" (1995, p. 7). Subsequently, the predominant response to situations charged with emotion will be a nonreflective one, readily disowned by the self. Naturally the absence of reflective function in such situations will give the appearance of rigidity to the person's behavior, as if only a singular pattern of response were accessible. Furthermore, the response may frequently be in conflict with social norms because the tendency to take the perspective of others has been abandoned in that context, and consequently the "moral emotions" used to make judgments about the consequences of actions and to regulate behavior (Arsenio and Lover 1995) are absent. The absence of

reflective function may further exaggerate an antisocial response by forcing the individual to see the other not as another intentional agent, but, rather, in nonhuman terms, as a body, as representing a social position or agency, or as an anonymous member of a group.

Maltreatment, or more broadly trauma, is seen as interacting with the domain- and situation-specific restrictions on reflective function at two levels: (*a*) As we have argued, maltreatment presents the young child with a powerful emotional disincentive for taking the perspective of others because of the actual hostility of the intentional stance of the abuser, as well as the constraints upon self-development imposed by an older person's failure to understand and acknowledge the child's budding sense of intentionality. (*b*) The child is deprived of the resilience provided by the capacity to understand a traumatic interpersonal situation (Fonagy et al. 1994). Thus individuals traumatized by their family environment are vulnerable in terms both of the long-term maladaptive effect of their reaction to the trauma and of their reduced capacity to cope with it. The predominantly nonmentalizing stance adopted in such situations therefore further impairs the individual and, in the extreme, their nonmentalizing approach will come to dominate all intimate interpersonal relationships. It is at this stage that severe developmental psychopathology—ultimately entrenched personality disorder—becomes the likely outcome.