THE DEVELOPMENT OF PSYCHOPATHOLOGY FROM INFANCY TO ADULTHOOD: THE MYSTERIOUS UNFOLDING OF DISTURBANCE IN TIME

PETER FONAGY
University College London

ABSTRACT: A model for the development of this mechanism is offered as well as evidence for it from five areas: (1) the nature of the association of early attachment and later cognitive functioning, (2) accumulating evidence for the association between secure attachment and the facility with which internal states are understood and represented, (3) the limited predictive value of early attachment classification, (4) the studies of the biological functions of attachment in other mammalian species, and (5) factor analytic studies of adult attachment scales that suggest the independence of attachment type and attachment quality. The author tentatively proposes that attachment in infancy has the primary evolutionary function of generating a mind capable of inferring and attributing causal motivations and epistemic mental states, and through these arriving at a representation of the self in terms of a set of stable and generalized intentional attributes thus ensuring social collaboration, whereas attachment in adulthood serves the evolutionary function of protecting the self representation from the impairments that social encounter inevitably create. Severe personality pathology arises when the psychological mechanism of attachment is distorted or dysfunctional and cannot fulfill its biological function of preserving the intactness of self representations.

RESUMEN: A partir de las siguientes cinco áreas, ofrecemos un modelo para el desarrollo de este mecanismo, así como pruebas de ello: (i) la naturaleza de la asociación entre la temprana afectividad y el funcionamiento cognitivo posterior, (ii) la acumulación de pruebas para determinar la asociación entre una afectividad segura y la facilidad con que los estados internos son comprendidos y representados, (iii) el limitado valor de predicción de la clasificación de la temprana afectividad, (iv) los estudios de las funciones biológicas de la afectividad en otras especies mamíferas, y (v) los estudios analíticos de factores de las escalas de la afectividad adulta que sugieren la independencia del tipo y calidad de la afectividad. El autor propone tentativamente que la afectividad en la infancia tiene como función primaria de evolución la de generar una mente capaz de deducir y atribuir estados mentales cualitativos, de motivos y de conocimiento.
mientos, y por media de esta llegar a una representación del propio ser en términos de un grupo de atributos inmateriales, establecidos y generalizables, pero sin integrar así la colaboración social, sino en tanto que la afectividad en la época adulta sirve a la función de evolución de proteger la representación propia de las impresiones que las encuentros sociales crean inevitablemente. Una grave patología de la personalidad aparece cuando el mecanismo sociológico de la afectividad está distorsionado o funciona mal y no puede llevar a cabo sus funciones biológicas de preservar inactivas las representaciones de uno mismo.

**RESUMEN.** Un modelo para el desarrollo de este mecanismo es efectivo así que la prueba que funciona, en cinco dimensiones: (i) la naturaleza de la asociación de la estabilidad precesada y del funcionamiento cognitivo subsecuente, (ii) provee de la asociación entre el aprendizaje y la facilidad con la que las etapas interiores son compuestas y representadas, (iii) la validez de predicción limitada de la clasificación del aprendizaje, (iv) los estudios de las funciones biológicas de la asociación entre otras especies mamíferas y (v) los estudios de la fase analítica de la etapa de aprendizaje de la etapa de aprendizaje. El autor propone provisoriamente que el aprendizaje de la pequeña infancia a la función evolutiva principal de generar un espíritu capaz de infine y de alterar los estados de motivación y los estados epistémicos causales y, a través son causales de aportar a una representación del seif en términos de un set de atributos intencionalmente estables y generalizables, preservado por la misma colaboración social, lo que el aprendizaje a la etapa adulta se forma de función evolutiva para proteger la representación del autor de los fenómenos que las encuentros sociales crean inevitablemente. Una patología crítica de la personalidad desliza cuando el mecanismo de aprendizaje psicosem es deformado y disfuncional y no puede recobrar su función biológica de preservación de la naturaleza inactiva de la representación de uno mismo.

**ZUSAMMENFASSUNG.** Ein Modell für die Entwicklung dieses Mechanismus wird angeboten und Beweise aus fünf Bereichen: (i) Das Wesen der Verbindung der frühen Bindung und der spätere, kognitiven Leistung, (ii) nehmende Beweise für die Verbindung zwischen stärkerer Bindung und der Einfachheit mit der inneren Zufriedenheit verstanden und abgebildet werden, (iii) die Begründung der Voraussagbarkeit der Klassifizierungen der frühen Bindung, (iv) die Studien der biologischen Natur der Bindung bei anderen Säugetieren und (v) die qualitativen Studien zur Bindung bei Erwachsenen, die eine Unabhängigkeit des Bindungstyps von der Qualität wahrscheinlich machen. Der Autor zeigte dazu vorzugsweise, dass die Bindung im Kleinen Kindesalter die primäre evolutives Funktion hat einen Geist zu entwickeln, der kausale motivationalen und epistemischen Grenzüberschreitungen herstellen und zusammenbringen kann. **Enseñanza de la representación del autor de los fenómenos que las encuentros sociales crean inevitablemente. Una patología crítica de la personalidad desliza cuando el mecanismo de aprendizaje psicosem es deformado y disfuncional y no puede recobrar su función biológica de preservación de la naturaleza inactiva de la representación de uno mismo.**

**Notes:** 1. 情報の発達はデジタルと、および以下の3つの形態からの連携が、提供される。1) 单一の愛着とそれが全般的な性質を内化した愛着。2) 内化された愛着と、内的な状態が理解され表象されることの容易さと、両の間の関連について、形成しつつある記憶。1) 単一の愛着の分野は主に自己を含み、また内化された愛着がそれを内化することを示唆する。成人愛着の出現を分析研究、著者は、仮に以下の観点を、推論することができる。そして当様と動機としてそれを実現する、この仮想的な動機を、乳児期の愛着を含む、そしてこれらを通じて、一案の安定して一般化した差別の特性の属性に該当する。一つの差別を、そのようににして社会的な共感が成果される、したがって、成人の愛着は、自己表現を、社会的な共感を不可避的に作り出す要素である、普通の機能をもたらす。発達的な病態を、発達の心理学的機能が、あらゆる考えであり、そので自己表現を無視して、また普通に理解するという、生物学的な機能が満たされないときに、生じる。
PREAMBLE

Where Has All the Mystery Gone?

Recently, very unusually for a clinician-academic like myself, I saw three new patients for assessment on the same day. The three new male cases were very different (a depressed journalist with sexual problems, a young man soon to be married but worried about his history of bipolar illness, and an adolescent with violent behavior problems), but they shared a single common theme.

My friend, the late George Morey, taught me to try during initial assessments to elicit the patient’s theory of their problem. Whether this is a good idea or not, I always ask something like: “Why do you think this has happened to you?” or “Why do you think people like you get depressed?” On this particular day I was surprised because all three men came up with identical answers: “I think it is well established to be a chemical imbalance caused by my genes,” said the groom with the bipolar disorder; “I think from my mother I inherited a tendency to look for the negative” answered the journalist, and “I’ve been told I have bad genes that make me hit people” replied the adolescent.

In each case, as they answered, I had the impression of time collapsing. There was no space between the moment their father’s sperm penetrated their mother’s ovum and the present moment. Of course, in each case I was able to call upon the natural human desire to create a meaningful life narrative and to explore how their experiences had assisted or hindered their capacity to cope with the difficulties they brought. But while they expounded their naïve nativist views, there was no room for dialogue. There was just one simple message: don’t ask what causes my problems, don’t probe my memories or thoughts or feelings; there is nothing to know, the answer lies in my genes. There was no room for human mystery!

THE DEMISE OF SOCIALIZATION: PARENTING VERSUS GENETICS

There are three primary agents of socialization of children in western society: families, peer group, and day-care centers or schools (Messcoby, 2000). The emphasis, both professional and cultural, has been on the family as an agent of socialization. For the best part of the past century, both psychological theories and common-sense psychological views were in agreement in identifying experience with parents as pivotal in shaping an individual’s values, beliefs, character and, naturally, dysfunctions in adaptation. It is interesting to note that, of the two psychological approaches which dominated the last century, learning theory and psychoanalysis, it was the latter that retained some emphasis on the constitutional limits to socialization (Freud, 1920).

