PERSECUTORY DELUSIONS: A REVIEW AND THEORETICAL INTEGRATION

Richard P. Bentall and Rhiannon Corcoran
University of Manchester, UK

Robert Howard and Nigel Blackwood
University of London, UK

Peter Kinderman
University of Liverpool, UK

ABSTRACT. Persecutory (paranoid) delusions are a frequently observed clinical phenomenon. In recent years, an increasing volume of research has attempted to explain these types of beliefs in terms of psychological mechanisms. Theories have emphasized early experience, perceptual abnormalities, motivational factors, and information-processing deficits. In this article we review relevant findings, including our own studies of the role of causal attributions and theory of mind deficits. We propose a new integrative model that builds on this work. The core of the model is an account of the way that causal attributions influence self-representations, which in turn influence future attributions: the attribution–self-representation cycle. We argue that biases in this cycle cause negative events to be attributed to external agents and hence contribute to the building of a paranoid world view. These abnormalities are amenable to investigation by functional neuroimaging, and recent studies have implicated specific areas of neuroactivation. However, these findings do not necessarily suggest that paranoid delusions are entirely biological in origin, and there is evidence that adverse early experience may play a role in determining the development of a cognitive vulnerability to paranoid thinking. © 2001 Elsevier Science Ltd. All rights reserved.

KEY WORDS. Psychosis, Persecutory delusions
INTRODUCTION

PERSECUTORY IDEAS, VARYING from vaguely held conceptions through to full-blown delusions, are seen both in psychiatric patients and in people who do not meet standard criteria for psychiatric disorder. They have received more attention than other kinds of abnormal beliefs, partly because they captured the interest of the early psychoanalysts (Freud, 1911/1950), and because they are very commonly observed in clinical practice. Garety, Everitt, and Hemsley (1988) suggested a simple five-way classification of delusional themes and, in a survey of 55 British psychiatric patients, found that persecutory beliefs were most common (35.2%), followed closely by abnormally negative beliefs about the self (31.5%) and abnormally positive beliefs about the self (25.9%). Less common were negative and positive delusions about the world (5.5% and 1.8%, respectively). In a similar survey of first admissions to a psychiatric hospital in Denmark, Jorgensen and Jensen (1994) found that 37 of 88 deluded patients had persecutory beliefs. These observations probably have some cross-cultural validity; in a survey of psychiatric in-patients from Europe, the Caribbean, India, Africa, the Middle East, and the Far East, Ndetei and Vadher (1984) reported that persecutory delusions were the most common type observed except in the Far East, where a small sample of patients more often reported sexual delusions. More recently, Stompe et al. (1999) reported a comparison between the delusions of schizophrenia patients in Austria and Pakistan, again finding persecutory delusions to be the most common in both countries.

In this article we review recent evidence on the psychological processes that lead to persecutory delusions and propose a theoretical model that attempts to integrate this evidence and that builds on our previous work. By way of a preface to the main theme of this article, we define in advance the criteria that we believe must be met by any adequate psychological model and allow the reader judge the merits of our proposals and compare them to other accounts.

What Must an Adequate Theory of Persecutory Delusions Achieve?

Models of the psychological processes involved in psychopathological phenomena have become commonplace, and it is not unusual to find several competing models of the same disorder or symptom. They vary from vaguely articulated accounts that amount to little more than the rewriting of folk psychology in psychological jargon, to well-researched theories that have evolved as they have confronted new data. We propose that psychological models should ideally have the following properties.

First, the phenomenon to be explained should be clearly described and defined, and be reliably measurable. Second, the psychological processes invoked in order to explain the phenomena should be clearly defined and reliably measurable. Third, research should support the hypothesized causal relationships between psychological constructs and psychopathological phenomena. Psychological models are often portrayed as boxes, representing constructs, connected by arrows depicting causal relationships. As arrows indicate functional relationships, they amount to experimentally testable theoretical statements. Fourth, empirical tests of models should involve behavioral and physiological data, as well as questionnaire and self-report measures. As the ability of individuals to report on their psychological processes may be subject to error, tests that involve other kinds of measures can only strengthen confidence in the validity of a model. Fifth, psychological models
should explain how phenomena change over time. Change is the hallmark of psychopathology, and symptoms rarely stay constant. Finally, adequate models will explain the etiology of psychopathological phenomena, by reference to either biological or environmental variables. If psychological processes (for example, cognitive and emotional abnormalities) are assumed to be the immediate precursors of symptoms, the origins of these abnormalities must be explained.

We do not claim that the integrative model of persecutory delusions that we propose later in this article adequately meets all of these criteria. In fact, we doubt whether any psychological model of psychiatric disorder presently does so. However, they provide a benchmark against which the proposals outlined in this article can be evaluated and compared with alternative accounts.

**Defining the Phenomenon**

The psychiatric literature contains numerous descriptions of abnormal beliefs that have a persecutory theme. Perhaps the most famous is the case of Daniel Schreber (1903/1955), a German judge who wrote an autobiographical account of his psychotic illness. Schreber’s beliefs are not easily summarized, but began with the conviction that his physician, Flechsig, was transforming him into a woman in order to submit him to sexual abuse. They later evolved into a complex theological system, in which “rays” from God, attracted by the highly excited nerves in his brain, were believed to bring about a series of “miracles” that would feminize his body. These included the “miracles of heat and cold” during which Schreber believed that blood was being pushed into different parts of his body, and “the head compressing machine” into which he believed that he was being forced by “little devils.”

Autobiographical accounts by modern patients record beliefs that, although perhaps less bizarre, similarly appear to defy rational analysis. For example David Zelt (1981), a young psychology student, has written a vivid account of his discovery that he was a Messiah, and his subsequent fear that the CIA wished to prosecute him for treason. Peter Chadwick (1992), another psychologist who suffered a paranoid illness, first became ill when he developed the conviction that he could achieve a transcendental “borderline” state of exceptional creativity. On discovering that he was unable to attain such a state, he felt tormented by supernatural forces that provoked him to attempt suicide. “Alison,” another patient described in Chadwick’s book, developed an infatuation for a doctor who rejected her overtures, believed that the doctor was following her around, and later became convinced that strangers on the street were trying to kill her. It is clear from these descriptions that persecutory delusions are variable in content and are often the end point of a complex series of difficulties. The imagined persecutors may be individuals known to the patient, ethnic or political groups such as Catholics or communists, or organizations such as the Freemasons or (as in Zelt’s case) intelligence services.

**Defining Delusions**

The difficulty of defining when a belief is delusional has taxed the minds of many psychopathologists. Although the concept of delusion has a long history (Berrios, 1991), modern attempts to clarify this distinction date from the work of Karl Jaspers (1913/1963). Jaspers observed that the beliefs of psychiatric patients are often held with extraordinary conviction, tend to be impervious to experiences or counter-
arguments, and often have bizarre or impossible content. However, he held that true delusions, in contrast to delusion-like ideas, are ununderstandable in the technical sense that they are inaccessible to empathy and therefore impossible to explain in terms of the patient’s background or personality. Echoing this distinction, Kurt Schneider, a later German psychopathologist, held that psychotic symptoms—including delusions—must be defined in terms of their form rather than content:

Diagnosis looks for the ‘How’ (form) not the ‘What?’ (the theme or content). When I find thought withdrawal, then this is important to me as a mode of inner experience and as a diagnostic hint, but it is not of diagnostic significance whether it is the devil, the girlfriend or a political leader who withdraws the thoughts. Wherever one focuses on such contents, diagnostics recedes; one sees then only the biographical aspects or the existence open to interpretation. (Quoted in Hoenig, 1982)

Perhaps because of the difficulties in operationalizing the criterion of ununderstandability, modern definitions of delusion have emphasized Jasper’s first three more general criteria. For example, in DSM-IV a delusion is defined as:

A false personal belief based on incorrect inference about external reality that is firmly sustained in spite of what almost everyone else believes and in spite of what usually constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not one ordinarily accepted by other members of the person’s culture or subculture (e.g., it is not an article of religious faith). (American Psychiatric Association [APA], 1994, p. 765)

Definitions of this sort have been criticized because it is not obvious what constitutes an “incorrect inference” or “incontrovertible and obvious proof or evidence” (Harper, 1992; Heise, 1988). However, as Strauss (1969) has pointed out, much of the confusion about the distinction between delusions and ordinary beliefs can be resolved by locating them along a continuum. Consistent with this suggestion, phenomenological studies show that delusional beliefs, like ordinary beliefs and attitudes, vary across a number of dimensions such as their bizarreness, the conviction with which they are held, the extent to which the patient is preoccupied by them, and the extent to which they lead to distress (Garety et al., 1988; Garety & Hemsley, 1987; Kendler, Glazer, & Morgenstern, 1983). Over short periods of time, the conviction with which delusional beliefs are held may vary considerably (Brett-Jones, Garety, & Hemsley, 1987; Garety, 1985), so that beliefs that are held to be absolutely true on one day may be described as possibly true only on the next. In the long term, however, the delusions of psychotic patients often persist for many years (Harrow, MacDonald, Sands, & Silverstein, 1995; Harrow, Rattenbury, & Stoll, 1988).

The dimensional approach to delusions implies that they might be found, perhaps in a less severe form, in people who have not sought or received psychiatric treatment. In fact, there have been few attempts to assess the prevalence of abnormal beliefs in the general population, the exception being investigations that have focused on magical and supernatural forces, which have generally found these beliefs to be remarkably common (Gallup & Newport, 1991; Ross & Joshi, 1992).

Van Os, Hanssen, Bijl, and Ravelli (2000) have reported a population survey of psychotic symptoms carried out in Holland. Seven thousand seventy-six men were questioned by trained interviewers using a standardized schedule, and participants showing evidence of psychosis were then further interviewed by psychiatrists. When abnormal beliefs secondary to drug taking or organic pathology were excluded, 3.3% were found to have “true” delusions and 8.7% had delusions that were not clinically relevant (that is, they were not associated with distress and did not require treat-
Poulton et al. (2000) found comparable results in a survey of 761 residents of Dunedin, New Zealand, born between April 1, 1972 and March 31, 1973, who were assessed between 3 and 26 years of age in the hope of discovering predictors of psychosis. In this sample, 20.1% were recorded as having delusions by the end point of the study, and 12.6% were judged to be paranoid.

The only study we are aware of in which the relative prevalence of different kinds of delusional beliefs was assessed in the general population employed the Peters Delusions Inventory (Peters, Joseph, & Garety, 1999), a self-report measure (Verdoux et al., 1998). Thirty-one family doctors in the Aquitaine region of Southwest France approached 1,053 patients (of whom only 11.5% had attended their doctor because of psychiatric difficulties) and asked them to complete the questionnaire. Approximately three-quarters responded. The most common ideas reported were that people were not who they seemed to be (69.3% of those with no history of psychiatric disorder); that the individual had experienced telepathic communication (46.9%); that seemingly innocuous events had double meanings (42.2%); that the individual was being persecuted in some way (25.5%); and that occult forces were at work (23.4%). Although the use of a questionnaire measure clearly limits the inferences that can be made from this study, the findings confirm that persecutory beliefs are quite widely held.

**When are Delusions Persecutory?**

Attempts to classify delusions according to their form have not been entirely successful. Although the DSM system places some emphasis on the difference between bizarre and non-bizarre delusions, Flaum, Arndt, and Andreasen (1991) found that the inter-rater reliability of this distinction was less than 0.40. Similarly, Spitzer, First, Kendler, and Stein (1993) calculated the reliability of three alternative definitions of bizarre delusions when 12 clinicians evaluated the beliefs of a large group of patients, reporting kappas ranging between 0.45 and 0.65. Classification according to content, although rarely researched, may prove more reliable: Forgus and DeWolfe (1974) concluded that inter-judge reliability for categorizing delusional themes ranged from .73 to .93.

Kraepelin (1899/1990) identified six subtypes of delusions defined according to their content: ideas of sin, ideas of persecution, ideas of influence (delusions in which people believe that they are being influenced or controlled), exalted ideas (grandiose delusions), sexual ideas, and ideas of reference. Wing, Cooper, and Sartorius (1974) proposed detailed criteria for many of these subtypes, describing a persecutory delusion as follows:

The subject believes that someone, or some organisation, or some force or power, is trying to harm him (sic) in some way: to damage his reputation, to cause him bodily injury, to drive him mad or to bring about his death. (Wing et al., 1974, p. 170)

It is sometimes difficult to distinguish between persecutory delusions and delusions of guilt or sin, in which the individual believes that he or she has committed some kind of terrible offence, so that punishment is either imminent or at least warranted. This kind of delusion is commonly seen in patients with a primary diagnosis of depression, and is often associated with the idea that others have extremely negative or threatening attitudes toward the patient (Kuhs, 1991; Thakur, Hays, Krishnan, & Rangar, 1999). Trower and Chadwick (1995) have made the interesting proposal that persecutory delusions can be subdivided into two main types: “bad me” delusions,
which persecution is believed to be deserved, and “poor me” delusions, in which the individual feels the innocent victim of unjustified persecution. The evidence for the validity of this distinction is currently weak, and it is possible that many patients who would be classified by Trower and Chadwick as suffering from bad me paranoia are in fact severely depressed and suffering from delusional guilt. However, we have more to say on this issue when we propose our integrative account of paranoid delusions.