The last quarter of the 20th century saw a dramatic realignment of developmental theories. The emergence of a cognitive mental science prompted the translation of some learning and many psychodynamic principles into the language of information processing, with presumed mental operations on past experience creating predictable biases and distortions in mental representations (e.g., Bandura, 1977). Cognitive behavioral approaches to development and psychopathology were ultimately saved from tutelage and circularity by two factors: (a) the theory inspired a whole series of brief and effective psychosocial interventions (CBT), and (b) the introduction of a dialectical model into developmental theory (e.g., Kagan, 1989).
The views of socialization that emerged from cognitive social learning theory have underscored the important role of the child in determining their socialization experience. Clearly, mothering an infant high in emotionality must elicit quite a different set of maternal behaviors than the mothering of a sociable, unmotional infant. This realization was critical in radically moderating the parent-blaming tendency of early psychopathologists. These transactional models of child-to-parent effects, however, while both highly desirable and totally appropriate in the way that they were originally proposed and stated, unfortunately became a major plank in the argument of those proposing a nativist revival—but more of this later. The ontogenetic frame of reference of cognitive social learning theory for the most part maintained the environmentalist tradition of psychoanalytic and learning theories.

Developmental psychopathology, permeated with the dialectic of social learning theory, came to dominate child psychiatric epidemiology, under the leadership of Norman Garmezy with other giants such as Michael Rutter, Alan Sroufe, Robert Emde, Dante Cicchetti, and others. The key research question came to be the mysterious unfolding of disturbance through time, the integration and interaction of person and environmental characteristics in the generation of psychological disturbance through ontogenetic development. Notwithstanding its explicit commitment to a dialectic transactional model (e.g., Garmezy, Masten, & Tellegen, 1984), developmental psychopathology always retained its emphasis on socialization, particularly intrafamilial socialization (e.g., Cicchetti, 1987; Sameroff, 1995). Attachment theory became one of the guiding frameworks of the approach, and John Bowlby was, to some degree posthumously, recognized by many as one of its pioneers (Sroufe, 1986). Thus, notwithstanding the dominance of cognitive psychology and social learning theory, developmental psychopathology remained a broad church, and many psychodynamic concerns were retained.

The principal concerns of developmental psychopathologists in the last quarter of the 20th century were mostly around risk factors, with risk factors associated with the family occupying a most important role. Developmental psychopathology of the early years of development was particularly concerned with social and cultural facets of risk, parent–infant relationships, epistemic and motivational mental states that influence parenting (e.g., Belsky, 1984), the interactions of economic and social disadvantage with parenting (e.g., McLoyd, 1990), the distorting influences of past experience on emotional and cognitive structures of the child (e.g., Fox, Platz, & Bentley, 1995), and parental behaviors as mediators of the gross social inequalities that become an increasing source of concern for social scientists of the Thatcher and Reagan years (e.g., Pettit, Bates, & Dodge, 1997). Permeating all these ideas was the notion that the unfolding of psychopathology occurred in the context of the child’s primary socialization environment: the family. The family, in particular the parents, provided the backdrop against which this unfolding occurred: their characteristics were crucial to the developmental choices the child would make, their actions and collaboration critical to both treatment and prevention. But alas, all this was too good to continue.

The Findings from Behavior Genetics

Over the last decade of the 20th century, perhaps partly triggered by the excitement of the human genome project, but also by research designs of increasing statistical sophistication, quantitative behavior genetics was unleashed on early development research. For some time, it seemed as if research in genetics had all but eliminated the place for classical socialization theories with an emphasis on parenting, such as attachment theory, and had refuted all theories that advocated the key role of early family experience (see Scarr, 1992). For example, the behavior geneticist Rowe (1994) wrote: “Parents in most working to professional class families may have little influence on what traits their children may eventually develop as adults.”
(p. 7). He went on to say that he doubted whether any undesirable trait displayed by a child could be significantly modified by anything a parent does.

The biological (genetic) movement of the 1990s highlighted a number of issues of particular relevance to early developmentalists:

1. The overall connection between early parenting and socialization outcomes turns out to be quite weak, and in longitudinal studies, parenting accounts for negligible proportions of the variance. There is very limited evidence that might link early relationship experiences to the development of psychopathology. Most observed associations between socialization and disorder have been reinterpreted in terms of reverse causality: the child’s disorder causes family dysfunction rather than the other way around. For example, critical parental attitudes are more commonly observed in children who have suffered from a psychological disorder for longer (Hooley & Richters, 1995), suggesting that the parents’ exposure to psychopathology increases the likelihood of parental criticism, rather than the other way around.

2. Correlations between characteristics of early parenting and later child behavior, even in prospective studies, can be reinterpreted in a model in which the child’s genetic characteristics are seen as determining the parent’s response, rather than assuming that parenting influences the child. For example, the observed associations between parenting sensitivity and attachment classification may be driven by the behavior of the child and accounted for by the child’s genetically determined predispositions (the so-called child to parent effects). It is also interesting that aspects of a family’s experience of its own interactions are genetically determined. Thus, according to the findings of the Colorado Adoption Project, the parents’ report of warmth and negativity in the family and the child’s report of achievement orientation appear to be genetically determined, suggesting that aspects of the family environment are susceptible to the influence of the child’s genetically rooted characteristics (Deater-Deckard, Fulker, & Plomin, 1999).

3. Parental warmth is influenced by the parents’ genetic endowment, thus the association between warmth and lack of pathology may well be spurious (Losoya, Culver, Rowe, & Goldsmith, 1997).

4. The relative contributions of genes and environment are estimated by examining the observed correlation between two siblings with the correlation that would be expected on the basis of the degree of genetic material the two have in common. Thus, MZ twins who share all genetic material should resemble each other (correlate on a trait) about twice as much as DZ twins. Behavior genetic models of twin and adoption studies partition variability into genetic and environmental components by subtracting the proportion of variability on a specific trait accounted for by shared genes ($h^2$) from 100 ($E = 100 - h^2$). In most domains $h^2$ is 50–60%, with less than half left to $E$. Two large-scale, high-quality, community-based studies—the Virginia Twin Study (Eaves et al., 1997) and the Non-shared Environment and Adolescent Development (NEAD) project (Reiss et al., 1995a)—have confirmed that most types of childhood psychopathology have quite substantial genetic components. For example, heritability estimates for ADHD range from .54 to .82. More or less the only psychological disorder of childhood with a negligible genetic component is separation anxiety. Even for this disorder, heritability estimates for girls are substantial (31–74%), the low figures come from boys (0–19%).

5. Behavior genetics research revealed that influences that had previously been considered environmental were actually genetically mediated (Kendler et al., 1996). Apparently
environmentally mediated family influences, such as children who are read to learning to read sooner than those who are not read to, are, in fact, mostly mediated by the shared genetic predisposition of caregiver and offspring, and are therefore in themselves unimportant (Harris, 1998; Rowe, 1994). Other studies have shown a number of peer influences to be actually genetically determined (e.g., Jacobson & Rowe, 1999). Most recently, an analysis of the Colorado adoption project showed that many of the milder adverse effects on social adjustment associated with parental divorce are in fact genetic: the “divorce gene” causes adjustment problems in children even if they are adopted into nondivorcing families (O’Connor, Caspi, DeFries, & Plomin, 2000).

5. In so far as behavior genetic studies showed family environment as matter, it was environment specific to each child within the same family (nonshared environment) that mattered. Environment may be partitioned into a shared and a nonshared component. If the trait under scrutiny has a shared environmental component, then both MZ and DZ twins would be significantly correlated on the trait, while if nonshared environmental factors are involved, then the siblings should not be correlated. Shared environmental influences may be estimated in adoption studies by comparing the correlation of adopted children and their adopted siblings with children in other households. If shared aspects of the environment such as parenting were indeed formative, then adopted siblings living in the same home should be significantly more alike than unrelated children across households. After the genetic and shared environmental components are estimated, what remains is the nonshared environment (\( r_{non} = 1 - h^2 - \delta^2 \)). The nonshared environment appears to be the backdrop of the environmental component—shared environment, an instance of which would be parental sensitivity, accounts for almost no variance (Plomin, 1994). Adopted children, it seems, are no more like their adopted siblings than unrelated children growing up in a different household (Plomin & Bergeman, 1991). Such findings are particularly striking in the case of twin studies where there are in-built controls for age, gender, temperament, and birth order. This is important because the relatively weak observed effects of the shared environment have been used to suggest that environments generally assumed to be toxic by developmental psychopathology (such as high level of parental conflict, divorce, in-consistent discipline, parental psychiatric disturbance, multiple moves, death of the parent or even relative social disadvantage and neighborhood effects) are either of less importance than previously thought, or, more likely, are actually genetically mediated (Plomin, Chipuer, & Neiderhiser, 1994). Plomin (1994) put this quite elegantly: “So often we have assumed that the key influences on children’s development are shared: their parents’ personality and childhood experiences, the quality of their parents’ marriage relationship, children’s educational background, the neighborhood in which they grow up, and their parents’ attitude to school or to discipline. Yet to the extent that these influences are shared, they cannot account for the differences we observe in children’s outcomes” (p. 23). More or less the only disorders with a substantial shared environmental component are oppositional defiant disorder and conduct disorder (Goldsmith, Buss, & Lemery, 1997).

6. It has been argued that even nonshared environmental effects may be better understood as genetic in origin. Genetically influenced aspects of children’s behavior may be responsible for provoking specific observed responses in parents and other people. This is sometimes termed evocative covariance, when children with different genetic predispositions exhibit complementary responses from the caregiver. Thus, the child’s nonshared (specific) environment may have sometimes been erroneously attributed to pa-
rental behavior rather than to his/her genes. As much as 20% of the variability in how parents treat adolescents may be due to the genetic characteristics of the adolescent (O'Connor, Hetherington, Reiss, & Plomin, 1995). There is evidence, for example, from studies of adopted children that authoritarian parenting may be elicited by the child’s resistive or distractible behavior (Ge, Canger, Cadoret, Neiderhiser, & Yates, 1999). The celebrated results of the NEAD study, showing that adolescents who are observed to receive preferentially negative treatment from the parents relative to their twin siblings are at greater risk of developing depression and antisocial symptoms, while the more positively treated siblings are actually protected from these disorders (Reiss et al., 1995b), may also be an instance of a pseudoenvironmental effect. The correlation between parental conflict-negativity and adolescent disorder could also be accounted for by genetic factors in the adolescent (Neiderhiser, Reiss, Hetherington, & Plomin, 1999).

These are just the highlights. It seems that over the past 10 years we, both developmental professionals and the lay public, have unconsciously switched from a primarily psychosocial model of child and adult development to a genetic-biological frame of reference that often a priori excludes consideration of child-parent relationships. In an informal study, we asked 20 consecutive parents referred to an outpatient child community mental health clinic about the likely cause of their child’s problems. It surprised no one that they all put brain chemistry at the top of the list. It was more surprising that “bad genes” came second, peers third, and early life experiences a poor fifth, just ahead of food additives. Why is this happening? The excitement of novelty, of scientific discovery, must have something to do with it. But there is more to it than that. The reduction of the mind to chemicals was appealing even to Freud. Although our consciousness, our free will, our mind is undoubtedly our most treasured possession, it is also the source of all our sadness, conflict, pain, suffering, and misery. The reduction of models of pathology to a principally genetic mode of causation is undoubtedly a relatively comfortable solution for all of us. But like all comforts, it comes at a price.

I now intend to argue that the case for reducing the emphasis on parenting, particularly the emphasis on the early attachment relationships, is based on false evaluations of behavior genetics data. I will also suggest that perhaps in the past our emphasis on the role of parenting was somewhat naïve in trying to see the parents’ influence simply in terms of relationship quality, or in terms of mechanisms of internalization, introjection, identification and so forth. I will try to show that (1) early attachment experiences may well be key moderators of the expression of individual genotype, and (2) that the primary evolutionary function of attachment may be its contribution to the ontogenetic creation of a mental mechanism that could serve to moderate psychosocial experiences relevant for gene expression.

SCRUTINIZING THE CASE FOR GENETICS

As we have seen, the case for genetic determinants rests on two pillars: the first is the weakness of the socialization evidence, and the second is the findings of quantitative behavior genetics.

The Current Case for Socialization

Classical reviews extensively cited by behavior geneticists tended to reveal weak correlations between parenting and socialization outcomes open to alternative, genetic interpretation. However, there have been substantial methodological improvements in studies of socialization, both in terms of the breadth and the depth of measurement, and effect sizes have increased correspondingly. For example, Martin Maldonado and his colleagues at the Menninger Child and
Family Center carried out operationalized clinical ratings of almost 150 infants. Over 70% of the children have so far been followed up, two to four years later. Infant behavior problems strongly predicted later behavioral difficulties in a number of areas. Parenting observed in infancy predicted preschool emotional difficulties. In particular, parents who were both neglectful and hostile had children with more behavioral problems. This remained true even when behavioral problems in infancy were controlled for. Observed neglect and hostility correlated .36 with the child's behavioral problems noted at four years, even when infant behavior was already controlled for. This implies that parenting had a predictable effect beyond that which could be explained in terms of the parent responding specifically to the precursors of the child's behavioral problems.

There is increasing support for the hypothesis that many of the effects of socioeconomic disadvantage, social adversities, family transitions, and other risk factors for the development of psychological disturbance are mediated by distortions in intrafamilial social relation- relationships, particularly in the relationships between the child and the parent. For example, a large-scale study of 9,000 children in the UK, living in various family configurations (step-, single-parent, and onestep families) found that the quality (degree of negativity) of children's relationship with their caregiver was related to adjustment outcome, regardless of the degree of the biological relationship between child and caregiver (Dunn, Deater-Deckard, Pickering, O'Connor, & Goldberg, 1998). An accumulation of findings links the course of childhood dis- order to characteristics of parent—child relations, specifically expressed emotion (EE). The extent of family criticism, hostility, and emotional overinvolvement predicted recovery rates one year after hospitalization in a sample of children hospitalized for depression irrespective of treatment method, illness duration or comorbidity (Asarnow, Tompson, Hamilton, Goldstein, & Guthrie, 1994). Studies of children with physical disorders also imply that the course of these disorders is likely to be affected by parenting characteristics. For example, Mrazek and colleagues (Mrazek et al., 1999) studied the association between early parental behavior and the later onset of asthma in 150 children at genetic risk of developing the disorder. Poor maternal coping and parenting problems were associated with the child developing the problem before the age of three years. Parenting characteristics appeared to accelerate the onset of the disease but did not increase the overall risk for the disorder. It appears that the over twofold increase of the risk of psychopathology in children with chronic physical illness is mediated by the quality of family functioning under the stress of the child's disorder (Cohen, 1999). Not surprisingly, then, family-based interventions tend to be effective in reducing the psychological complications of chronic physical illness in children (Walker, Johnson, Manion, & Coutuier, 1996). Family conflict is associated with poor adherence in both asthma and diabetes (Diamont et al., 1995; Weinstein & Faust, 1997) and decreasing family conflict appears to consistently improve illness outcome (Weinstein, Faust, McKee, & Padman, 1992).

Better-controlled studies are accumulating, and current reviews of the socialization liter- ature tend to yield more encouraging conclusions (Macoby, 2000). In the meantime, the question has shifted from attempts to establish the nature or extent of parenting influence upon child development to questions concerning the range of factors that influence parent—child relationships. Characteristics that have been reliably demonstrated to be significant in this context include: (1) parents' childhood relationships or experiences with their own parents (Simons, Beaman, Conger, & Chao, 1993) with adverse early experiences increasing the like- lhood of family conflict and divorce (O'Connor, Thorpe, Dunn, & Goldberg, 1999); (2) eco- nomic and social disadvantage that is linked to poorer parenting practices (e.g., Hughes, Deater- Deckard, & Cutting, 1999); (3) the parent's selection of a partner with a history of deprivation and disadvantage has a negative impact that is increased by that parent's own history of adverse early experience (e.g., Krueger, Moffitt, Caspi, Bleske, & Silva, 1998); (4) length of time
parents have spent in their current family setting is likely to be related to the quality of their relationships with their children (Hetherington, Bridges, & Insabella, 1998); (5) biological relatedness (the experience of "ownership" in relation to a child) enhances emotional relatedness between parent and child (Dunn, Davies, O'Connor, & Sturgeess, 2000); (6) parent–child relationships are strongly influenced by the relationship history of the other parent (Dunn et al., 2000) as well as by aspects of the sibling relationships (Dunn et al., 1998), which suggests that a systemic perspective is essential to arriving at an adequate account. These are simply some illustrative recent findings, but they stress the complex character of the interaction between the social environment provided by the family and aspects of parent–child relationships with a potential to influence childhood outcome.