Freeman and Garety (2000) have recently proposed operational criteria for classifying delusion as persecutory, arguing that the imagined perpetrator must clearly intend to cause harm (the belief that the Earth is about to be struck by a meteorite would not qualify on this basis). Noting that different researchers have varied in the significance they have attributed to the target of the perpetrator’s intentions (whether to the individual or also to the individual’s friends and acquaintances), the timing of the persecution (past, present or future) and the severity of the harm intended, they also suggest that, for the purposes of research, a delusion should only be considered persecutory if the individual believes that harm is directed at him or herself, and is ongoing or anticipated in the future. In the light of the difficulty of distinguishing between persecutory delusions and delusional guilt, we suggest that a further criterion is also important: the intended harm should be believed to be out of all proportion to any sins or offences committed by the patient in the past.

**Diagnostic Issues**

Delusions have been associated with a variety of psychiatric diagnoses. Because diagnostic systems have changed as psychopathology has evolved as a science, there have been many opportunities for confusion. For example, the term paranoia was used by the Ancient Greeks to mean crazy or mad, and was reintroduced to describe a type of delusional disorder by Heinroth in the first half of the 19th century (Lewis, 1970). Reflecting this new usage, Kraepelin classified the non-affective psychoses into dementia praecox, catatonia and dementia paranoides in his early writings, but later collapsed these into the single category of dementia praecox (Kraepelin, 1899/1990). However, in the 5th edition of his famous textbook and thereafter he argued that paranoia could be used as a distinct diagnosis in cases in which delusions were the only prominent symptom.

In the middle of the last century, Adolf Meyer’s (1951) enthusiasm for psychodynamic theory led to a loss of interest in psychiatric classification among American psychiatrists, who often used the term paranoid to describe persecutory beliefs that were not necessarily delusional. Consequently, the term was incorporated into ordinary language where it now refers to “A tendency to suspect or distrust others or to believe oneself unfairly used” (Brown, 1993, p. 2097). Although the publication of DSM-III (APA, 1980) reflected a return to categorical diagnoses as favored by Kraepelin (Klerman, 1978), this tension between a narrow and broad definition of paranoia remains evident. In the current fourth edition of the DSM (APA, 1994) delusions are given prominence as a symptom of schizophrenia, and the term delusional disorder is used instead of paranoia to describe a syndrome characterized exclusively by delusions, usually of a persecutory nature. However, paranoid personality disorder is listed as an axis-2 diagnosis, and is characterized as an enduring pattern of cognition, affectivity, interpersonal functioning and impulse control in which the individual exhibits a “pervasive distrust and suspiciousness of others such that their motives are interpreted as malevolent” (p. 633). In the rest of this article,
therefore, we often use the term paranoid either as an alternative to persecutory, or as shorthand to indicate individuals who suffer from persecutory ideas.

Positive symptoms of psychosis are commonly observed in people who have a primary diagnosis of depression or bipolar disorder. The most severely ill patients in these groups are especially likely to experience delusions (Goodwin & Jamison, 1990; Kuhs, 1991; Lattuada, Serretti, Cusin, Gasperini, & Smeraldi, 1999). As we have already noted, although these often have a persecutory theme, the anticipated persecution is often seen as a consequence of believed misdeeds. Observations of patients with a combination of psychotic and affective symptoms led Kasanin (1933) to propose the diagnosis of “schizoaffective disorder” to describe conditions in which there seemed to be a mixture of schizophrenic and affective symptoms, a proposal that has been embraced by the DSM system. Other theorists, including contemporaries of Kraepelin (Berrios & Beer, 1995), have used the same observations to argue for an einheitspsychose (unitary psychosis) concept, a proposal that still has champions today (Brockington, 1992; Crow, 1991). Taking a different approach, Zigler and Glick (1988) have proposed redrawing Kraepelin’s famous distinction between the schizophrenic and affective disorders to group paranoid syndromes with the latter. They have argued that, in both depression and paranoid conditions in contrast to the other psychoses, pre-morbid social competence is typically good (Glick, Zigler, & Zigler, 1985; Lewine, Watt, Pretky, & Fryer, 1980; Zigler & Glick, 1986; Zigler & Levine, 1981), and symptoms reflect excessive salience of the self.

Partly because of these persisting disputes and the poor demonstrable validity of all categorical systems of diagnosis, some authors have suggested that research into psychopathology should focus on specific types of behaviors and experiences, or symptoms to use conventional psychiatric terminology (Bentall, 1990, 1998; Costello, 1992; Frith, 1992; Persons, 1986). While the arguments pertaining to this debate are largely beyond the scope of this article (see Mojtabai and Rieder, 1998, for a dissenting view), our specific focus on persecutory delusions is consistent with this general strategy. A major strength of this approach is that the interpretation of findings does not depend on assumptions about how symptoms cluster together, or about which of various systems of psychiatric classification is most scientifically plausible.

Psychological Research on Persecutory Ideation

Psychological investigators studying persecutory delusions have tended to focus on two main types of processes (Winters & Neale, 1983). Some have argued that abnormal beliefs are the consequence of fundamental perceptual or reasoning deficits which cause the individual to misunderstand what is happening in the world, whereas others have held that they are motivated beliefs—beliefs that, despite their apparent bizarreness, serve some intra-psychic function for the individual.

Theories of the first type have usually focused on delusions in general, rather than on any one particular type of delusion. A classic example was von Domarus’ (1944) proposal that schizophrenic patients are fundamentally illogical. Modern examples include Maher’s (1974, 1988, 1992) proposal that delusions reflect rational attempts to explain anomalous experiences, and Hemsley and Garety’s (1986) suggestion that delusions arise through defective Bayesian reasoning. Both of these last accounts suggest that deluded patients reach conclusions about the world in much the same way that scientists reach theoretical beliefs, but assume that the processes leading from evidence to deduction are somehow dysfunctional in the psychotic patient.
In contrast, most motivational theories have focused specifically on persecutory beliefs, perhaps because these kinds of delusions appear to reflect some underlying insecurity about the self in relation to others. The prototypical example is Freud’s (1911/1950) celebrated analysis of the autobiography of Daniel Schreber, in which he maintained that the judge harbored homosexual feelings for his father, a well-known physician, which he had then displaced on to his psychiatrist, Flechsig. According to Freud, in a further series of defensive maneuvers, Schreber’s love for his psychiatrist was then transformed into Flechsig’s imagined hatred of Schreber, and then eventually into complex delusions about God’s bizarre manipulations of Schreber’s body. Although later writers inspired by psychoanalysis have continued to assume that persecutory ideas serve a defensive function, their accounts have usually been rather less exotic. For example, Colby, Weber, and Hilf (1971) created a simple computer simulation of persecutory delusions, which assumed that paranoid individuals are highly sensitive to threats to their self-esteem, but protect themselves from feelings of personal inadequacy by blaming disappointments on malevolent others. Of course, these two types of theories are not mutually exclusive. It is possible that both psychological deficits and abnormal motivations contribute to delusions, and we argue that this is indeed the case for persecutory beliefs.

One limitation of many of these models is that they have not been based on an understanding of the processes involved in ordinary belief acquisition. Yet, one implication of a dimensional model of delusions is that such processes may well play a role in the beliefs expressed by psychiatric patients, no matter how strange. Bentall (1990) proposed a simple heuristic model that attempted to highlight the stages that might be important when beliefs are acquired (see Figure 1), and which therefore provides a framework for considering psychological research on delusions. The model assumes that events have to be perceived and attended to if they are to elicit the generation of a belief (events that are not noticed cannot alter an individual’s convictions). Once an individual has noticed an event or a series of events, various kinds of inferential strategies may be employed to determine their meaning and significance. Finally, the search for further information may lead to data that confirm a belief, refute it, or require its modification.

It is important to note that this heuristic model is not a model of delusional ideation; it is a way of organizing data that has been collected in the attempt to

![Figure 1. A Simple Heuristic Model of the Processes Involved in Acquiring and Maintaining a Belief (from Bentall, 1990).](image-url)
explain delusions. In principle, abnormalities at each of the four stages might contribute to the formation and maintenance of unusual beliefs.

**Paranoia-Inducing Events**

When Henry Kissinger chastised the Israeli Prime Minister Golda Meir for being “paranoid” about the Arabs during peace talks in 1973, she retorted that “Even paranoids have enemies” (Schneider, 1998). Unfortunately, the proposition that persecutory beliefs may be the consequence of genuine persecutory experiences has rarely been considered by researchers. This is probably because obtaining objective accounts of the life experiences of deluded individuals is fraught with difficulty; although paranoid patients may describe events that imply a genuine history of persecution, the nature of the phenomenon undermines researcher’s confidence that these accounts can be treated as veridical.

One of the most compelling attempts to explain persecutory beliefs in terms of actual persecution experiences is Niederland’s (1959, 1960) reinterpretation of the Schreber case, based on historical data about the judge’s family (see also Schatzman, 1973, for further elaborations of this reinterpretation). Neiderland’s historical research revealed that Schreber’s father had unusual views about child-rearing, which he proselytized in a series of publications. Believing that children should learn to hold rigid postures when sitting, walking or even sleeping, he devised a series of braces to force them to adopt the desired positions. For example, one brace was designed to make children sit upright while eating. Bolted to the table in front of the child, it consisted of an iron bar that extended up to the child’s face, which it gripped by means of a leather strap under the chin. Although later research by Lothane (1992) has suggested a more positive view of the Schreber family than that painted by Neiderland, it seems likely that the younger Schreber spent much of his childhood restrained by these kinds of devices. Moreover, the bizarre ideas he developed later in his life seem to reflect these experiences. For example, a delusion about a “chest compression miracle” seemed to relate to his father’s practice of securing the child to his bed by means of a specially designed strap that ran across the child’s chest. Of course, even if Neiderland’s account is correct, psychological processes need to be appealed to in order to explain why Schreber’s experiences led to beliefs that were so obviously bizarre.

It would be wrong to generalize from a single case study to an overarching theory of paranoid ideation without considering other types of data. Although few serious attempts have been made to obtain objective evidence that a history of persecutory experiences can lead to delusions, at least three findings are consistent with this hypothesis. First, taking a sociological approach, Mirowsky and Ross (1983), surveyed residents of El Paso in the United States and Juarez in Mexico in order to explore relationships between different types of psychopathology and social circumstances. They argued that experiences of victimization and powerlessness, associated with low socioeconomic status and a lack of educational opportunities, would foster feelings of external control and mistrust which, in turn, would engender paranoid thinking. Constructing a path analysis between these variables as measured in their sample by questionnaire, they were able to show that external control, mistrust and persecutory beliefs were connected to socioeconomic status and educational attainment in the manner they had expected.
The second relevant finding has emerged from the literature on life events and psychopathology. Harris (1987) argued that the events most likely to lead to psychotic and especially paranoid symptoms can be characterized as intrusive. Such events involve someone not a close relative according to Harris’ definition—imposing demands or experiences on the patient. Examples would include threats from landlords, police inquiries, burglaries and unwanted sexual propositions. In a sample of psychotic patients living in the community, Harris found a twenty-fold difference in the number of these events in a comparison of patients who were about to relapse and ordinary people. This finding was replicated in some but not all of the centers participating in the WHO study of the Determinants of Outcome of Severe Mental Disorders (Day et al., 1987), although the ratings of intrusiveness in this study were based on only limited information about the context in which the life events occurred. More recently, Fuchs (1999) collected biographical data from elderly patients with diagnoses of late paraphrenia (now sometimes known as very late-onset schizophrenia-like psychosis (cf. Howard, Rabins, Seeman, & Jeste, 2000) and depression, finding a significantly higher frequency of reports of discriminating, humiliating and threatening experiences in earlier life in the paranoid group.

Finally, recently emerging evidence of an increased vulnerability to positive symptoms, especially paranoid delusions, in migrant groups is also consistent with the hypothesis that paranoia is associated with actual persecution experiences. This association first became apparent with the discovery that Afro-Caribbeans living in Britain have an unexpectedly high chance of being diagnosed as suffering from paranoid schizophrenia (Carpenter & Brockington, 1980; Hemsi, 1971). Although it has been suggested that this phenomenon reflects the culturally insensitive diagnostic practices of white, middle-class British psychiatrists (Littlewood & Lipsedge, 1989), careful epidemiological studies (Bhugra et al., 1997; Harrison, Owens, Holton, Neilson, & Boot, 1988) and attempts to verify diagnoses by consulting Afro-Caribbean psychiatrists (Hickling, McKenzie, Mullen, & Murray, 1999) have confirmed that the increase in risk is a real one. As rates of psychotic illness among Afro-Caribbeans living in the Caribbean have been found to be lower than the corresponding rates for Afro-Caribbeans living in Britain (Bhugra et al., 1996; Bhugra, Leff, Mallett, & Mahy, 1999) the obvious inference is that living in an alien and sometimes racially intolerant culture increases the likelihood that an individual will experience paranoid ideation. Interestingly, this effect does not seem to be confined to Britain; Selten et al. (2001) have recently reported an increased risk of psychotic illness among immigrant groups living in the Netherlands.