The recognition that the parenting environment varies in significance across developmental stages and even within each developmental phase across time introduces a considerable degree of further complexity into the interpretation of social environment to behavioral outcome correlations. Therapeutic experience teaches us of the exceptional influence of certain "key moments" of interaction between therapist and patient, when experience brought into clear relief by a confluence of circumstance and intrapsychic factors suddenly enables therapeutic change. This idea is fully elaborated in the work of Daniel Stern, Ed Tronick, Karlen Lyons-Ruth, and the Boston group on therapeutic process (Stern et al., 1998). They argue that change in therapy might be a function of special moments of attunement between therapist and patient. More generally, the same could apply to special moments of influence of parents on children—naturally both with a positive and negative valence—moments of parenting influence, which, however, key and formative, might be the needles in a socialization haystack. It is hard to imagine how observational research, focusing on aggregate or time-sampled behaviors, can hope to capture significant numbers of such key moments. The correlations between observed parenting and child outcome may never reflect the true influence of parenting.

In any case, however, the associations between parenting and socialization, the possibility that the more parsimonious explanation is genetic cannot be ruled out. It is the apparent strength of genetic findings, the massive proportion of variability accounted for, no matter how specific the trait, which casts such a dark shadow over developmental psychopathology. The room they leave for socialization, for the mysteries and uncertainty of development, is limited. But are the genetic findings as unequivocal as they seem? The evidence from behavior genetics should be interpreted with caution. The reasons are: (1) methodological, (2) conceptual, and (3) empirical.

**METHODOLOGICAL PROBLEMS WITH THE EVIDENCE FROM BEHAVIOR GENETICS**

Methodologically, the contrast of identical and fraternal twins confounds genetic similarity and environmental influence. It has been claimed that identical twins have more similar environments than fraternal ones (e.g., they have more friends in common, they are treated more similarly by the parents) (Reiss, Niederbisser, Hetherington, & Plomin, 2000). The status of inferring environmental effects by a process of subtraction using the additive model has also been questioned (Turkheimer, 1998). In particular, E is often estimated by subtraction, with no attempt at a direct measure of environmental factors. If the estimate for heritability is high, E must be low. In reality, G and E combine to generate a phenotype. In simple additive models, however, this interaction would be pooled with genetic effects. Further, measurement error is normally pooled with the nonshared environment category. This could artificially increase the proportion of variance attributed to the nonshared environment, especially when the measures used happened to be poor.
The source of data about pathology significantly influences heritability estimates. The agreement between raters and methods in the Virginia Twin study was only moderate, with parent-child agreement rarely above .3 for most clinical conditions (K Hewitt et al., 1997). The use of parents as a source of data about themselves as well as their children creates an inherent genetic bias. Heritability estimates are inflated by the use of parents’ reports of child behavior rather than behavioral observation or self-reports. It is not surprising that when parents rate a child’s aggression, the correlation with the parent’s aggression should increase (Miles & Carey, 1997). It is more challenging to explain why the heritability of internalizing disorders would be 60–72% when the informants were the parents and 11% when the informant was the child (Eaves et al., 1997).

CONCEPTUAL PROBLEMS WITH THE EVIDENCE FROM BEHAVIOR GENETICS

At a conceptual level, we may certainly question the notion of nonshared environment, because it merely refers to interindividual differences, not to their environment. In fact, shared environments could easily serve to make children in the family different from one another as to increase intrafamilial similarity, because shared environments may be experienced very differently by two children. A further conceptual problem concerns heritability estimates based solely on individual differences. Such estimates remove shared environmental effects such as secular trends, and are strictly restricted to the environment studied. Height, IQ, and the prevalence of a number of psychological disorders have increased markedly over the last century, undoubtedly as a consequence of environmental changes, yet current behavioral genetic methods of estimating environmental effects preclude consideration of these.

EMPIRICAL PROBLEMS WITH THE EVIDENCE FROM BEHAVIOR GENETICS

Empirically, we could point to studies where environmental determinants revealed substantial effects after genetic influences had been excluded (Johnson, Cohen, Brown, Sniselski, & Bernstein, 1999). We could raise questions about the actual (rather than assumed) differential responsiveness of caregivers to siblings. Evidence on just how differently siblings are treated is actually quite mixed. In one of the only behavioral genetic studies to actually look at the child’s environment rather than simply infer it, Reiss, Plohm, Hechertang, and colleagues found direct evidence for the notion of the nonshared environment (Reiss et al., 2000). The difference between the degree of coerciveness of parenting between two twins was more predictive than the absolute level of negativity (Reiss et al., 1995b). However, Judy Dunn’s naturalistic observational studies of siblings actually suggest that whilst cross-sectionally parents may appear to be treating siblings differently, looked at longitudinally children at various ages receive comparable treatment (Dunn & McGuire, 1994). Regardless of the ultimate conclusion concerning the differential treatment of siblings, the fact that studies of social development tended to look at single children implied that they have on the whole underestimated the impact of parenting and other shared environmental influences. As we shall see later on, there may be specific pressures in family systems for differential responses of siblings as part of the need for each person within the system to have a unique role. Interestingly, the pressure for difference may be greater when genetic differentiation is least. Further, experimental manipulations of the environment as part of treatment and prevention interventions have yielded relatively large effects on occasions. It is worthy of note that neither of the two major attacks on the importance of family on socialization (Harris, 1998; Rowe, 1994) cover parent training. The average effect
size of parent training for children with ODD is around 1 (Serketich & Dumas, 1996). More relevant in this context, accumulating evidence supports the usefulness of experimental interventions with parents such as home visitation (e.g., Olsd et al., 1998), with long-term beneficial effects in reducing risk of criminality and delinquency. Of course, the impact of environmental manipulation is often not as large as one would hope; moreover, long term follow-ups in treatment studies are relatively rare and even quite impressive changes initiated by experimental interventions dissipate (Sonagy, Target, Cotrell, Phillips, & Kurtz, 2000).

As clinicians, our main objection to behavior genetic data would not be methodological, conceptual, or empirical but rather pragmatic. Genetic effects may well be indirect as well as direct. Even a high genetic loading for a certain environmental hazard does not mean that the consequences associated with that risk factor would necessarily be genetically rather than environmentally mediated. If child abuse, for example, were found to have a large genetic component, its toxic effects would still be via the destruction of trust in the world for the abused child, rather than via a purely genetic process. The implications of behavior genetic data for clinical intervention are thus quite limited.

THE ROLE OF EXPERIENCE IN THE EXPRESSION OF THE GENOTYPE

It is universally agreed that developmental psychopathology involves a gene–environment interaction. Empirically, this interaction term has proved to be quite hard to find. Plomin’s systematic review of the literature now admitted somewhat dated, only found evidence for relatively isolated examples (Plomin, DeFries, McLearn, & Rutter, 1997). Some quantitative behavior genetic studies strongly suggest interactive processes in which a genetic vulnerability is triggered by environmental exposure. For example, the classic Finnish adoptive family study of schizophrenia suggests that children with a schizophrenic biological parent were more likely to develop a range of psychiatric problems if, and only if, they were adopted into dysfunctional families (Tenfesari, Wynne, Morling, Labri, & Naarich, 1994). Brideman (1996) reported that criminality only appeared to be associated with a genetic risk if children whose biological parents were criminals were adopted into dysfunctional homes. Genetic risk may or may not become manifest, depending on the quality of the family environment so which a child is exposed. But if this is such a pervasive process, why is the quantitative behavior genetic evidence so sparse?

The obvious answer is that the environment that triggers the expression of a gene is not objective. The child’s experience of the environment is what counts, and this is a function of appraisal. The manner in which environment is experienced will act as a filter in the expression of genotype into phenotype. Here we touch on the pivotal role of parenting for genetic research, particularly attachment theory. The primary concern of attachment theory is with the interaction of multiple layers of representations in generating developmental outcomes. Data from genetics call for exactly such sophistication in understanding the way genes may or may not be expressed in particular individuals. This idea was beautifully elaborated by a founding father of WAIMH, Robert Emde (1988), a decade ago.