**Perceptual and Attentional Factors**

The idea that delusions might be the product of rational attempts to make sense of anomalous perceptual experiences was considered by Kraepelin, but has been more recently developed by Brendan Maher (1974, 1988, 1992). Maher’s theory consists of two separable hypotheses. The first, to be considered here, is that delusions are always attempts to explain some kind of unusual experience. (Note that, according to this model, experiences do not have to be the consequence of perceptual deficits, they just have to seem anomalous to the individual.) The second, which is considered in detail later, is that delusions are never the product of abnormal reasoning processes. The second hypothesis does not follow inevitably from the first, although Maher sometimes seems to suggest that it does.
Whether this theory is sufficient to account for all delusional ideas seems doubtful. In the only systematic attempt to address the relationship between delusions and anomalous perceptions, Chapman and Chapman (1988) interviewed students who scored high on their questionnaire measures of schizotypy. They found that some expressed delusional ideas but reported no anomalous experiences, and that others reported anomalous experiences in the absence of delusional ideas. In only a few cases did there appear to be an obvious causal connection between anomalous experiences and unusual beliefs.

Nonetheless, several lines of evidence suggest that perceptual and attentional abnormalities may contribute to persecutory delusions. For example, it has been argued that persecutory ideas might follow the slow onset of deafness if individuals attribute their inability to hear others to conspiratorial whispering. Consistent with this hypothesis, Zimbardo, Andersen, and Kabat (1981) found that hypnotically induced deafness led to paranoid ideation in normal individuals. Unfortunately, although early studies of clinical samples seemed to show the expected association between deafness and the onset of paranoid symptoms late in life (Cooper & Curry, 1976; Kay, Cooper, Garside, & Roth, 1976), these findings were not replicated by later, more carefully conducted investigations (Moore, 1981; Watt, 1985).

Two studies have suggested that enhanced perceptual skills may play a role in the genesis of persecutory convictions. LaRusso (1978) persuaded a group of normal volunteers to allow her to film their expressions as she gave them brief electric shocks, and also as they pretended to receive electric shocks. Showing the films to paranoid patients and to another group of normal volunteers, she found that the paranoid patients were better able to tell the differences between the genuine emotional responses and the pretend responses. Similar findings have recently been reported by Davis and Gibson (2000) who studied schizophrenia patients’ ability to recognize genuine and posed facial expressions. Whereas most of their patients were poor at recognizing posed expressions, their paranoid patients were better at recognizing genuine expressions of surprise and negative emotion than ordinary people. These observations raise the intriguing possibility that one factor that may contribute to paranoid thinking is patients’ ability to tell when other people are being disingenuous.

Ullmann and Krasner’s (1969) hypothesis that selective attention to threat-related stimuli maintains paranoid ideas can be regarded as a variant of the anomalous perception theory. In a test of this account, Bentall and Kaney (1989) administered an emotional Stroop test to paranoid, depressed, and normal participants. In this kind of test, words are printed in a variety of ink colors and participants are asked to ignore the meaning of the words and name the colors; slow color-naming, indicative of strong response competition between reading and color-naming, is evidence of selective attention to the words. Bentall and Kaney used four types of stimuli: threat-related words (for example, “deceit,” “follow”), depression-related words (for example, “defeat,” “failure”), neutral words (for example, “bud,” “recipe”) and meaningless strings of Os. The paranoid patients were unique in showing slowed color-naming for the threat-related words in comparison with the neutral words.

This finding was replicated by Fear, Sharp, and Healy (1996) in a group of delusional disorder patients with paranoid ideation; in this study anxiety-related words were also included but no effect specific to patients with paranoia was found with these. In an interesting variant of this approach, Leafhead, Young, and Szulecka (1996) studied a woman who was suffering from the Cotard delusion (she believed that she was dead; see Young and Leafhead, 1996). The patient also suffered from the
delusion that her brother and mother were impostors. Testing their patient on three occasions using words relating to death (for example, “coffin,” “dead”), depression-related words, words relating to duplicates (for example, “copy,” “double”), paranoia-related words, anxiety-related words and emotionally neutral words, they found that an abnormal Stroop effect was only present when the Stroop words were congruent with currently held delusions, and was no longer present after the delusions had remitted.

If people attend excessively to information relating to their delusions they should also recall this kind of information particularly easily. Two studies have reported evidence that this is indeed the case. Kaney, Wolfenden, Dewey, and Bentall (1992) asked paranoid, depressed and normal participants to recall stories that contained threat-related and neutral information; the paranoid patients selectively recalled propositions from the stories that were threat related. In a later study, Bentall, Kaney, and Bowen-Jones (1995) administered a free-recall task with threat-related, depression-related and neutral words to similar groups of participants. The paranoid patients selectively recalled the threat-related words. Nonetheless, paranoid patients show poor recall of autobiographical information (Kaney & Bentall, 1999), a phenomenon also found in depressed people (Williams & Broadbent, 1986), especially those who have been sexually abused (Kuyken & Brewin, 1995), and also in patients suffering from post-traumatic stress disorder (McNally, Lasko, Macklin, & Pitman, 1995).

A further consequence of selective attention to threat-related information should be high estimates of the frequency of negative events; this follows because individuals judge frequency according to the availability of relevant information in memory, a phenomenon known as the availability heuristic (Tversky & Kahneman, 1974). This prediction was tested by Kaney, Bowen-Jones, Dewey, and Bentall (1997), who asked depressed-paranoid patients, non-depressed-paranoid patients, depressed patients, and normal controls to estimate the frequency of selected negative, neutral and positive events in the past and future for both themselves and others. All three clinical groups gave high estimates for negative events affecting themselves. However both groups of paranoid patients, in contrast to the depressed patients, also gave high estimates of the frequency with which negative events are experienced by other people; in other words, they showed a high self–other consensus effect for negative events, indicating a paranoid world view in which everyone is seen as vulnerable to threat.

It seems unlikely that these attentional and memory effects can be accounted for by a tendency for deluded patients to spend excessive time studying stimuli relating to their delusions. In a series of studies, Phillips and colleagues (Freeman, Garety, & Phillips, 2000; Phillips & David, 1997a, 1997b) measured the eye movements of patients as they looked at pictures, some of which showed people making threatening gestures. Patients with paranoid delusions, compared with patients with other symptoms, spent less time looking at the threatening parts of the pictures. Phillips and colleagues argued that paranoid patients are quick to identify stimuli in their environment that could be threatening, and then move rapidly to checking whether threats can be found elsewhere.

In summary, it seems that a variety of perceptual and attentional factors may contribute toward the development or maintenance of persecutory ideation. Patients with persecutory delusions attend selectively to threat-related information, are excessively sensitive to others’ expressions of negative emotion, preferentially recall threat-related information, and give high estimates of the frequency with which those
events occur to themselves and others. However, they spend less time looking at threat-related information than ordinary people.

**Inferential Processes**

The specific content of persecutory beliefs suggests that processes involved in social inference may play an important role in their development. In this review, we focus on two types of social inference processes: those involved in inferring the causes of social interactions, and those involved in understanding the intentional states of other people.

Attributions. Individuals readily infer causes of events that are salient to them; Zullow, Oettingen, Peterson, and Seligman, (1988) have estimated that an attribution (a proposition containing or implying the word “because”) can be found in every few hundred words of recorded speech. The role of these kinds of propositions in psychopathology has been most systematically explored in the context of depression, following Abramson, Seligman, and Teasdale’s (1978) proposal that depressed people make excessively internal, stable and global attributions for negative events. A very substantial research program has broadly confirmed this hypothesis (Abramson, Metalsky, & Alloy, 1989; Robins & Hayes, 1995; Sweeney, Anderson, & Bailey, 1986) and has yielded some evidence that an abnormal attributional style in euthymic individuals confers vulnerability to future depression (Alloy et al., 1999).

In a series of publications some of the present authors have argued that paranoid ideation may be a product of abnormal attributional processes (Bentall, 1994; Bentall & Kinderman, 1998, 1999; Bentall, Kinderman, & Kaney, 1994). Early evidence in favor of this hypothesis was available from a study of psychiatric patients’ locus of control, a construct that is more general than attributional style but which bears some similarity to it. Rosenbaum and Hadari (1985) used Levenson’s (1974) multidimensional locus of control scale, which has separate scales for internality, chance factors and the influence of powerful others, and demonstrated that paranoid patients excessively believe that the course of life is influenced by powerful others. This finding has been replicated in several later studies (Kaney & Bentall, 1989; Lasar, 1997).

The first study to demonstrate an abnormal attributional style in paranoid patients was reported by Kaney and Bentall (1989), who used Peterson et al.’s (1982) Attributional Style Questionnaire (ASQ) to compare schizophrenic patients with persecutory delusions, depressed patients and normal participants. This measure requires participants to generate attributions for hypothetical positive and negative events and then to self-rate their attributional statements on three bipolar scales: internality (anchor points: the cause was totally due to other people and circumstances versus totally due to me), stability (anchor points: the cause will never be present versus will always be present) and globalness (anchor points: the cause influences just this particular area versus influences all areas of life). Kaney and Bentall reported that paranoid patients, in common with depressed patients, made excessively stable and global attributions for negative events. However, they excessively attributed positive events to internal causes and negative events to external causes, a tendency that appeared to be an exaggeration of the self-serving bias observed in ordinary people (Campbell & Sedikides, 1999). Candido and Romney (1990) compared paranoid, depressed-paranoid, and depressed patients on the ASQ, finding results comparable to Kaney and Bentall. In this study depressed-paranoid patients, like paranoid patients, made excessively external attributions for negative
events but their internality scores for positive events fell between those of the paranoid patients and the depressed patients. Fear et al. (1996) replicated Kaney and Bentall’s findings with paranoid patients diagnosed as suffering from delusional disorder. In a more recent study, Won and Lee (1997) reported similar attributional abnormalities in patients with persecutory delusions living in South Korea.

Several studies have attempted to find evidence of an abnormal attributional style in deluded patients using measures other than the ASQ. Kaney and Bentall (1992) asked paranoid, depressed, and normal participants to judge the extent to which their own skill contributed to the outcomes of computer games that had been rigged to yield either negative or positive outcomes; the paranoid patients again showed an excessive self-serving bias, that is, they tended to attribute good outcomes to their skill and poor outcomes to chance factors.

In their study of frequency judgements (see above), Kaney et al. (1997) later tested a counter-intuitive implication of paranoid patients’ excessively external attributions for negative events, which they derived from attribution theory. Ordinary people often believe that others have beliefs and experiences similar to themselves, a phenomenon known as the “false consensus effect.” It has been argued that this effect often arises as a consequence of making external attributions (Marks & Miller, 1987); for example, an individual failing an unfair exam should expect others to fail it also. Kaney et al. therefore predicted that paranoid patients would demonstrate a consensus effect for negative events, a prediction that was born out by their data.

In an attempt to compensate for some of the psychometric shortcomings of the ASQ (see below), Kinderman and Bentall (1996a) developed the Internal Personal and Situational Attributions Questionnaire (IPSAQ) to assess three attributional loci highlighted in Levenson’s locus of control typology: internal causes, external-situational causes (in which the event is said to be mainly determined by circumstances), and external-personal causes (in which the event is said to be mainly determined by the actions of another person). The sub-scales of this questionnaire were found to have superior reliability than the ASQ. In a subsequent study, patients with paranoid delusions were shown to have an excessive tendency to make external-personal (other blaming) as opposed to external-situational attributions for negative events (Kinderman & Bentall, 1997).

Craig, Hatton, and Bentall (2001) recently explored yet another approach to assessing attributions in a study comparing acutely ill paranoid patients, Asperger’s syndrome patients and normal controls. Asperger’s patients were believed to be a particularly interesting comparison group because this disorder seems to reflect a disability in inferring the intentions of other people (see below). Rather than use a questionnaire, Craig et al. verbally presented their participants with attributional scenarios and then recorded their verbal responses, codifying the attributions obtained this way using a modified version of the Leeds Attributional Coding System (Stratton et al., 1986). The findings broadly replicated those obtained in previous research, so that the deluded participants made more external-personal attributions for the hypothetical negative events.

In contrast to those studies that have compared patients with paranoid delusions and non-psychotic patients, several studies have examined attributions within different groups of psychotic patients. Sharp, Fear, and Healy (1997) compared patients with different delusional systems, finding that excessively external, stable and global attributions for negative events were specific to those with persecutory or grandiose beliefs. Other studies have found less clear-cut evidence that attributional abnor-
malities are specific to paranoid ideation. Silverman and Peterson (1993) found that paranoid schizophrenia patients, in comparison with non-paranoid schizophrenia patients, made more internal attributions for positive events. Kristev, Jackson, and Maude (1999) examined a group of patients who had recently recovered from their first episode of psychosis using a variety of attribution measures; although many of the expected differences between paranoid and non-paranoid patients failed to emerge, suspiciousness was significantly predicted by external attributions for negative events on an ASQ-like measure. Finally, Martin and Penn (in press) conducted comparisons of schizophrenic patients with and without persecutory delusions and normal participants on the ASQ and the IPSAQ. Although many of the expected differences that emerged from the comparisons were not statistically significant, when independent judges assessed the participants’ attributional statements, the paranoid patients were found to make more external-personal attributions than the controls, and a linear association was observed between clinical ratings of suspiciousness and a tendency to attribute positive events to internal causes.