The pathway between genes and phenotypes is a tortuous one, along which genetics and the environment constantly interact. Internal and external stimuli, steps in the development of the brain, hormones, stress, learning, and social interaction, alter the binding of transcription regulators (Kandel, 1998). For example, although risk factors operate in combination, there is substantial individual variability in response to stress and adversity. Much of this variability is poorly understood (Rutter, 1996), but it underscores the potential importance of intrapsychic variables. Whether or not specific environmental factors trigger the expression of a gene vary
depend not only on the nature of those factors, but also on the way the child experiences them, which in turn, in many instances will be a function of attachment and other intrapsychic experiences. The experiential filter that attachment provides may, in turn, be a function of either genetic or environmental influences, or of their interaction (Kandel, 1998). Thus, intrapsychic representational processes are not just consequences of environmental and genetic effects—they may be the critical mediators of these effects.

This point has substantial clinical significance, because a child’s understanding of his/her environment is more readily modifiable than the environment itself, or the genes with which the environment interacts (Emde, 1988). In attachment theory, intrapsychic perspectives may be helpful in considering, not just what precipitates a disorder, but also what processes influence the course of the disorder for better or worse. Up until the last five years this idea was theory—now the collaboration of molecular geneticists and attachment theory is making it a reality. I shall give just one example of this powerful paradigm.

An important result has recently been published from the Budapest Infant-Parent Study (Lakatos, Toth, Nemoda, Ney, Sasvari-Stekely, & Gervai, 2000). These workers found an association between the DRD4 receptor exon polymorphism and disorganized attachment classification in 12-month-old infants. Considerable evidence over the years has linked behavioral problems in both children and adults with the seven-repeat allele of the DRD4 gene. In particular, ADHD has been implicated (Parsons et al., 1999) although not all studies concur (Casella et al., 1999). The review by Swanson and colleagues (Swanson et al., 2000) confirmed the likely role of the seven-repeat allele of this gene in making the postsynaptic receptor subtype sensitive. Corr et al. (1999) report findings related to impulsive, compulsive, addictive behaviors that indicate a greater complexity than a sole focus on the seven versus non-seven alleles of the DRD4 gene. In view of recent findings that have linked disorganized attachment in infancy to clinical conditions in middle childhood, it may be particularly important that in this study 71% of the infants classified as disorganized were found to have at least one seven-repeat allele, in contrast with only 29% of the non-disorganized group. Thus, infants classified as disorganized were more than four times more likely to be carrying this allele.

This finding is consistent with observations that neurological (Pipp-Saegel, Saegel, & Dean, 1999) and neonatal behavioral organization (Spangler, Fremouw-Booth, & Grossman, 1996) may anticipate a disorganized attachment classification. It might at first sight seem at odds, however, with the classical observation that disorganized infant attachment was linked to unresolved loss or trauma in the mother (Main & Hesse, 1990). A recent prospective study led by Pat Hughes, in which I had the privilege of participating, confirmed that mothers with a history of perinatal bereavement were far more likely to have disorganized infants than controls. Whereas almost 45% of the mothers who had lost their first baby during pregnancy had infants classified as disorganized at one year, only 20% of the control mothers matched for age, SES, and education did so. Adult attachment interviews collected before the birth of the child picked up the risk for disorganization. Lack of resolution of mourning mediates the association of still-birth experience and disorganization of the “replacement infant” in the Strange Situation.

But only 62% of the mothers with unresolved AAI classification had infants classified as disorganized, although specificity was relatively high (over 80%) of disorganized infants had unresolved mothers). It seems that lack of resolution of mourning may be a necessary but not sufficient condition for disorganization. Because only a third of the children in the Budapest study with the seven-repeat allele showed disorganized attachment, checking for the presence of the seven-repeat allele might, of course, explain the discrepancy. It is possible that the abnormalities in infant–mother interaction, assumed to be associated with lack of resolution of bereavement, may impact more on individuals whose mesolimbic dopamine system is func-
tioning less efficiently, for which the seven-repeat allele of the D1 receptor may be a marker. It has been suggested that the mesolimbic dopamine system controls behavior motivated by reward (Robbins & Everitt, 1999), and less sensitive D1 dopamine receptors (Van Tol, Wu, Guss, Chara, Buxrow, & Cirelli, 1992) could further distort the signal value of the mother’s respons. The review by Swanson et al. (2000) suggests that dopamine underactivity compromises transitional systems that might exaggerate the impact of subtle anomalies of the mother’s behavior in relation to her infant (e.g., momentary dissociation, frightened or frightening behavior, etc.; see Solomon & George, 1999). This is clearly speculative, but is a readily testable hypothesis that is in line with the general internaional model that we propose.

To summarize, a posttraumatic state in the mother might possibly interact with another dopamine receptor, causing dysfunctional attachment organization and subsequent psychological disturbance. All these findings are rather tenuous at the moment, but all three are consistent with the notion of early experience with the caregiver triggering gene expression. Taken together, they suggest quite a fruitful line of investigation, which, given the relative facility with which samples can be collected and analyzed, might become an important adjunct to most of our work in developmental psychopathology.

THE GENESIS OF AN APPRAISAL MECHANISM

We have seen that the importance of family environment may have been underestimated in behavioral genetics research, for methodological, conceptual, as well as empirical reasons. I have also tried to construct a primate case for a representational system that forms an active filter between the genotype and the phenotype. In other words, the mental processing of experience is critical for the expression of genetic material, and these lie substantial interactions between gene and environment. I would like to conclude by suggesting that the genesis of the appraisal mechanism is intrinsically linked to human attachment. Perhaps more than shaping the quality of subsequent relationships (for which evidence is lacking), the early relationship environment serves to equip the individual with a processing system. The creation of this representational system is arguably the most important evolutionary function for attachment to a caregiver. Adopting this perspective helps redress the prevailing bias against the equality of the family as the major force in socialization, but it also shifts the emphasis from contents of experience to psychological structure or mental mechanism, and involves expanding on current ideas of the evolutionary function of attachment.

John Bowlby, a major Darwin scholar, was impressed by the obvious selection advantages of infant protest at separation, i.e., protection from predators (Bowlby, 1969). Given that phylogenetically and ontogenetically, infancy is a period of extreme risk, natural selection would inarguably favor individuals with a capacity for attachment. There has been a revolution in evolutionary theory since Bowlby’s time. We now realize that “survival of the fittest” cannot guarantee the natural selection of a behavior. Only the reproduction of genetic material can achieve this (Hamilton, 1964). This is the theory of inclusive fitness. One does not need to survive and reproduce oneself for one’s genes to be replicated. For example, some organisms will forgo reproduction to ensure the reproductive potential of their genetically close relatives. The concept of “inclusive fitness” places attachment theory at the center stage of evolutionary sociobiology as a key behavioral mechanism mediating the establishment of genetic proximity. Attachment is the process that ensures that we know whose survival will advantage the reproduction of our genes. Of course it may have additional evolutionary functions. It is possible that attachment marks individuals with whom we should not mate because of the biological risks associated with interbreeding and incest. Adult attachment may also be a marker for reciprocal altruism. Altruism and cooperation (Axelrod, 1984), the “quid-pro-quo” strategy
of helping nonkin if, and only if, they have done something for one, might also be underpinned by the mechanisms of attachment. Attachment is likely to minimize the adverse effects of "cheaters," individuals who do not reciprocate equitably in groups over time and to whom we are unlikely to become attached. This would be a good example of a further interesting facet of evolution, how a mechanism that evolved for one purpose (the protection of the vulnerable infant) may be put to good biological use in the context of the adoptive problems of subsequent developmental phases. But all these potential biological functions would apply as readily to animal models of attachment as the human infant. If the biological function of attachment is to be a pillar in our argument for the importance of parenting, we need to restrict ourselves to uniquely human capacities.

The generally recognized components of attachment behaviors that serve to establish and maintain proximity are: (1) signals that draw the caregivers to their children (e.g., smiling); (2) aversive behaviors (such as crying), which perform the same function; and (3) skeletal muscle activity (primarily locomotion) that brings the child to the caregiver. There is a fourth component that provides a better evolutionary rationale for the entire enterprise of human attachment, going beyond the issue of physical protection. According to Bowlby, at about the age of three, behaviors signifying a goal-directed partnership begin to emerge. The central psychological processes that mediate goal-directed partnerships are the internal working models.

Bowlby's original concept has been thoughtfully elaborated by some of the greatest minds in the attachment field, and no attempt to duplicate this will be undertaken here. However, it might be helpful to summarize the four representational systems that are implied in these reformulations: (1) expectations of interactive attributes of early caregivers created in the first year of life and subsequently elaborated; (2) event representations by which general and specific memories of attachment-related experiences are encoded and retrieved; (3) autobiographical memories by which specific events are conceptually connected because of their relation to a continuing personal narrative and developing self-understanding; (4) understanding of the psychological characteristics of other people (inheriting and attributing causal motivational mind states such as desires and emotions and epistemic mind states such as intentions and beliefs) and differentiating these from those of the self. Thus, a key developmental attainment of the IWM is the creation of a processing system for the self (and significant others) in terms of a set of stable and generalized intentional attributes, such as desires, emotions, intentions, and beliefs inferred from recurring invariant patterns in the history of previous interactions. The child comes to be able to use this representational system to predict the other's or the self's behavior in conjunction with local, more transient intentional states inferred from a given situation.

Classically, in attachment theory this phase change from behavior to representation is generally regarded as a modification of the attachment system propelled by cognitive development (Mervin & Briner, 1999). Our contention here is the reverse: rather than seeing the biological role of attachment shifting ontogenetically as a consequence of other, biologically driven, maturational changes, we propose that a major selective advantage conferred by attachment on humans was the opportunity it afforded for the development of social intelligence and meaning making. The capacity for "interpretation," which Bogdash (1997) defined as "organism making sense of each other in contexts where this matters biologically" (p. 10), becomes uniquely human when others are engaged "psychologically in sharing experiences, information and affects" (p. 94). The capacity to interpret human behavior—to make sense of each other—requires the intentional stance: "treating the object whose behavior you want to predict as a rational agent with beliefs and desires" (Dennett, 1987, p. 15).

The capacity for interpretation in psychological terms—let's call this the Interpersonal Interpretive Mechanism or IIM—is not just a generator or mediator of attachment experience,
it is also a product of the complex psychological processes engendered by close proximity in infancy to another human being, the attachment figure. It is not the IWM. It does not contain representations of experiences, and is not a repository of personal encounters with the caregiver. Rather, it is a mechanism for processing new experiences. We could call it an internal working model, if the meaning of this term was clarified: Bowlby applied IWM interchangeably to a processor of experience—a computational module—and as a repository of experiences.

The IIM overlaps the notion of a “Theory of mind,” the ability to attribute independent mental states to self and others to explain and predict behavior. It has been suggested that this ability arises from a dedicated domain-specific and possibly modular cognitive mechanism (Leslie & Thaiss, 1992). Studies that have identified selective impairments in theory of mind in patients with Asperger’s and autism support this notion (Happe & Frith, 1996). Recent attempts to link theory of mind to specific areas of the brain (Frith & Frith, 1999) have been consistent with viewing this unique human capacity as a hard-wired constitutionally firmly established maturational function, such as language or vision. We have taken an alternative view, linking theory of mind more closely to social than to cognitive development. Theory of mind is in many ways an unsatisfactory term, and tends to be operationalized in terms of relatively simple experimental tasks that call for the child’s appreciation of a puppet’s putative perspective (e.g., Wimmer & Perner, 1983). Others refer to the capacity as mentalization (Fonagy, Steele, Mossin, Steele, & Higgitt, 1991; Morton & Frith, 1995), implying a process whereby mental representations of mental states are created. In previous work we have referred to mentalization that is concerned with emotionally charged (attachment) relationships as reflective function (Fonagy & Target, 1997). This term, however, conflates the measurement of this capacity on the basis of attachment related narratives (Fonagy, Steele, Steele, & Target, 1997) with an assumed psychological capacity that may be involved to greater or lesser extent in generating narratives variably imbued with mental state contents. In any case, mentalization is just one aspect of the IIM. We conceive of the IIM as that collection of neurocognitive mechanisms that are naturally selected to evolve under the influence of early social interactions (predominantly) with the attachment figure, and that serve to mediate the impact of the quality of early relationships into adult personality functioning. These mechanisms include the second-order representation of affect and through this its regulation, the regulation of attention, particularly effortful control, alongside aspects of mentalization, both implicit and explicit (Fonagy & Target, in press). However, it is likely to have many more components than these, all of which have the function of taking forward early relationship experiences into adult personality functioning.

THE ONTOGENESIS OF THE IIM

How is the IIM created out of the secure base? Clearly, there must be biological preparedness, but in our view this is not separable from the infant’s experience of the caregiving environment. Jon Allen (personal communication) drew attention to the possibility of integrating the innate and constructivist perspectives in the development of representational redesignation suggested by Karmiloff-Smith (1992) in her book, Beyond Modularity. Representational redesign involves making what is in the mind available to the mind via some kind of endogenous feedback mechanism. She proposes that the innate equipment drives attention in the service of module construction. She elaborates the principle in the context of language acquisition but it applies with equal force to the development of an interpersonal representational system: “There must therefore be some innate component to the acquisition of language—but, to reiterate, this does not mean that there has to be a ready-made module. Attention biases and some innate predispositions could lead the child to focus on linguistically relevant input and, with time, to
build up linguistic representations that are domain-specific. Since we process language very rapidly, the system might at time close itself off from other influences—i.e., become relatively modularized" (p. 36).

In answering this question we are drawing upon George Gergely’s and John Watson’s model (Gergely & Watson, 1996, 1999), which we have elaborated in a coauthored monograph (Fonagy et al., 2002). The Gergely-Watson model assumes that the human infant has an attentional bias or filter for detecting contingencies between its own behavior and the responses in the environment that are contingent with it. This mechanism works in the service of representation building for the mentalization capacity.

Our core idea is that the attachment context provides the setting in which the infant can develop a sensitivity to self-states, through what Gergely has termed “psychofeedback” or social biofeedback. This is basically the development of a second-order symbolic representational system for motivational and epistemic mind states. Secondary representations of self-states are established through this social biofeedback loop, even though the meta-representational capacity itself may be an innate potential waiting to be “fed” by (tailored) environmental input. The intensification of the mother’s mirroring response of the infant’s distress (caregiving behavior) comes to represent an internal state. Empathic emotion comprises social biofeedback: insofar as the mother’s expressiveness is yoked to the infant’s emotional state. The infant internalizes the mother’s empathic expression by developing a secondary representation of his emotional state with the mother’s empathic face as the signifier and his own emotional arousal as the signified. The mother’s expression tempers emotion to the extent that it is separate and different from the primary experience, although crucially it is not recognized as the mother’s experience, but as an organizer of self-state. It is this “intersubjectivity,” which is the bedrock of the intimate connection between attachment and self-regulation.

It should be noted that intersubjectivity in this context may be a misnomer. The infant at this stage 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. At this level of human proximity the other’s subjective state is automatically referred to the self. In infancy the contingent responding of the attachment figure is thus far more than the provision of reassurance about a protective presence. It is the principal means by which we acquire understanding of our own internal states, which is an intermediate step in the acquisition of an understanding of others as psychological entities—the intentional stance.

In the first year, the infant only has primary awareness of being in a particular, internal, emotional state. Such awareness is conscious or epiphenomenal in that it is not put to any functional use by the system. It is in the process of psychofeedback that these internal experiences are more closely attended to and evolve a functional role (a signal value) and a function in modulating or inhibiting action. Thus, it is attachment processes that ensure the move from primary awareness of internal states to functional awareness. In functional awareness a feeling of anger may be used to simulate and so to infer the other’s corresponding mental state or may be used to serve a signal value to direct action. A final level of awareness is reflective awareness, where the individual can make a causal mind state become the object of attention before, and without it causing action. Whereas functional awareness is intrinsically coupled with action, reflective awareness is separate from it. It has the capacity to move away from physical reality and may be felt to be no more real.