Two important questions about the interpretation of the attributional data are addressed here before proceeding. First, the measurement of attributional processes remains fraught with difficulty. The ASQ in particular has been criticized for its poor reliability (Reivich, 1995; Tennen & Herzenberger, 1985), and doubts have been cast about the ability of ordinary people to understand and respond appropriately to questions about internality (White, 1991). In an attempt to explore attributional judgments more fully, Kinderman, Kaney, Morley, and Bentall (1992) asked independent judges to rate paranoid, depressed, and normal individual’s attributional statements on the ASQ, and found that they were often discrepant from the participants’ self-ratings. In particular, paranoid patients tended to self-rate as external attributions for negative events that were rated by the judges as internal, a result which was not replicated by Fear et al. (1996). Unfortunately, it is not obvious which type of rating — by the individual who makes the attributional statement or by an independent judge — is most meaningful, as self-ratings may reflect self-presentation biases and independent ratings may be adversely affected by the failure to take into account background information known but not articulated by the participant. Discrepancies between the two types of ratings could be interpreted as evidence that attributional judgements often require two stages, an initial stage in which a cause is inferred, and a second stage in which a meta-judgement is made about the cause (an issue we consider in more detail later). Resolution of these issues requires further research into the psychological processes involved in generating and classifying attributions.

Second, it might be asked whether the attributional abnormalities of paranoid patients are evident only when they try to account for their own experiences rather than the experiences of others. Bentall, Kaney, and Dewey (1991) asked paranoid, depressed, and normal individuals to make attributions about vignettes describing positive and negative interactions between third parties. In this study, paranoid patients made excessive personal attributions for negative interactions (that is, they tended to attribute the negative behavior of an actor toward a target to the actor, rather than to the target or circumstances). However, using a variation of this procedure, Young and Bentall (1997a) failed to find the same effect. More recently, Lee (2001) carried out two experiments, one comparing students scoring high and low on a paranoia measure, and one comparing paranoid patients, depressed patients, and normal individuals, in which attributions were assessed for situations in which the self was target and situations for which another was target. No
differences were observed in either study when another was the target, but the paranoid participants in both studies showed excessive self-serving biases when the self was the target. The implication of this finding is that abnormal attributional responses in paranoid patients are only triggered when there is a perceived threat to the self.

Overall, the findings support the claim that patients with persecutory delusions make abnormal attributions, the most common observations being an exaggeration of the self-serving bias, and/or a tendency to attribute negative events to powerful others. However, differences between persecuted and non-persecuted psychotic patients have generally been less robust than differences between persecuted patients and controls with other diagnoses. Interestingly, correlational analyses within psychotic groups seem to yield better evidence of a relationship between persecutory ideation and attributions than group comparisons, perhaps reflecting the fact that such ideation is common within this population but varies in severity.

Mentalizing or “Theory of Mind.” The second type of inferential abnormality that is considered here concerns the ability to understand others’ mental states. Following Premack and Woodruff’s (1978) seminal work on this ability in primates, individuals who can infer the beliefs, attitudes and intentions of other people, or “mentalize,” are said to possess a “theory of mind” (ToM). Interest in ToM was sparked by the discovery that autistic children (Baron-Cohen, Leslie, & Frith, 1985) and people with Asperger’s syndrome (Happé, 1991) have deficits in this domain, which appear to account for the problems they experience in exercising imagination, communicating and forming relationships with others (Happé, 1994). Because ToM skills may be deficient in people whose IQ is average or above average, and because people with Down’s syndrome perform reasonably well on theory of mind tests (Baron-Cohen, Leslie, & Frith, 1986), many researchers have argued that they reflect the functioning of a modular cognitive system which, in normal children, only becomes operational during the fourth year of life (Baron-Cohen, 1995). However, this account remains controversial. Some investigators have claimed that ToM skills are the consequence of the development of more general executive functions (Ozonoff, Pennington, & Rogers, 1991). Still others have argued for a social developmental view of ToM, claiming that these skills can be detected earlier in life when naturalistic tasks are employed, and emerge in the context of interactions with caregivers (Boyés, Giorano, & Pool, 1997). Consistent with this last account, there is evidence that the development of ToM skills is delayed in individuals who are insecurely attached to their parents (Fonagy, Redfern, & Charman, 1997; Meins, 1997).

Frith (1994, 1992) has argued that a dysfunctional ToM may be implicated in psychotic symptoms including persecutory delusions. He argued that paranoid patients have a normally functioning ToM when asymptomatic but that, during acute illness, impairments to ToM lead to the belief that others have malevolent intentions. According to this theory, when ill, paranoid patients still know that other people are guided by internal mental states but, as they are no longer able to accurately infer what others’ beliefs and intentions actually are, they assume that others are hiding their intentions and, therefore, that their intentions are not good.

Corcoran, Frith and their colleagues attempted to test this theory by conducting a series of studies in which ToM skills were assessed in groups of schizophrenia patients whose symptoms were classified using the Present State Examination (Wing et al., 1974), and who were grouped according to a hierarchical model of symptoms which placed patients with third-person hallucinations and those with persecutory delusions
together. In their first study, using a “hinting” task which explored the ability to infer intentions behind veiled speech acts, Corcoran, Mercer, and Frith (1995) found that patients in this paranoid subgroup were poorer at inferring intentions than normal adults and psychiatric controls suffering from anxiety and/or depression; schizophrenia patients whose symptoms were in remission performed the task as well as the normal controls.

Frith and Corcoran (1996) subsequently presented participants with brief stories accompanied by line drawings to test their ability to understand situations in which individuals hold false beliefs or intend to deceive. In this study, one question following each story required simple memory for the story whereas another required the listener to infer the mental state of one of the characters. In order to ensure that inability to infer intentions was not simply the consequence of poor comprehension or memory, the ToM questions were only scored if the memory questions were answered correctly. The main finding was that schizophrenia patients with persecutory delusions, compared to normal controls and psychiatric controls, were impaired on the ToM questions. The findings therefore supported the idea that people with paranoid delusions have a mentalizing deficit that is not attributable to general intellectual dysfunction or to logical memory impairment. Again it was noted that patients in remission performed well on the tasks.

A further study conducted by Corcoran and Frith (1996) explored knowledge of Grice’s (1975) maxims — conversational rules that insure that acts of communication are as meaningful as possible. They contrasted adherence to the maxims with the ability to use subtle contextual information in order to judge when politeness and tact are called for. The findings indicated that people with persecutory delusions had difficulties only when they attempted to answer the questions assessing politeness and tact. Again the patients in remission performed normally. This was interpreted as evidence that paranoid patients, although conversant with social conventions, had a specific mentalizing deficit that was a consequence of their inability to recognize and act on the contextual cues that normally guide behavior in social settings.

In their final study, Corcoran, Cahill, and Frith (1997) investigated mentalizing skills using a task that drew on both visual and linguistic information processing. Two sets of 10 humorous cartoons were shown to participants. In order to “get” the joke in one set, the participant needed to infer the mental state of one of the characters. Appreciation of the other set of jokes required only understanding of the physical or behavioral events portrayed. While schizophrenia patients in remission understood both types of joke equally well, and with the same level of sophistication as the control participants, those with persecutory delusions clearly found the ToM jokes more difficult.

Studies reported since the completion of this series have supported the view that ToM is impaired in psychotic patients, but not that there is a specific association between ToM deficits and persecutory delusions. For example, in their picture sequencing studies of ToM, Sarfati, Nadel, Chavalier, and Widlocher (1997) found strong associations only between ToM dysfunction and formal thought disorder, while Langdon et al. (1997), also using a picture sequencing task, found evidence of ToM problems in groups of patients with negative features of alogia, social withdrawal and flattened affect, but not in people with persecutory delusions. Doody, Gotz, Johnstone, Frith, and Cunningham-Owens (1998) found ToM difficulties in patients with both negative and positive symptom clusters defined using the Positive and Negative Symptom Scale (Kay, Opler, & Fiszbein, 1986). Using a series of ToM tasks, Drury, Robinson, and Birchwood (1998) also found evidence of deficits in currently
ill patients suffering from multiple positive and negative symptoms but no difference between patients with persecutory delusions and those without. Murphy (1998) similarly found evidence of ToM impairments in schizophrenia patients recruited from a secure psychiatric hospital compared with controls diagnosed as suffering from personality disorder, but patients with persecutory delusions tended to perform better than patients with behavioral signs (negative symptoms, incongruent affect and incoherent speech). It is possible that these differences from the results obtained by Frith and Corcoran reflect different approaches to selecting patients for study. Pickup’s (1997) doctoral research used the same grouping methodology as the original studies and did find ToM difficulties in patients with persecutory delusions, although there was some evidence that this related to their current level of IQ. However, by carefully examining a range of patients, Walston, Blennerhassett, and Charlton (2000) were able to identify a small group who had encapsulated persecutory delusions (a specific group of persecutors were implicated) but who had no other symptoms, and demonstrated that they scored highly on ToM tests.

More recent studies conducted by the present authors have found evidence that ToM deficits can be found in acutely ill patients with persecutory delusions and, to a lesser degree, in patients whose delusions were in remission (Randall, Day, Corcoran, & Bentall, 2001), and that the deficits in acute patients are as severe as those seen in Asperger’s patients (Craig et al., 2001). However, these studies did not employ controls with non-persecutory psychotic symptoms, and therefore could not test the hypothesis of a specific link between ToM and paranoia; we review these studies in detail later. What is clear from the mixed research findings is that any ToM difficulty associated with persecutory delusions must be subtle in nature and is not easily demonstrated or characterized by the experimental methods currently available. A possibility, which we consider in detail later, is that these kinds of deficits, which may not be specific to patients with persecutory delusions, may nonetheless indirectly influence persecutory beliefs via mechanisms not considered in Frith’s original theory.

As with the attributional data, inconsistencies in the evidence may reflect our relative ignorance of the psychological processes involved in these kinds of judgements. For example, some researchers have argued that adults make mental state inferences by means of an automatic modularized system (Leslie, 1993), a hypothesis sometimes known as the theory–theory, whereas others have argued that these inferences are made by running on-line cognitive simulations of others’ cognitive processes (Harris, 1990), a hypothesis sometimes known as simulation theory. More recently, Corcoran (2000) has argued that individuals make mental state inferences by referring to autobiographical memories of their own mental processes in similar circumstances, and using analogical reasoning to assume similar processes in those they are observing. On this view, impoverished ToM in paranoid patients may be linked to biased retrieval of events from autobiographical memory. In fact, as we have already seen, Kaney and Bentall (1999) have found impaired autobiographical memory in patients with persecutory delusions. Of course, it is possible that all of these strategies are utilized at different times by ordinary people, and that different ToM tests may be sensitive to the different strategies. For example “theory–theory” might best account for judgements based on information held in a ToM semantic store, which might contain hard and fast rules like the Gricean maxims. Novel mentalizing situations, on the other hand, for which no directly relevant information is present in the ToM store, may require individuals to simulate or to reason by
analogy. It is therefore likely that any progress in this area will derive from a more detailed analysis of the sub-components of ToM.

**Search for Further Evidence**

Research on deluded patients’ ability to test hypotheses and search for evidence pertaining to their theories was stimulated by Hemsley and Garety’s (1986) suggestion that abnormal beliefs might arise from faulty Bayesian reasoning, or an inability to alter the weight of conviction in a hypothesis in the light of new evidence. Pioneering studies of this phenomenon were subsequently reported by Huq, Garety, and Hemsley (1988) and by Garety, Hemsley, and Wessely (1991) using an abstract “beads in jars” task in which participants were shown two jars of beads, one containing 85% of one color and 15% of another, and another in which these ratios were reversed. Participants were then shown a sequence of beads apparently drawn from one of the jars and were required to estimate which one they came from. Although a variety of methods of measuring the participants’ responses were employed, the findings can be summarized as follows. First, when participants were asked to keep drawing beads until they were certain which jar they came from, deluded patients in comparison with non-deluded psychiatric patients and normal controls guessed after fewer trials—a phenomenon Garety and Freeman (1999) refer to as the “jumping-to-conclusions bias.” Second, and more paradoxically, in an experiment in which the sequence of beads first favored one jar and then changed to favor the other, deluded patients more readily changed their hypotheses than controls, a phenomenon Garety and Freeman refer to as the “disconfirmation bias.” This last observation seems counter-intuitive as delusions have often been defined in terms of their imperviousness to counter-argument or disconfirmatory evidence (APA, 1994; Jaspers, 1913/1963).

There have been a number of successful (Dudley, John, Young, & Over, 1997a; Fear & Healy, 1997; Peters, Day, & Garety, 1997) or partial replications (Young & Bentall, 1997) of these findings using the same procedure, although one correlational study failed to find the expected relationship between the jumping-to-conclusions bias and a measure of severity of delusions (Mortimer et al., 1996). The jumping-to-conclusions bias has also been demonstrated with several other tasks, including a game in which participants were allowed up to 20 questions before guessing the nature of a hidden object (John & Dodgson, 1994) and, in a comparison of students scoring high and low on a measure of delusional ideation, a hypothesis-guessing task (Linney, Peters, & Ayton, 1998). Using a somewhat different approach in which participants’ hypotheses on a multi-trial visual discrimination task were tracked as they were exposed to confirmatory and disconfirmatory information, Young and Bentall (1995) found that deluded patients in comparison with controls generated fewer hypotheses and were less inclined to stick to their hypotheses following confirmatory information. Overall, therefore, the evidence that delusions are associated with abnormal reasoning about hypotheses seems quite strong. Although most of these studies have not distinguished between patients according to the content of their delusions, several have focused specifically on patients with persecutory beliefs (Dudley et al., 1997a; Young & Bentall, 1995, 1997b), raising the possibility that this type of abnormality plays an important role in paranoid ideation.

The psychological origins of the jumping-to-conclusions and disconfirmation biases remains unclear. Several studies found that the former is less evident when
participants were asked to make probability estimates on a fixed number of trials, rather than drawing beads until they felt able to make a decision (Dudley, John, Young, & Over, 1997b; Young & Bentall, 1997b). However, the bias seems to reflect neither impulsive decision-making, a general tendency to respond only to the most recently available evidence, or working memory limitations, as several investigations have shown that deluded patients become more cautious if the ratio between the colored beads is made closer to 50:50 (Dudley et al., 1997b; Young & Bentall, 1997b) and results are unaffected by provision of a memory aid (Dudley et al., 1997a). Moreover, when given a choice of different experiments to test hypotheses, patients with persecutory delusions seem to make sensible decisions that do not differ from those of ordinary people (Bentall & Young, 1997). Taken together, these findings are consistent with Garety and Freeman’s (1999) suggestion that the jumping-to-conclusions bias is, in fact, a data-gathering bias. However, it is not clear that the disconfirmation bias can be explained in the same way.

One possibility is that both biases reflect motivational rather than information processing abnormalities. Two studies have shown that both are amplified when paranoid patients are asked to reason about personally salient information (Dudley et al., 1997a; Young & Bentall, 1997), an observation that is clearly consistent with a motivational account. A type of motivation that might be important in this regard is need for closure, which has been defined by Kruglanski as, “the desire for a definite answer on some topic, any answer compared to confusion and ambiguity” (Kruglanski, 1989, p. 14). In a series of studies, Kruglanski and Webster (1996) have shown that individuals vary in their tolerance for ambiguity and their emotional reactions to uncertainty, and that this disposition can be measured by questionnaire.

It is not difficult to see how a high need for closure might explain the jumping-to-conclusions bias, but it is perhaps less obvious how the disconfirmation bias can be accounted for in this way. However, as most researchers will be aware from personal experience, when confronted with evidence that counters a cherished belief, individuals often react by questioning the evidence, a cognitively effortful activity that often increases uncertainty rather than reduces it (evidence, no matter how discredited, can rarely be reinterpreted as entirely consistent with the challenged belief). For this reason, individuals with a high need for closure may prefer to accept a new hypothesis in conditions in which the initial hypothesis is not highly valued or emotionally salient. (If this argument is correct, we would expect people with a high need for closure, like deluded patients, to vigorously resist challenges to belief systems that are highly systematized, and which appear to bring order to an uncertain world.)

A study by Roberts (1991) reported findings that are consistent with these speculations. Roberts asked a group of patients if they would welcome disproof of their delusional beliefs, which were often highly distressing to them. The majority said that they would not, leading him to conclude that their delusional worlds were often preferable to real life because they were more predictable. More recently, Bentall & Swarbrick (2001) assessed need for closure in patients with persecutory delusions, patients whose persecutory delusions were in remission, and also normal controls, using a simplified version of Kruglanski’s questionnaire. Both the currently ill and remitted patients scored highly on this measure, suggesting that a high need for closure may be a trait that causes vulnerability to paranoid ideas. However, as Garety’s beads task was not employed in this study, there was no direct evidence that need for closure determined performance on that task as hypothesized here. There is
clearly a need for further studies to determine the extent to which the jumping-to-conclusions and disconfirmation biases are motivated phenomena.

AN INTEGRATIVE MODEL

The large volume of psychological data from patients with persecutory delusions that has become available over the last decade represents remarkable progress compared to the situation that was apparent when Oltmanns and Maher (1988) conducted an earlier review of the literature on abnormal beliefs. However, these data must be integrated into a unitary framework if the criteria we have set at the outset for an adequate model are to be met. Before introducing our integrative account, we first discuss the shortcomings of our previous attempts to construct such a model.

These attempts have evolved over more than a decade (Bentall, 1990, 1994; Bentall & Kinderman, 1998; Bentall et al., 1994), and have been driven by the attributional data reviewed earlier. Our general approach has been to assume, like Colby et al. (1971) before us, that paranoid patients, in common with patients suffering from depression, have latent negative beliefs about the self that are vulnerable to activation by negative life events. However, in contrast to patients with depression, we have assumed that patients with persecutory delusions attempt to avoid the activation of negative beliefs about the self by attributing threatening events to the actions of another person. This set of assumptions is consistent with data from normal participants that indicate that the self-serving bias is most evident under conditions in which the self is threatened (Campbell & Sedikides, 1999).

Bentall et al. (1994) attempted to describe the mechanism involved in this kind of defensive projection using the framework of Higgins’ self-discrepancy theory (SDT; Higgins, 1987), which assumes that individuals have multiple self-representations, and that negative emotion arises from discrepancies between current beliefs about the self (the actual self, in the terminology of the theory) and standards of self-evaluation (or self-guides in SDT), which include the ideal self and the self as it ought to be. We argued that external attributions for negative events reduce discrepancies between the actual self and ideals but (because they implicate the actions of other people) cause the paranoid individual to believe that others have malevolent intentions, resulting in discrepancies between the actual self and the believed views of others about the self (or self–other representations). This account as it stands has two major shortcomings. The first, which has been discussed in the literature, concerns the self-esteem of paranoid patients. The second, which has been almost neglected until now, concerns the stability of attributional style.

Persecutory Delusions and the Self

Critics of the attributional model, especially Garety and Freeman (1999), have argued that low self-esteem is common in paranoid patients and that this is incompatible with the hypothesis that paranoia serves a defensive function. In their own study of a cohort of psychotic patients taking part in a randomized controlled trial of cognitive-behavior therapy, they found that self-esteem in the majority of their paranoid patients, as measured by Robson’s (1989) questionnaire, was lower than previously reported norms, although a sub-group reported high self-esteem. However, no data from appropriately matched normal participants was reported.
In fact, the evidence on self-esteem in paranoid patients is inconsistent, probably because of differences of opinion about how self-esteem should be measured. Robson (1989), the author of the questionnaire used by Freeman et al. (2000) remarked that, “Self-esteem is an idea rather than an entity and the term signifies different things to different people”; his questionnaire therefore sampled a broad range of self-esteem related constructs, and included items reflecting respondents’ beliefs that they are disliked by others (for example, “Most people would take advantage of me if they could”; “I often worry about what other people are thinking of me”; “I often feel humiliated”). As paranoid patients can be expected to endorse such items it is not surprising that they have been assessed as having low self-esteem on this measure. Bowins and Shugar (1998) similarly reported that deluded patients in general suffered from low self-esteem on a questionnaire measure, and speculated that this was inconsistent with a defensive model of paranoia. However, Candido and Romney (1990) assessed global self-esteem in paranoid, depressed-paranoid, and depressed participants by means of the same scale, reporting high self-esteem compared to norms in the paranoid group, low self-esteem in the depressed group, and intermediate scores in the depressed-paranoid group; like Freeman et al., they did not report data from matched controls. Havner and Izard (1962) found evidence of unrealistic self-enhancement in paranoid patients, and Lyon, Kaney, and Bentall (1994) reported that paranoids scored normally on Rosenberg’s (1965) self-esteem scale, which measures perceived self-worth. Finally, several studies have attempted to assess the relationship between self-consciousness and paranoid ideation, although not with consistent results. Following his earlier distinction between two different types of self-consciousness (Fenigstein, Scheier, & Buss, 1975)—private self-consciousness (defined as awareness of inner aspects of the self, for example thoughts and feelings) and public self-consciousness (defined as awareness that the self is a social object under the scrutiny of others)—Fenigstein and Vanable (1992) reported high levels of self-reported public self-consciousness in normal participants who scored high on a paranoia scale. However, Smari, Stefansson, and Thorgilsson (1994) found that clinical ratings of paranoia from psychotic patients correlated with private but not public self-consciousness.

On the assumption that unidimensional accounts of self-esteem fail to capture the complexity of self-representational abnormalities in patients with persecutory delusions, Kinderman and Bentall (1996b) measured discrepancies between different kinds of self-representations using the framework of self-discrepancy theory. In this study, participants were asked to describe the actual self, their ideal self, and the believed views of their parents about the self. (The parent perspective was chosen because, in pilot work, many patients expressed difficulty reporting the believed views of friends because they had no friends.) Patients with persecutory delusions, like normal participants and in comparison with depressed controls, were found to have actual self-representations that were consistent with their ideals. However, considerable discrepancies were observed between their actual self-representations and the believed views of their parents, an observation that was interpreted as consistent with Bentall et al.’s (1994) proposal that patients with persecutory delusions maintain consistency between their beliefs about themselves and their ideals at the expense of attributing negative beliefs about the self to others.

As Bentall et al. (1994) argued that paranoid patients have latent negative beliefs about the self, studies that show a discrepancy between implicit indices of self-esteem and explicit measures might be thought to be particularly strong evidence for a
defensive model. Kinderman (1994) administered paranoid, depressed, and normal participants a self-concept checklist containing both high self-esteem and low self-esteem words. The normal and paranoid participants endorsed far more high self-esteem words than low self-esteem words, whereas the depressed participants endorsed approximately equal numbers of both. (This finding is clearly consistent with the hypothesis that paranoia is associated with overt high self-esteem.) However, on a subsequent emotional Stroop test, the paranoid patients showed slow color-naming for the high self-esteem words, and were especially slow for the low self-esteem words. Lee (2000) was able to replicate this effect in two experiments, one comparing students who scored high and low on a paranoia scale, and one comparing clinically paranoid, clinically depressed and normal participants. In the first study, the paranoid students showed excessive interference for low self-esteem words. In the second, paranoid patients showed marked interference for both positive and negative trait words.

Lyon et al. (1994) attempted to obtain evidence of covert negative self-representations in paranoid patients by means of the Pragmatic Inference Test (PIT), an implicit attributional style measure disguised as a memory test, which had been developed by Winters and Neale (1985) for this purpose. In this study, it was assumed that implicit attributional judgements would reflect latent self-representations. On a questionnaire measure of attributional style similar to the ASQ, paranoid patients made excessively external attributions for positive events and internal attributions for negative events, as found in previous studies. However, on the PIT, in common with depressed controls they made excessively internal attributions for negative events and external attributions for positive events. Unfortunately, Kristev et al. (1999) and Martin and Penn (in press) were unable to replicate these findings in their studies comparing schizophrenia patients with and without paranoid delusions.

A final approach to assessing latent or implicit self-representations that we must consider here involves the measurement of self-schemas as defined by Beck’s (1976) cognitive model of depression. These are considered to be enduring rules and standards for self-evaluation (a concept that clearly overlaps with Higgins’ concept of a self-guide). Depressed patients typically score high on these measures, which indicate vulnerability to low self-esteem (Gotlib & Hammen, 1992). Fear et al. (1996) reported that delusional disorder patients suffering from persecutory delusions scored highly on the Dysfunctional Attitude Scale (DAS), a widely used self-schema measure. This finding was replicated by Bentall and Kaney (1996); by comparing depressed-paranoid and non-depressed paranoid patients, Bentall and Kaney were able to show that raised DAS scores in the paranoid group were not attributable to depression. In a more recent study, Bentall and Swarbrick (2001) attempted to distinguish between two types of schemas identified by Beck (1983): sociotropy (roughly, the tendency to evaluate self-worth in terms of relationships with others) and autonomy (roughly, the tendency to evaluate self-worth in terms of autonomy and academic success). They predicted that remitted patients would tend to be preoccupied about the attitudes that others held toward them, and would therefore score high on sociotropy, whereas currently deluded patients would tend to be rejecting of others and would therefore score highly on autonomy. Using Robins et al.’s (1994) Personal Style Inventory, which distinguishes between these two types of schemas, data was obtained that was consistent with these predictions. However, differences from normal and depressed controls largely disappeared when depression was included as a covariate.
It would appear that the most consistent findings have emerged from studies that have attempted to assess latent self-representations, which generally suggest that paranoid patients harbor beliefs about themselves that are more negative than those held by ordinary people. However, attempts to measure explicit self-representations have yielded results that are much less consistent, some reporting high self-esteem in paranoid patients and some reporting low self-esteem. Several explanations for these inconsistencies are possible. First, as Trower and Chadwick (1995) have argued, it is possible that there is more than one type of paranoia: “poor me paranoia,” in which persecution is assumed to be unwarranted and malevolent, and “bad me” paranoia, in which persecution is believed to be deserved because of previous misdeeds. Although, Trower and colleagues have so far failed to operationalize this distinction in a way that would allow psychological mechanisms to be investigated, this proposal clearly merits further attention from experimental psychopathologists, who have largely ignored it to date.

A second possible explanation for the inconsistencies in the data on the self and paranoia, which we have already alluded to, concerns different conceptualizations of the self, associated with different measurement approaches. As Robson (1989) points out, these different approaches are often inconsistent and are therefore likely to lead to different conclusions. Within normal samples, even correlations between self-ideal discrepancies and global self-esteem are moderate, as individuals may believe that they fall short of their ideals but nonetheless remain persons of significant worth (Arnold, 1988). For this reason, global measures of attitudes toward the self may be much less informative than measures that distinguish between different types of self-representations. Given the nature of paranoia, measures that confound self-perceptions and the believed perceptions of the self by others are likely to be particularly misleading.