Many studies provide evidence consistent with this model. For example, a study carried out by us showed that the rapid soothing of distressed six-month-olds could be predicted on the basis of ratings of emotional content of the mother’s facial expression during the process of soothing; mothers of rapid responders showed somewhat more fear, somewhat less joy but most typically a range of other affects in addition to fear and sadness. Mothers of rapid-
sponders were far more likely to manifest multiple affect states (complex affects). We interpreted these results as supporting Gergely and Watson’s notion of the mother’s face being a secondary representation of the infant’s experience—the same and yet not the same. This is functional awareness with the capacity to modulate affect states.

A further set of studies, performed by Gergely and his colleagues in Budapest as well as our lab in London with an ongoing replication in Topeka, KS, explored one-year-olds’ understanding of conflicting affect. In one study recently presented at ICBS, 12-month attachment classification, particularly secure and disorganized, was found to be predicted by infant behavior at 6.5 months in a modified still paradigm (Koós, Gergely, Gervai, & Tóth, 2000). The paradigm involves the mother being instructed according to the still-face protocol, but facing a mirror where the infant has a choice between looking at the mother’s face or looking at a perfectly contingent image (themselves). Infants classified as securely attached six months later engaged in significant amounts of active testing of their mirror self-image only when their mother became temporarily inaccessible (the still-face period). By contrast, babies who went on to manifest disorganized attachment six months later were drawn to the image of their fully contingent self-movements all through the laboratory testing. Interestingly, the Koós et al. study also demonstrated that following the still-face period the infants who engaged in more contingency testing looking at their self-image showed more positive affect following the procedure. This led to more successful affect regulation in disorganized infants than in secure ones. Yet seeking for perfect contingency in attempts to detect internal states in the context of human interaction will be of limited effectiveness in the long run. It characterizes the dissociative style of attention organization that is typical of disorganized attachment.

EVIDENCE FOR THE INTERPERSONAL INTERPRETIVE MECHANISM

Is there any evidence for an IIM that evolves out of the attachment relationship, with its efficiency conditioned by attachment security? First, there is unequivocal evidence from two decades of longitudinal research that secure attachment in infancy is strongly associated with the precocious development of a range of capacities that depend on interpretive or symbolic skills, such as exploration and play, intelligence and language ability, ego resilience and ego control, frustration tolerance, curiosity, self-recognition, social cognitive capacities, and so on. Attachment security foreshadows cognitive competence, exploratory skill, emotion regulation, communication style and other outcomes. In our view, this is not because of the general impact of attachment security on the child’s self-confidence, initiative or ego functioning, or other broader personality processes, but rather because attachment processes provide the key evolutionary prepared paths for an interpersonal interpretive capacity to develop. Thus, it is not the first attachments that are formative, it is not attachment security per se that predicts good outcome on this dazzling array of measures; rather, the features of the interpersonal environment that generate attachment security during the first year of life also prepare the ground for the rapid and concrete ontogenetic evolution of interpersonal interpretation. One problem with attempting to trace some of the long-term outcomes of secure attachment in infancy has been the apparently conservative strategy of controlling for numerous aspects of this interpretive capacity. Controlling for verbal fluency or even IQ removes a part of the variability in which the attachment relationship arguably plays a causal role. But this is not an issue which this article has scope to explore.

Second, a number of specific findings in the literature link attachment to the development of an IIM. Laser and Thompson (1998) reported that securely attached children have higher competence in understanding negative emotion. A unique study by Jude Cassidy and colleagues
(Cassidy, Kinsh, Scelton, & Parke, 1996) found that securely attached kindergarteners were less likely to infer hostile intent in stories with ambiguous content and this bias appeared to mediate their superiority in sociometric status. In the London Parent–Child Project, Minami and Howard Steele and Juliet Hodler and I reported precarious performance on theory of mind tasks amongst five-year-olds with a history of secure attachment in infancy. Since then, other investigators have also reported this finding (Meins, Perny, Buss, & Clarks-Carter, 1998).

Third, in a relatively full exploration of findings linking early attachment and later development, Ross Thompson (1999) concludes that "the strength of the relationship between infant security and later socio-personality functioning is modest" (p. 289). The associations are stronger contemporaneously than they are predictively. Within the context of the present theory, it is not how the content of internal working models is determined by early experience that is of interest; rather, the extent to which early experience can jeopardize the very existence of a model, the processing skills required to deal with interpersonal interactions, the robustness of the model, the extent to which this interpretative mechanism can function under stress and process emotionally charged information, is the focus of the present enquiry. Thus, attachment classification might or might not be stable from infancy through middle childhood to adolescence. As prediction comes from the IIM, not from attachment security per se, this is no great concern.

The focus of study should not be attachment security, which achieved significance as a correlate of the IIM, but has little stability, and possibly little predictive value. Rather, the interpersonal interpretive mechanism, which is a genetically defined capacity, probably localized in the medial prefrontal cortex, is the mechanism of predictive significance. Studies with patients with orbital-frontal and medial-frontal lesions have repeatedly suggested specific deficits in tasks that call for thinking about mental states in others (Channon & Crawford, 2000; Stuss, Gallup, & Alexander, 2001). Both PET and fMRI studies where subjects were asked to make inferences about the mental states of others found activity associated with mentalizing in the medial prefrontal cortex (Frith & Frith, 1999; Gallagher, Happe, & Brunswick, Fletcher, & Frith, 2000). In addition, activity was elicited in the temporo-parietal junction (Gallagher et al., 2000).

There is independent evidence for the developmental vulnerability of this structure from PET scan studies of Romanian adoptees who were deprived of the interpersonal experiences that we think might generate the IIM (Perry, 1997). Independently, we, of course, know that the attachment classification of these adoptees remains disorganized at age three and their social behavior is abnormal at age eight (e.g., Chisolm, 1998). We also have evidence that the mentalizing capacity of individuals maltreated in early childhood continues to have significant limitations (Bogdanski & Cicchetti, 1994). The frontal lobes have a central role to play in social behavior, personality, personal memories, and self awareness (Adolphs, Tranel, Damasio, & Damasio, 1995; Channon & Crawford, 2000; Rogers et al., 1999). Damage to these frontal areas has been consistently associated with social and personality deficits that are consistent with the notion of the loss of interpersonal interpretive capacity, viz. impaired social judgement, impaired pragmatics, deficient self-regulation, and impoverished association of social situations with personal affective markers (e.g., Craik, Moroz, Moscovitch, Stass, Wiper, & Tulving, 1999).

Fourth, Myron Hofer’s work with rodent pups identified regulatory interactions within the mother–infant relationship that have clear analogies to what is proposed here (Polan & Hofer, 1999). Hofer’s work over three decades has revealed that the evolutionary survival value of staying close to and interacting with the mother goes way beyond protection and may be expanded to many pathways available for regulation of the infant’s physiological and behav-
Hofre's view is analogous to ours in that he proposes that the attachment "relationship provides an opportunity for the mother to shape both the developing physiology and the behavior of her offspring through her patterned interactions with her infant" (Polar & Hofre, 1999, p. 177). Attachment is not an end in itself—it is a system adapted by evolution to fulfill key ontogenetic physiological and psychological tasks.

Hofre's reformulation of attachment in terms of regulatory processes, hidden but observable within the parent–infant interaction, provides a very different way of explaining the range of phenomena usually discussed under the heading of attachment. The traditional attachment model is clearly circular. The response to separation is attributed to the disruption of a social bond, the existence of which is inferred from the presence of the separation response. What is lost in "loss" is not the bond but the opportunity to generate a higher order regulatory mechanism: the mechanism for appraisal and reorganization of mental contents. We conceptualize attachment as a process that brings complex mental life into being from a complex and adaptable behavioral system. Some, but by no means all, of such mental function is unique to humans. The mechanisms that generate these (the attachment relationships) have evolutionary continuity across nonhuman species. Just as in rat pups the ontogenetic development of biological regulators crucially depends on the mother–infant unit, so in human development, psychological interpretive capacity evolves in the context of the repetitive interactions with the mother.