The third and perhaps most important explanation that can be offered to account for inconsistencies in the self-esteem data concerns researchers’ failure to take into account dynamic aspects of the self. There is evidence that fluctuations in global self-esteem may be as important in psychopathology as average level of self-esteem. Kernis and colleagues, for example, have shown that fluctuations in self-esteem are associated with a tendency to make extreme attributions about daily hassles (Greenier, Kernis, & Waschull, 1995) and that individuals with high but highly fluctuating self-esteem have an excessive tendency to attribute negative experiences to external causes (Kernis, 1993). These observations raise the possibility that self-representations in patients with persecutory delusions may be unstable over time, and suggest that these patients might be locked into an intense struggle to maintain positive self-representations that often fails. The high private self-consciousness ratings found in association with clinical paranoia by Smari et al. (1994) are perhaps consistent with this hypothesis.

**Instability of Attributions**

A second problem implicit in most attributional models of psychopathology, including Bentall et al.’s (1994) account of paranoia, concerns the stability of attributional processes. Although the concept of attributional style suggests a stable trait, and although researchers have often treated it as such, there is considerable evidence that attributional judgements vary according to time and circumstances. Most of this evidence has been collected in the context of studies of depression.
First, some longitudinal studies have indicated that the pessimistic attributional style that is believed to play a causal role in depression is more evident during depressed mood than during periods of euthymia (Lewinsohn, Steinmetz, Larson, & Franklin, 1981; Miranda & Persons, 1988; Segal & Ingram, 1994). Second, in a naturalistic study of a psychotherapeutic intervention, detailed analysis of the depressed patient’s verbal responses indicated that internality attributions changed during treatment sessions and that these changes were predictive of changes in mood (Peterson, Luborsky, & Seligman, 1983). Third, in a series of laboratory studies conducted with normal individuals, Forgas, Bower, and Moylan, (1990) found that mood induction by means of contrived success and failure experiences resulted in marked shifts in attributions; negative mood induced in this way resulted in more internal and stable attributions for negative events and external and unstable attributions for positive events whereas positive mood results in exacerbation of the normal self-serving bias toward internal and stable attributions for positive outcomes. Therefore, although it would appear that different groups of people have characteristic attributional styles when tested under controlled circumstances (so that attributional style can be said to be stable in the long term), the attributional responses, at least of depressed patients and normal participants, show substantial instability over the short and medium term.

These findings mirror the data on self-esteem in paranoia considered earlier, in that they suggest the possibility that attributional responses may be highly labile in these patients. Together with the findings on global self-esteem they suggest the need for a more dynamic account of paranoid ideation than proposed in previous explanatory models of this phenomenon.

THE ATTRIBUTION–SELF-REPRESENTATION MODEL

Much of the evidence we have already considered can be accommodated by recognizing that attributions and self-representations are cyclically coupled, so that attributions are partially determined by the availability of self-representations, which in turn influence the availability of self-representations in the future (see Figure 2). It is widely accepted that attributions can affect mood by affecting self-representations, so that pessimistic attributions for negative events reduce self-esteem, thereby inducing feelings of dysphoria. Less widely recognized is the possibility that changes in self-representations that are brought about in this way affect subsequent attributions. In the following discussion we argue that this model of the attribution–self-representation cycle more accurately captures dynamic relationships between attributions, self-representations and symptoms than previous attributional models of psychopathology. We begin by considering some of the processes involved in the generation of attributions. The failure to consider these processes is a major shortcoming of previous attempts to explain the role of attributions in psychopathology.

Of course, events often contain information that points toward one attribution rather than another. However, there is usually some degree of freedom about the kind of attribution that can be made for any given situation. It therefore seems likely that the generation of an attribution involves a process of cognitive search that terminates when a suitable explanatory construct is found. We assume that this search begins with currently available self-representations. If self-representations that match the event to be explained are readily available, an internal attribution will be
generated (for example, a person passing an examination is likely to attribute the cause of her success to herself if she already believes herself to be intelligent; otherwise some other kind of attribution is more likely to be generated). For this reason, people with currently low global self-esteem tend to make internal attributions for negative events (Flett, Pliner, & Blankstein, 1995), whereas people with high self-worth tend to generate internal attributions for successful outcomes (Ickes & Layden, 1978).

If the cognitive search process fails to generate an internal attribution, continued searching may yield explanatory constructs that imply an external cause. However, there is good reason to believe that some kinds of information pertaining to external causes are processed preferentially compared to others. In a series of studies, Gilbert, Pelham, and Krull (1988) found that, in normal individuals, a cognitive load reduces the number of external-situational attributions generated and increases the number of external-personal attributions, in which the cause of an event is attributed to a trait of some agent. The implication of this finding is that external-personal attributions

are the “default option” when attributing events to external causes, and that an
effortful further cognitive search is required to identify situational information that
may modify this verdict. This observation has considerable implications for theories
of paranoia, as patients with persecutory delusions make many external-personal
attributions but few external-situational attributions (Kinderman & Bentall, 1997).

The model of the attribution–self-representation cycle proposes that attributions
affect future self-representations. In common with many other models of psycho-
pathology, for example Beck’s (1976) theory and the various attributional models of
depression (Abramson et al., 1978, 1989), we assume that negative self-repre-
sentations in turn lead to negative mood, especially when they are discrepant with
enduring schemas or self-guides (Higgins, 1987). However, we have previously
proposed that different kinds of attributions will have different effects on the current
availability of self-representations according to the extent to which they are able to
prime stored knowledge about the self (Bentall & Kinderman, 1998; Bentall et al.,
1994). We assume that such stored knowledge is encoded in a variety of ways, for
example as internalized descriptions of the self learned in childhood and also as
autobiographical memories of emotionally salient events. Internal attributions for
negative events will tend to prime stored representations of the self (beliefs and
autobiographical memories) that match the attribution, and hence increase the
discrepancy between beliefs about the self and ideals. For example, a person who
blames him/herself for failing an examination is likely to feel more stupid as a
consequence, but only if she already has some implicit understanding that she is
capable of being stupid. External personal attributions, on the other hand, are not
likely to have this effect but are likely to influence our beliefs about what other people
think about us. Specifically, this kind of attribution for a negative event may lead to
the belief that others have negative attitudes toward the self, and hence to discrep-
ancies between self-representations and the believed views of others about the self
(Kinderman & Bentall, 1997). Finally, external-situational attributions for negative
events, although requiring more cognitive effort than external-personal attributions,
are likely to be psychologically benign, neither priming negative self-representations
nor negative perceptions of others’ attitudes toward the self.

Because, according to this model, attributions affect future self-representations,
which in turn affect future attributions, nonlinear changes in self-esteem and the
kinds of attributions generated are clearly possible over a period of time. However,
whether or not this occurs is likely to depend on the precise functional relationships
between different components of the cycle, and also on circumstantial factors that
affect the specific attributions generated at any point in time.

**Individual Differences in the Functioning of the Attribution–Self-Representation Cycle**

A variety of factors are likely to influence both the generation of attributions and
their impact on self-representations during the attribution–self-representation cycle.
These include: (a) relatively enduring factors, such as individual differences in stored
knowledge about the self, in motivational biases toward selecting some types of
attributions rather than others, in the tendency to selectively attend to specific types
of information, and in the ability to understand the beliefs and motives of others; and
(b) circumstantial changes in the relative availability of information pertaining to
internal, personal or situational causes. Before considering how these processes may
lead to the generation of external-personal attributions for negative events in the case
of the paranoid patient, it is first useful to consider their influence in normal and depressed people.

As we have already noted, normal individuals have a general tendency to make internal attributions for positive events and external attributions for negative events: the self-serving bias (Campbell & Sedikides, 1999). There is also evidence that normal individuals with high self-esteem have predominantly positive beliefs about the self (Campbell, 1990). We believe that these two observations are connected and that normal individuals maintain positive beliefs about the self because of the kinds of attributions they tend to make. On this view, self-representational homeostasis is the consequence of the self-serving bias, which ensures that all but the most serious threats to self-esteem, although provoking some degree of mood disturbance, are followed by a rapid resumption of normal functioning.

There is abundant evidence that the stored knowledge about the self in people who are depressed or vulnerable to depression tends to be more negative than the stored knowledge of ordinary people (Gotlib & Hammen, 1992), and that such patients are more likely to make internal attributions for negative events (Robins & Hayes, 1995). In terms of the attribution–self-representation cycle, depression is therefore associated with an absence of homeostasis, so that negative life events precipitate a spiral of increasingly negative beliefs about the self coupled to increasingly internal attributions for negative events. This account explains why a pessimistic attributional style is difficult to detect in most patients prior to the onset of depression (Robins & Hayes, 1995) although, when present, predicts depression in the future (Alloy et al., 1999).

It seems likely that a number of factors are responsible for the paranoid patient’s tendency to make external-personal attributions for negative events in preference for either internal attributions or external-situational attributions. We have previously assumed that this bias partly reflects a defensive avoidance of making internal attributions (Bentall & Kinderman, 1998; Bentall et al., 1994), which might otherwise lead to the kind of catastrophic escalation of discrepancies between self-representations and ideals found in depressed patients. The evidence on the self-representations of patients with persecutory delusions reviewed earlier is clearly consistent with the hypothesis. As we saw, the most clear-cut findings were obtained from implicit measures, which presumably reflect stored knowledge about the self. On this evidence, paranoid patients harbor knowledge about the self that is as negative as the stored knowledge about the self of depressed people, and which provides a powerful motive for avoiding self-blame when threatened by failure or humiliation.

The most likely explanation of the less consistent results obtained from global self-esteem measures is that current self-representations are highly unstable in paranoid patients, reflecting the varying extent to which priming of this underlying stored knowledge is avoided during successive iterations of the attribution–self-representation cycle. Although no data is currently available on the stability of self-esteem in patients with persecutory delusions, this hypothesis is consistent with (Kernis, 1993) observation that extreme variability in self-esteem is most common in individuals whose average self-esteem is low. This account also helps to explain why persecutory ideation is so often associated with negative mood, as observed by (Zigler & Glick, 1988). It also suggests that Trower and Chadwick’s (1995) two types of paranoia (‘‘poor me’’ and ‘‘bad me’’) might be different manifestations of the same mechanism, the latter occurring when attempts to avoid internal attributions for negative events fail to prevent the individual from being overwhelmed by negative beliefs about the self. Whether most paranoid patients eventually settle into one of
these two types of paranoid thinking, or oscillate between them, remains a matter for further research.

Despite these observations, it is clear that avoidance of self-blame is not a sufficient explanation of the paranoid patient’s abnormal attributions, as external-situational attributions would achieve this without the heavy cost of assuming a paranoid world view. In order to explain the peculiar functioning of the attribution–self-representation cycle in paranoid patients, it is therefore also necessary to explain why patients excessively generate external attributions that implicate the intentions of other people. One possibility is that this is a consequence of the type of attributional information readily available to the paranoid patient.

In fact, we have already seen that there is considerable evidence of attentional biases in paranoid patients, and these biases are certainly likely to make information about the actions of others more available than in ordinary people. Thus, patients with persecutory delusions attend selectively to (Bentall & Kaney, 1989; Fear et al., 1996) and preferentially remember (Bentall et al., 1995; Kaney et al., 1992) threat-related information, and are excessively sensitive to negative emotional expressions (Davis & Gibson, 2000; LaRusso, 1978) despite spending less time than ordinary people physically inspecting threatening stimuli (Phillips & David, 1997a, 1997b). One implication of this account is that provoking attention to others during negative experiences should facilitate paranoid ideation in ordinary people. This effect has been demonstrated by Bodner & Mikulincer (1998) in a series of experiments in which participants were asked to explain their performance following a contrived failure experience. When their attention was focused on the experimenter (for example, by having a video camera obviously pointing at the experimenter and a monitor showing the experimenter’s image in clear view) the participants were especially likely to make paranoid attributions.

A second, but not incompatible, possibility is that the paranoid patient’s excessive tendency to make external-personal attributions results from the failure to process situational information. Recall that Gilbert et al.’s (1988) findings suggest that this requires considerable cognitive effort, so that initial judgements about the causes of the behavior of other people often implicate supposed traits, before being adjusted to take into account situational factors. Failure to make this adjustment might occur because appropriate cognitive resources are not available. However, the kinds of jumping-to-conclusions reasoning biases demonstrated by Garety et al. (1991) and others, or an excessive need for closure (Bentall & Swarbrick, 2001), may also be implicated, leading to premature termination of the attributional search before a situational adjustment can be made.

These observations lead us to reconsider the relationship between theory of mind and paranoid symptoms, which we reviewed earlier in this article. In that review, we found at best equivocal evidence of a specific link between ToM deficits and paranoid symptoms, although psychotic patients as a whole seem to experience difficulty when attempting to infer the intentions of other people. The model of the attribution–self-representation cycle does not suggest any direct link between ToM and paranoia, but it does suggest that ToM may indirectly influence attributions under some circumstances. This will happen when the ability to infer the point-of-view of another person may point toward situational factors, thereby reducing the probability of an external-personal attribution. For example, after being ignored by a friend, our initial response might be to attribute some negative characteristic to the friend (“He’s rude”). However, if we are able to infer something about the friend’s mental state, we may
revise this attribution in favor of a more situational account of his behavior (‘‘He must be upset by the break-up of his marriage’’; ‘‘He’s been under a lot of stress lately’’).