Fifth, in a series of studies at the Menninger Clinic we explored the factor structure of a number of self-report measures of adult attachment. These studies cannot be described in detail here, but on both community and clinical samples we found very similar results across three investigations. In the first study (Aller, Hooman, Fultz, Stein, Foggy, & Evans, 2000) two measures of adult attachment style, the Relationship Questionnaire (Bartholomew & Horowitz, 1991) and the Adult Attachment Scale (Collins & Read, 1990), were administered to 235 individuals (99 female trauma patients and 154 community controls). In a principal component analysis a two-component solution accounted for all eigen values greater than 1 and 67.2% of the total variance. The factor space provided a reasonable two dimensional solution with a secure–fearful axis and a dismissive–preoccupied axis. We found the same two factors: a secure–fearful axis and a dismissive–preoccupied axis in a replication study by Stein et al. (2002), which used five adult attachment questionnaire measures, again on a mixed population. When we plotted the subjects in the sample, both patients and community controls, on the same two principal components it was clear that although the secure–fearful axis was excellent at discriminating the community sample from the patient group, the preoccupied–dismissive axis did not distinguish the groups well. What was also clear was a somewhat unexpected relationship between component scores. Although the overall correlation between the two scales was negligible as one might expect, the discrimination between preoccupied and dismissive was somewhat greater towards the middle point of the secure–fearful dimension.

One way of interpreting these data is to assume that security represents an experience of safety in closeness, while fearfulness relates to a disorganization of attachment. Fearfulness appears to be specific to attachment relationships, as nonattachment relationships rarely score highly on this dimension. The dismissing attachment style appears to offer protection to the self by isolation, whereas in enmeshed preoccupation self-protection is perhaps afforded by an amplification of the other, by a denial or subjugation of the self.

We would argue that the safety to fearfulness dimension corresponds to the quality of functioning of the Interpersonal Interpretative Mechanism. At the high end, individuals are well able to represent complex internal states of the other and of the self. With a well-established higher order capacity for distinguishing psychological states of the other and the self, they need no additional strategies for conducting productive interpersonal relationships. When the psy-
chological mechanism crucially underpinning attachment is somewhat weaker (as a function of attachment history or biology), the capacity for sustaining a clear distinction between self and other also becomes weaker. In such a situation the individual will require specific strategies to accommodate to interpersonal encounters. The two prototypical strategies are the avoidant and resistant strategies.

But why are such strategies necessary? Both serve to protect the self in the context of intense interpersonal relations. We assume that these strategies may be necessary because the self, which is as we have seen the product of the other, forever remains vulnerable to social influence. To avoid such instability, against a background of a relatively insecure internal working model, the individual can either deliberately withdraw and enhance the self-representation relative to the other representations (dismissing), or protectively overemphasize and exaggerate the other representation (preoccupied). In either case, the strategies in representational terms are about deliberately separating the other from the self-representation.

Neither of these strategies is inherently pathological, although both signal a certain degree of weakness. At the extreme end of the safety to fearfulness dimension, there can be no strategy because the attachment system is not there to sustain a consistent set of defenses. In these cases, the lack of the interpretive mechanism that sustains social relations functions so poorly that the capacity to arrive at representations of the motivational or epistemic mind states of the other independent of those of the self are profoundly compromised. This is attachment disorganization, or rather, the absence of the mental function that sustains attachment. Thus, we conceive of attachment disorganization as at the opposite end of attachment security and as an indicator of the regular failure of the interpersonal interpretive mechanism.

The concept of empathy has been variously defined but in this context the availability of a mechanism that allows one individual to take another’s perspective and infer to some degree experience their emotional state of mind is central. We believe that a psychological mechanism, which we have labeled the IIM-a, is responsible for this. Other workers have independently come to very similar theoretical conclusions (e.g., Corcoran, 2000). Darwin considered sympathy to be the core moral emotion as it involved the automatic experiencing of the other’s distress, which gives rise to altruistic attempts to offer comfort or relief. Some severely antisocial psychopathic individuals lack an adult moral sense, hence the term “moral imbecility.” These individuals cannot feel a sense of right and wrong, guilt or remorse, because the neural mechanisms (IIM-a) underlying these experiences are impaired and these individuals cannot feel empathy or sympathy (Blair, Jones, Clark, & Smith, 1997). We have for a number of years maintained that the ability to predict and experience the emotions of others is mediated by the functional connection of cognitive representations of other minds in the medial prefrontal cortex (IIM-c) with emotional control centers (IIM-a) in the OFC and temporal lobe. We argue that these functional connections are impaired in individuals with antisocial personality disorder, some of whom have no IIM-a function (causing feel empathy at all) but most of whom may simply have poor connections between the IIM-a and IIM-c systems. The purely cognitive task of identifying belief states in the other has been associated with activation of medial prefrontal foci around BAs using SPECT (Baron-Cohen, Ring, Moriarty, Schmitz, Costa, & Ell, 1994, PET (Blair et al., 1997), and fMRI (Gallagher et al., 2000). Innate emotional responses to facially expressed emotion, particularly fear or sad expressions, underpin the acquisition of empathy and morality, or the moral emotions (Blair et al., 1997). Extreme ASPD (high psychopathy) individuals do not respond autonomically to distressed faces. They appear to be insensitive to the distinction between moral transgressions that result in harm to someone and conventional transgressions that are socially disruptive but do not harm an individual. Responsiveness to distress is necessary for the acquisition of this distinction. This might be lost at at least two levels. Some have suggested that this response is ensured by the amygdala.
(AM) and the impairment of this structure accounts for the limited mortality of the ASPD individuals (Blair, Morris, Frith, Perrett, & Dolan, 1999). Others have suggested that early damage to the OFC is associated with severely impaired and immature moral judgment and evidence of OFC lesions are associated with loss of autonomic responsiveness to social stimuli (Anderson, Richara, Damasio, Tranel, & Damasio, 1999).

The infant’s understanding of their own emotional responses (a precursor of empathy by definition) itself arises out of a complex process of early mirroring between caregiver and child, which leads to the creation of a second-order representation of the child’s emotional state (Gergely & Watson, 1996, 1999). I suggest that some individuals with limited capacity for empathy will show limited AM responses to depiction of distress in children’s faces, while others will show no AM deficit but demonstrate limited OFC activation. Those with AM impairment are likely to have failed to acquire a proper understanding of their own emotional responses that directly led to a failure of empathy. Those with insensitive caregivers (perhaps disorganized early attachments) would not have acquired the second order representation of emotional states and this level of incapacity manifests in OFC dysfunction. A third level of failure may occur as a consequence of a disconnection between the medial frontal and the orbital frontal areas (HIM-a and HIM-c), whereby the individual might experience affect in relation to another’s distress but this is inappropriately or inadequately linked to a representation of the belief and intention state of that person. This latter pattern of functional disconnection might be most likely to arise as a function of severe social adversity at later stages of development.

To summarize, at least five strands of converging evidence suggest that a key selective advantage of attachment might be the development of an understanding of internal states: (1) secure attachment is associated with favorable outcomes across a wide range of relevant tasks; (2) secure attachment predicts precocious performance in tasks specifically calling for symbolic capacity; (3) the class of early attachment classification has less predictive weight than whether attachment experiences occurred; (4) attachment has been demonstrated to have other ontogenetic biological functions in mammalian species that have analogies or that may parallel the evolutionary function for attachment proposed here; (5) the factor structure of adult attachment scales separates out a factor to do with type of attachment (perhaps the Internal Working Model) and the quality of attachment (perhaps the Interpersonal Interpretive Mechanism).

CONCLUSIONS

Although what is proposed here may sound radical, when closely scrutinized it actually contains sadly very little that is new. I suggest that infant attachment functions, in part at least, to facilitate the development of an interpersonal interpretative capacity. The quality of the early relationship plays a major role in determining the robustness of that capacity but attachment security per se is less relevant to later development. The interpretative capacity, in turn, has a key role in the processing of social experience. The level of functioning of the IMM will be reflected in an individual’s ability to function in close interpersonal relationships without needing to have recourse to strategies for amplifying the distinction between self and other representations. The unfolding of disturbance over time is conditioned by the interpretative capacity—we speculate that the expression of pathogenic genotypes is made more likely by the poor functioning of a mechanism designed to differentiate the psychological state of self and other.

This is a function of immense importance, as the labile move from genotype to phenotype is conditioned this way. A full understanding of the interaction between individual mentalized representations of life experience and the expression of genetic predispositions is the task of the developmental psychopathology of the next decades. Eric Kandel (Kandel, 1999)
cites Francois Jacob, who wrote in *Of Fliers, Mice and Men* (1998): “The century that is ending has been preoccupied with nucleic acids and proteins. The next one will concentrate on memory and desire. Will it be able to answer the questions they pose?”

REFERENCES