If this hypothesis is correct, poor performance on ToM tasks should be associated with a tendency to make external-personal attributions. This predicted relationship was found in a study by Kinderman, Dunbar, and Bentall (1998), in which student participants were divided into two groups according to their performance on a ToM task. In a more recent study comparing acutely paranoid patients, remitted paranoid patients and normal controls, Randall et al. (2001) found the expected negative correlation between ToM performance and external-personal attributions when the participants’ attributions were classified by independent judges, but not when attributions were classified by the participants themselves. However, contrary to expectation, in a comparison of acutely ill paranoid patients, Asperger’s patients and normal controls, Craig et al. (2001) found no specific association between ToM and external-personal attributions, but a negative correlation between ToM skills and the self-serving bias. They therefore suggested that confusion about the thoughts and intentions of other people may create circumstances in which the paranoid patient tends to respond defensively.

Overall, the findings considered here point to complex relationships between attributions and other kinds of processes implicated in delusions. There seems to be evidence that both excessive attention to the actions of others during negative experiences, and an impaired ability to generate situational attributions, contribute to the abnormal functioning of the attribution–self-representation cycle observed in paranoid patients. These relationships should provide a rich source of hypotheses for future research.

**Tests of the Model**

We have seen that the theory of the attribution–self-representation cycle does not lend itself to precise predictions about self-esteem in paranoid patients. Indeed, we have speculated that some paranoid patients may eventually reach a relatively stable state in which global self-esteem is high, others may reach an equally stable state in which self-esteem is low, and still others may oscillate between periods of high and low self-esteem. Critics of our earlier attributional accounts of paranoia, for example Garety and Freeman (1999), who have assumed that our approach assumes high self-esteem in paranoid patients, would be right to regard this lack of precision as a fundamental weakness of the new model, except for two considerations. First, many psychiatric phenomena including symptoms (Tschacher, 1996; Tschacher, Scheier, & Hashimoto, 1997) and self-esteem (Kernis, 1993) appear to be inherently unpredictable over time. Indeed, as meteorologists and population biologists have long known, unpredictability is often an expected characteristic of complex systems such as that described by our model. Second, we have been able to devise other experimental tests of the model that do not require us to be able to make precise predictions about self-esteem.

The model has implications for the speed with which attributions are generated. Normal individuals, who have almost exclusively positive self-representations available to them, should tend to make internal attributions more rapidly for positive than for negative events, whereas individuals with low global self-esteem, who tend to have a mixture of positive and negative self-representations available, should not show this bias. In a test of this prediction, we measured the time taken to generate attributions by normal participants, patients with persecutory delusions and depressed patients
As expected, response latencies for internal attributions were significantly quicker for positive than for negative events in the normal participants but not in the depressed patients. Patients with persecutory delusions responded similarly to the depressed patients, a finding that is consistent with the hypothesis that they have more negative stored knowledge about the self than normal participants.

In two further experiments carried out with normal participants, we studied dynamic aspects of the model (Kinderman & Bentall, 2000). In both experiments, discrepancies between the actual self and ideals and discrepancies between the actual self and the believed views of others about the self were measured before and after the participants made attributions for negative events. As expected, discrepancies between the actual self and ideals at the outset predicted the types of attributions made, with high discrepancies being associated with internal attributions for negative events. Controlling for this effect, when the participants were categorized according to the kinds of attributional responses they predominately made, it was found that changes in their self-representations were also largely as predicted. External-situationnal attributions were benign, leading to few changes in either kind of discrepancy. However, external-personal attributions lead to a reduction in discrepancies between the actual self and ideals, but an increase in discrepancies between the actual self and the believed views of others about the self. Unexpectedly, internal attributions for negative events led not only to the predicted increase in discrepancies between the actual self and ideals, but also to an unpredicted increase in self–other discrepancies (which was not as large as the increase attributable to external-personal attributions), presumably because participants with very negative self-representations assume that others share their hostile view of themselves.

A final aspect of the model that we have tested concerns the stability of attributions. In contrast to previous attributional theories of psychopathology, the model of the attribution–self-representation cycle predicts that, although group differences in attributional style will be observed under most conditions, in some circumstances it will be possible to detect shifts in the kinds of attributions generated by clinical groups. Specifically, individuals who have stored negative information about the self should make more internal attributions for negative events when those beliefs are primed. Bentall and Kaney (2001) asked paranoid, depressed, and normal participants to make attributions for hypothetical negative events before and after attempting an insoluble anagram task that, it was assumed, would prime feelings of failure and thus temporarily increase the availability of negative self-representations. Although previously reported group differences in attributions were observed at the outset, as predicted, both clinical groups made markedly more internal attributions for negative events following this experience so that, in the paranoid group, their attributions were no longer more external than those of the normal controls. Interestingly, no such changes were detected in the normal control group, presumably because they had fewer negative self-representations that could be primed.

**The Functional Neuroanatomy of the Attribution–Self-Representation System**

A comprehensive model of persecutory delusions would identify the specific neuroanatomical systems that implement the cognitive processes described in the model (Blackwood, Howard, Bentall, & Murray, 2001). Initial clues about these systems are available from studies that have sought to correlate patterns of regional cerebral
blood flow in schizophrenia patients with either the “reality distortion” symptom cluster (delusions and hallucinations) (Ebmeir et al., 1993; Kaplan et al., 1993; Liddle et al., 1992) or “delusions” as a single entity (Sabri et al., 1997). These studies largely concurred in identifying positive correlations between these constructs and blood flow in the left lateral pre-frontal cortex, ventral striatum, superior temporal gyrus and parahippocampal region. However, such “first guess” approaches to localizing functional anatomy are neutral with respect to the cognitive operations that these regions serve.

A more theory-driven approach has been made possible by adapting experimental paradigms originally designed to study cognitive abnormalities in patients. For example, attentional bias has been investigated by asking subjects to determine the potential self-relevance of neutral and threatening statements (Blackwood et al., 2000). Attention to threatening statements deemed relevant to the self activates a network involving the left lateral inferior frontal cortex, ventral striatum and anterior cingulate. These areas have been previously implicated in “deep” levels of semantic processing which would be expected for processing material related to the self (the left lateral inferior frontal cortex; Symons & Johnson, 1997; Wagner et al., 1998), egocentric memory (ventral striatum; White, 1997) and assessments of the motivational content of stimuli (anterior cingulated; Devinsky, Morrell, & Vogt, 1995).

In a substantial program of research, Davidson (2000) has shown that affective style is anatomically represented in the prefrontal cortex, which represents emotion in the absence of immediately present incentives, the amygdala, anterior cingulated and related areas. Left prefrontal activation is associated with positive affect whereas right prefrontal activation is associated with negative affect. Relative left versus right anterior activation has been specifically associated with attributional style but not attitudes toward the self (study by Davidson reported in Gotlib & Abramson, 1999), an observation that is of some significance given the left anterior activations observed in deluded patients.

In an attempt to localize the cognitive mechanisms specifically involved in paranoid thinking, Blackwood et al. (2000) studied attributional bias in normal participants during fMRI scanning by asking them to make attributional decisions about hypothetical events derived from Kinderman and Bentall’s (1996a) IPSAQ. In this study, attribution of events in a manner that was not self-serving (internal attributions of negative events and external attributions for positive events) activated the left pre-central gyrus, an area involved in the executive control task of response inhibition (Bush et al., 1998; Carter, Mintun, Nichols, & Cohen, 1997). One possible explanation for this observation is that, in normal individuals, the self-serving attributional bias is a motivationally pre-potent response and the pre-central gyrus activation may therefore represent inhibition required to produce attributions that are not self-serving.

Theory of mind has been investigated using differing experimental tasks (stories and non-verbal cartoons requiring theory of mind attributions, decisions about object function that require the modeling of another person’s knowledge) that have the common requirement of asking subjects to report on the mental states of others (Brunet, Sarfati, Hardy-Bayle, & Decety, 2000; Fletcher et al., 1995; Gallagher et al., 2000; Goel, Grafman, Sadato, & Hallett, 1995). These studies have identified a so-called “mentalising” network involving the medial pre-frontal cortex, temporoparietal junction and lateral inferior frontal cortex (Frith & Frith, 1999). Such areas appear to be concerned with the representation of mental states of the self (medial prefrontal cortex; Lane, Fink, Chua, & Dolan, 1997), representation of the actions
and goals of self and others (the “mirror neurons” of the lateral inferior frontal cortex; Gallese & Goldman, 1998; Iacoboni et al., 1999), and representation of the behavior of others (Hietanen & Perrett, 1993).

New paradigms are required to investigate the neural networks serving self-representations, in terms of stored information, access to that information and the influence of the store on the encoding and interpretation of “noisy” social data. Functional neuroimaging approaches have so far largely focused on self-face recognition (Keenan, Nelson, O’Connor, & Pascual-Leone, 2001; Kircher et al., 2000), self-evaluation (Craik et al., 1999), and autobiographical memory (Fink et al., 1996) and the inter-relationships between these aspects of the self at the neuroanatomical level have yet to be defined.

THE ETIOLOGY OF PERSECUTORY DELUSIONS

Our model of the attribution–self-representation cycle appears to account for many of the observations about paranoid delusions recorded in the research literature. Moreover, our account of the theory has attempted to address many of the criteria we set for psychological models of psychopathology at the outset. Specifically, in proposing the theory we have attempted to: (a) define the phenomenon of interest; (b) explain the phenomenon by references to psychological processes which, for the most part, can be reliably measured; and (c) review the psychological literature to provide evidence that supports hypothesized functional relationships between components of our model. Although much of the evidence has been based on self-report data, we have: (d) occasionally appealed to other kinds of data such as response times, performance on tests of attention and memory and even data obtained from neuroimaging experiments. We have also shown how, (e) our model can account for changes in symptoms over time, for example the major changes in well-being that may follow significant life events and shorter-term fluctuations in attributions and self-representations. To be sure, none of the criteria we set has been perfectly met by the model as developed so far, but we believe that in each respect it sufficiently approximates these criteria to be worthy of further research.

So far we have hardly considered the final criterion: the need to specify etiological factors that lead to the cognitive abnormalities that we believe are responsible for persecutory delusions. This is a formidable problem, as almost every variable known to influence human behavior has, at one time or another been singled out as a cause of psychotic behavior (Bentall, Jackson, & Pilgrim, 1988). For example, biological investigators have implicated factors as diverse as abnormalities in brain biochemistry (Carlsson, 1995), abnormal neuroanatomy (Woodruff & Lewis, 1996), genetic factors (Torrey, Bowler, Taylor, & Gottesman, 1994), and viruses (Torrey & Yolken, 1995) leading to abnormal neurodevelopment (Keshavan & Murray, 1997). Environmental factors that have been implicated include the stresses of living in urban environments (Faris & Dunham, 1939) or racist societies (Bhurgra, Mallett, & Leff, 1999), intrusive life events (Harris, 1987) and dysfunctional family relationships (Singer & Wynne, 1965). To some extent, this plethora of theories must reflect the heterogeneous samples selected for research by broad diagnostic criteria such as those contained in the DSM. To disentangle these various influences is a project that extends well beyond the present article, a project that in any case would be hampered by a lack of etiological research focused on specific symptoms of psychopathology. In the
remaining section of this article, we attempt no more than to sketch how the theory of the attribution–self-representation cycle might be extended to encompass etiological variables, focusing on genetic and environmental factors.

Is There a Genetic Contribution to Paranoid Thinking?

Although life-span developmental studies of paranoid thinking have yet to be conducted, two observations suggest that the cognitive biases underlying persecutory convictions may be established many years before the appearance of frank delusions. In a high-risk study conducted in Israel, in which the children of psychotic parents were followed to adulthood, it was found that an external locus of control in adolescence was a predictor of future psychotic illness (Frenkel, Kugelmass, Nathan, & Ingraham, 1995). As we have seen, a tendency to attribute negative events to external causes appears to be a critical element of the paranoid style of thinking, and this observation suggests that it may substantially predate the establishment of beliefs that are sufficiently abnormal to attract a psychiatric diagnosis. More recently, in a study in which patients attending a psychiatric service for children were followed-up into adulthood, it was found that abnormal suspiciousness, sensitivity, and relationship difficulties with peers predicted a later diagnosis of schizophrenia (Cannon et al., 2001), suggesting the same conclusion.

One possible interpretation of these findings is they represent the earliest expression of a genetic vulnerability to paranoid ideation. However, some commentators have argued that it may be wrong to expect genetic contributions to specific psychotic symptoms. In their thoughtful critique of the symptom-orientated approach to psychopathology, Mojtabai and Rieder (1998) reviewed studies in which the heritability estimates for broad diagnostic categories have been compared with the apparent heritability of symptoms, concluding that there is more substantial evidence of genetic effects for the former. However, this case for retaining categorical diagnoses as explanatory constructs in psychopathology is not as strong as it first seems. Most genetic researchers have interpreted the results obtained from traditional genetic research strategies, for example, family, twin and adoption studies, as indicating that psychosis is a polygenic phenomenon (Asherson, Mant, & McGuffin, 1995), a conclusion that has been reinforced by the failure of recent linkage studies to yield replicable associations between particular genes and psychotic illness (Crow, 1997; Kidd, 1997; Moldin, 1997). Moreover, the studies cited by Mojtabai & Rieder (1998) were based on patient samples selected by diagnostic criteria requiring multiple symptoms. Assuming that different genes confer vulnerability to different symptoms, which are also influenced by nongenetic factors, individuals selected for multiple symptoms are likely to be carriers of multiple psychosis genes. Because close relatives of these patients will share only some of these genes, heritability estimates for diagnoses are almost bound to appear higher than estimates for each symptom, but the inference that different symptoms share a common genetic origin may not be warranted.

To obtain a clear picture of the heritability of persecutory ideation, it is necessary to turn to genetic studies in which this symptom has been the focus of research. Unfortunately, studies of this kind have rarely been carried out, and the only useful data is available from studies that have focused on the diagnostic categories of delusional disorder and paranoid personality disorder. Perhaps the most important series of studies of this kind have been reported by Kenneth Kendler and colleagues. In an early review, Kendler and Davis (1981) concluded that paranoid symptomato-
logy ran true in families, and that twins who were concordant for a diagnosis of schizophrenia also tended to be concordant for a paranoid versus non-paranoid subtype. In two later studies, Kendler and Hays (1981) and Kendler, Masterson, and Davis (1985) reported that the prevalence of schizophrenia in the relatives of patients with a diagnosis of delusional disorder was significantly lower than in the relatives of schizophrenia patients. In a reanalysis of data collected by Kety in the Danish–American adoption studies, Kendler, Gruenberg, and Strauss (1981) also found that schizophrenia spectrum disorders could be observed in the biological relatives of schizophrenia patients but not in the biological relatives of adoptees who were diagnosed as suffering from delusional disorder.

Kendler’s observation that paranoid psychosis may share little genetic variance with more typically schizophrenic psychoses has been supported by later studies. For example, in an analysis of data from the Copenhagen high-risk study, Jorgensen et al. (1987) examined the psychiatric status of children of psychotic mothers after they had reached a mean age of 24 years, finding that the offspring of mothers diagnosed as paranoid were less likely to be diagnosed as schizophrenic than the children of non-paranoid schizophrenic mothers. Dworkin et al. (1988) found that affected probands in twin pairs who were concordant for schizophrenia tended to have fewer paranoid symptoms than affected probands from discordant pairs. Onstad, Skre, Torgersen, and Kringlaen (1991) similarly observed that a diagnosis of schizophrenia was more likely in the monozygotic twin of a patient diagnosed as suffering from non-paranoid schizophrenia than in the twin of a patient suffering from paranoid schizophrenia. More recent studies of patients with a DSM diagnosis of personality disorder are also relevant in this regard. These studies have generally yielded lower estimates of heritability for the diagnosis of paranoid personality disorder than for other Axis-2 diagnoses (Coolidge, Thede, & Jang, 2001; Dahl, 1993; Torgersen et al., 2000).

Two general conclusions seem to emerge from these findings. First, the heritability of paranoid ideation, relative to the heritability of other psychotic symptoms, appears to be quite low. Second, despite this, there is some evidence that psychological characteristics associated with paranoid ideation run in families. For example, in Kendler’s studies, paranoid personality disorder was more common in the close relatives of delusional disorder patients than in the relatives of controls (Kendler et al., 1985) and these relatives were also likely to suffer from inferiority feelings (Kendler & Hays, 1981). A family study by Schanda et al. (1983) observed that the first-degree relatives of patients with a diagnosis of delusional disorder, but not of those with a diagnosis of schizophrenia, tended to be “suspicious, secretive, jealous, or showed delusions of some kind of paranoid disorder.”

If the current difficulties in replicating the results from the molecular genetic approach can be overcome, it is possible that future linkage studies will yield specific genes that contribute vulnerability to paranoia. Indeed, at least two tentative observations of this kind have already been reported (Gelernter, Kranzler, Satel, & Rao, 1994; Serretti, Macciardi, Catalano, Bellodi, & Smeraldi, 1999). However, this prospect must be qualified in two ways. First, given the existing heritability estimates for paranoid ideation, it is likely that any such genes will make a relatively minor contribution to vulnerability, and will form part of a complex of factors that accumulate in the expression of paranoid ideas. Second, the task of discovering pathways from genes to symptoms is likely to dwarf in complexity the problems of decoding the 30,000 genes that comprise the human genome. Genes code for proteins, not behaviors. Environmental factors play an important role in
determining the expression of genes in terms of particular neuroanatomical structures (Elman et al., 1999), which in turn provide the hardware that implements the cognitive processes implicated in particular psychiatric disorders. Under these circumstances, it may be fruitful to examine linkages between genes and particular cognitive processes rather than between genes and symptoms, although the problem of measuring cognitive processes with the necessary precision and reliability presents obstacles in this regard. Interestingly, a twin study has suggested that there may a genetic contribution to attributional style (Schulman, Keith, & Seligman, 1991).

Environmental Determinants of Paranoid Delusions

The apparent evidence of developmental antecedents of paranoid thinking, together with the observation that suspiciousness and other indices of paranoia run in families, might equally well be interpreted as reflecting environmental effects, especially given the rather limited evidence of a specific genetic contribution.

We have already reviewed evidence that certain environmental conditions are associated with paranoid thinking. For example, we saw that persecutory convictions tend to thrive in social circumstances that facilitate victimization and powerlessness (Mirowsky & Ross, 1983) such as exposure to an alien culture (Bhurgra et al., 1999). We also saw that there is tentative evidence that exposure to intrusive life events facilitates the development of paranoid ideation (Harris, 1987). The common feature of all of these circumstances is that they involve aversive experiences instigated by other people. As events of this sort are likely to direct attention toward those other people and thereby elicit external-personal attributions, their impact on paranoid ideation can be readily understood in the context the theory of the attribution–self-representation cycle. However, it is unlikely that they are sufficient for the emergence of frank paranoid delusions. After all, many people suffer these kinds of experiences without becoming paranoid, and the evidence that suspiciousness and an externalizing attributional bias precedes the onset of psychotic symptoms suggests that a vulnerable cognitive style is also necessary.

One possibility that must seriously be considered is that early experience, and particularly relationships with caregivers, facilitates the development of this kind of cognitive vulnerability. In fact, this idea is hardly new, and underlies much psychoanalytic speculation about the origins of paranoid ideation, for example, the ideas of the early object-relations theorists Melanie Klein and W.R.D. Fairburn (Hughes, 1989). In a program of research conducted by Alfred Heilbrun and colleagues in the 1960s and 1970s, attempts were made to test a theory of this kind in a series of experiments conducted mainly with male college students. Briefly, Heilbrun (1975) argued that adolescent males who are vulnerable to developing paranoid thinking have what he termed an “open cognitive style” characterized by “perceptual vigilance and social approach tactics that are intended to elicit positive evaluative responses and protect...vulnerable self-esteem” and that this style is a response to “unclear and devious maternal communication of control and rejection messages” (Heilbrun & Brosnson, 1975, p. 422). Some observations collected from young males classified as having this cognitive style were judged to broadly support this account. For example, it was reported that they had increased ability to recognize words that they had previously been told came from a maternal source (Heilbrun, 1973), and that, when asked whether negative comments ostensibly collected from a group of
mothers described either themselves or their peers, they excessively attributed the comments to their peers (Heilbrun, 1972).

However, although the assumptions about the cognitive processes involved in paranoid thinking bear some similarities with those outlined in the present article, Heilbrun’s program of research was inconclusive because evidence relevant to theory was not sought from patients suffering from persecutory delusions (crucially, no attempt was made to establish that paranoid patients had actually experienced aversive maternal control); because some predictions from the model were not supported by experimental data (for example, young males with an open cognitive style did not show excessive concern about others’ apparent negative attitudes toward them, and did not account for such attitudes by the excessive use of external attributions; Heilbrun & Brosnson, 1975); and because the theory seemed unable to account for paranoid thinking in females.

On the basis of more recent research, three arguments can be constructed to suggest that family relationships play a role in the development of paranoid ideation. The first involves extrapolating the findings from research with individuals who do not suffer from paranoia. For example, as we have already seen, several studies have reported that the development of theory of mind skills is impaired in normal children who are insecurely attached to their parents (Fonagy et al., 1995; Meins, 1997). Moreover, Seligman et al.’s (1984) observation that attributional style in 11-year-old children correlates with maternal but not paternal attributional style suggests that mothers play a special role in the development of causal reasoning. More importantly, perhaps, substantial evidence that cognitive style is influenced by family relationships has emerged from the program of research into depression conducted by Lyn Abramson, Lauren Alloy and their colleagues. Rose, Abramson, Hodulik, Halberstadt, and Leff (1994) sub-divided a large group of depressed patients into those who had a pessimistic attributional style and those who did not, finding that the former group were more likely to report a history of emotional and sexual abuse. More recently, the Temple-Wisconsin Cognitive Vulnerability to Depression project (Alloy et al., 1999) identified a group of euthymic students who had a pessimistic attributional style and dysfunctional attitudes toward the self, and who showed high rates of depression during a 2-year follow-up period. When Alloy and colleagues obtained information about their family of origin from the students and also from their parents, they found that the high-risk participants had experienced high rates of parental emotional abuse in comparison to the low-risk controls. In a subsequent study, Alloy et al. (2001) conducted attributional assessments of the parents of their high-risk sample to test three social learning mechanisms by which a pessimistic style might be learned: modeling (so that the child copies the parent’s attributional style), negative feedback from parents about the causes of stressful events (so that parents teach their children to blame themselves for negative events), and neglectful or emotionally abusive parenting practices. Although evidence supporting all three mechanisms was obtained, the most robust findings supported the hypothesis that parents teach pessimistic attributions by giving inappropriate feedback following negative events. Overall, the findings from both normal individuals and individuals vulnerable to depression suggest that relationships with caregivers affect the development of the very cognitive domains that appear to be important in paranoia.

The second argument is based on observations of abnormal attachment relationships in paranoid patients. Using the Adult Attachment Interview (which requires participants to recall early relationships with parents), Dozier, Stevenson, Lee, and
Velligan (1991) found that patients with a diagnosis of schizophrenia, in comparison with patients with affective disorders, often had the kind of insecure attachment style that is described as dismissing or avoidant. The dismissing-avoidant person devalues the importance of attachments and often cannot recall specific details about relationships with parents. In a later study, Dozier and Lee (1995) found that patients with this attachment style were especially likely to suffer from delusions, hallucinations and suspiciousness. Perhaps this is unsurprising as the emotional theme underlying the dismissing style is lack of trust. However, Dozier’s findings have been supported by two studies in which romantic attachment style and psychiatric symptoms were measured in large community samples (Cooper, Shaver, & Collins, 1998; Mickelson, Kessler, & Shaver, 1997), which also reported an association between paranoid ideation and insecure attachment. Kinderman and Bentall’s (1996b) data on self-discrepancies in paranoid patients, described earlier, are also consistent with these findings, since the patients reported excessive discrepancies between their positive beliefs about themselves and the negative attitudes that they attributed to their parents.

The final line of argument concerns prospective investigations of the impact of dysfunctional family relations on positive symptoms. Several studies have indicated that disruption of early relationships with parents seems to confer high risk of psychosis. For example, in an analysis of data from the Copenhagen high-risk study, Cannon, Mednick, and Parnas (1990) identified those in the sample who had predominately positive or negative symptoms. In the positive group there was strong evidence of inadequate parenting (six of eight had been separated from their mothers for more than one-and-one-half years during early childhood, one had been separated from her mother for more than 5 years, and one had been physically abused by her father). Similar evidence was absent in the negative symptom group, many of who had evidence of neuropsychological impairment. In a study of 38 British Afro-Caribbean psychotic patients, Bhugra et al. (1997) reported that 12 (34%) had suffered separation from their mothers for a period of 4 years or longer during childhood, and 19 (53%) had suffered a similar period of separation from their fathers. In an analysis of outcome data at 28 years for more than 11,000 children recruited into the Northern Finland 1966 birth cohort study, Myhrman, Rantakallio, Isohanni, Jones, et al. (1996) found that unwantedness (as assessed by questioning the parents prior to birth) led to a four-fold increased risk of later psychosis, even after adjusting for confounding socio-demographic, pregnancy, and perinatal variables.

While each of these arguments on its own is not particularly strong or specific to paranoia, and although the different lines of evidence are open to multiple interpretations, together they make a powerful case for supposing that families play a role in the development of vulnerability to paranoid thinking. Specifically, it seems likely that both a negative attitude toward the self and a tendency to respond to threats to self-esteem by making external attributions are fostered during early interactions with parents. If this hypothesis is seen as scapegoating the parents of people with severe mental illness, it is worth noting that less than ideal parenting is very common, and that parents who fail to provide their growing children with an emotionally nurturing environment are often laboring under their own psychological difficulties. Diamond and Doane (1994) and Paley, Shapiro, and Worrall-Davies (2000) have reported that parents of severely ill patients who were classified as “high expressed emotion” (highly critical and emotionally over-involved) tend to be insecurely attached to their own parents.
Finally, the model of paranoid thinking that we have outlined has many gaps. Future research based on the model must prioritize a number of issues, of which the following seem to be the most important. First, at the psychological level it will be necessary to find ways of investigating longitudinal changes in symptoms and underlying cognitive processes. Mathematical modeling techniques may be useful in determining whether symptoms are likely to settle into the two types of paranoid ideation described by Trower and Chadwick (1995) or whether, on the contrary, rapid fluctuations in the self-esteem of paranoid patients are to be expected over the long term. At the biological level, it will be important to continue to explore the functional neuroanatomy of the cognitive processes underlying paranoia, and to investigate similarities with the functional neuroanatomy of other attributional disorders, especially depression. With regard to etiology, the most fruitful lines of inquiry would seem to concern the impact of early relationships on social-cognitive development. In all likelihood, the model will require revision in the light of the findings generated. However, we believe that it provides a useful framework for integrating existing findings and guiding future research.

REFERENCES


