COGNITIVE NEUROPSYCHIATRY
2007, 12 (6), 495–510

Psychology Press
Taylor & Francis Group

Persecutory delusions and the conditioned avoidance paradigm: Towards an integration of the psychology and biology of paranoia

Michael Moutoussis

Tolworth Hospital, Surbiton, UK

Jonathan Williams

Department of Child & Adolescent Psychiatry, Institute of Psychiatry, London, UK

Peter Dayan

Gatsby Computational Neuroscience Unit, London, UK

Richard P. Bentall

School of Psychology, University of Wales, Bangor, UK

Introduction. Theories of delusions often underplay the role of their content. With respect to persecutory delusions, taking threat as fundamental suggests that models of threat-related, aversive learning, such as the Conditioned Avoidance Response (CAR) task, might offer valid insights into the underlying normal and abnormal processes. In this study, we reappraise the psychological significance of the CAR model of antipsychotic drug action; and we relate this to contemporary psychological theories of paranoia.
Methods. Review and synthesis of literature.
Results. Anticipation and recall of aversive events are abnormally accentuated in paranoia. Safety (avoidance) behaviours may help perpetuate and fix persecutory ideas by preventing their disconfirmation. In addition, patients may explain negative events in a paranoid way instead of making negative self-attributions (i.e., in an attempt to maintain self-esteem). This defensive function only predominates in the overtly psychotic patients. The "safety behaviours" of paranoid patients, their avoidance of negative self-attributions, and the antiparanoid effect of antipsychotic medication all resonate with aspects of the CAR.
Conclusions. The CAR appears to activate some normal psychological and biological processes that are pathologically activated in paranoid psychosis.

Correspondence should be addressed to Michael Moutoussis, Tolworth Hospital, Red Lion Road, Surbiton KT6 7QU, UK. E-mail: michael.moutoussis@swlstg-tr.nhs.uk

http://www.psypress.com/cogneuropsychiatry DOI: 10.1080/13546800701566686

Paranoid psychological defences may be a result of basic aversive learning mechanisms, which are accentuated during acute psychosis.

INTRODUCTION

Following Schneider (1949/1974; cf. Hoenig, 1982), psychiatrists have traditionally distinguished between the form of psychopathology and its content. The phenomenological tradition of the twentieth century therefore focused on the abnormal formal inferences found in delusions, declaring that their content is culturally determined, rather arbitrary and thus of little interest (Berrios, 1991). This position is now considered extreme for two reasons. First, it is difficult to make a clear distinction between ordinary beliefs and delusions either in terms of form or content (David, 1999) and delusions must therefore be seen as lying at the end of a dimension or series of dimensions of belief attributes (Kendler, Glazer, & Morgenstern, 1983); this principle is especially important in the case of paranoid delusions, as subclinical beliefs about persecution appear to be fairly common in nonpsychiatric samples (Freeman et al., 2005). Second, it is now clear that the affective processes associated with the content of abnormal ideas play important roles in their genesis and maintenance (Bentall, 2003; Freeman & Garety, 2005; Raune, Bebbington, Dunn & Kuipers, 2006). In this paper we will argue that the content of persecutory delusions can be explained by brain processes that process threat-related emotional information, and that these processes also account for the fixity of these kinds of beliefs.

The content of abnormal beliefs typically reflects a small range of core themes, such as persecution, grandiosity, and jealousy, that reflect concerns about the individual's place in the social universe (Bentall, 1994). Research in many cultures has consistently found that the most common type of delusion involves the belief that the self is being threatened by malevolent others (Garety & Hemsley, 1987; Jorgensen & Jensen, 1994; Ndetei & Vadher, 1984; Stompe et al., 1999). By way of illustrating this point, Table 1 shows previously unreported symptom data from a cohort of 255 first-episode schizophrenia spectrum patients recruited to the SoCRATES trial of cognitive-behaviour therapy for early psychosis (Tarrier et al., 2004). The symptoms were assessed with the Positive and Negative Syndrome Scale (PANSS; Kay & Opler, 1987) within 14 days of admission. If scores of 3 or more on both subscale P1 (delusions) and subscale P6 (suspiciousness) are taken as evidence of persecutory delusions, it seems that more than 90% of this highly representative sample experienced significant paranoid ideation. It is apparent from these observations that an adequate understanding of delusions must include an account of how the specific content (e.g., persecution, or perception of threat) is related to form (unwarranted derivation, fixity).

TABLE 1
Positive symptoms of patients recruited to the SoCRATES study

Symptom	*N/255 PANSS ≥3*	*% PANSS ≥3*	*Mean score*	*Median score*
Delusions (P1)	250	98.0	5.26	5.0
Suspicion (P6)	235	91.8	4.53	5.0
Delusions and suspicion (P1 and P6)	230	90.2		
Hallucinations (P3)	177	69.1	3.41	4.0
Formal thought disorder (P2)	144	56.5	2.70	3.0
Agitation (P4)	179	70.2	3.03	3.0
Hostility (P7)	97	37.9	2.30	2.0
Grandiosity (P5)	98	38.6	2.25	1.0

Patients were first-episode DSM-III-R diagnosed schizophrenia spectrum patients (schizophrenia, schizophreniform disorder, schizoaffective disorder, delusional disorder, or psychotic disorder not otherwise specified; total $N = 255$). SoCRATES is the Study of Cognitive Realignment Therapy in Early Schizophrenia (Tarrier et al., 2004). Patients were recruited from over 26 months from 11 mental health units serving three geographically defined English catchment areas: Liverpool, Manchester and Salford, and North Nottinghamshire). Assessments were conducted by trained psychiatrists within 14 days of admission using the Positive and Negative Syndrome Schedule. Symptoms are ranked in order of frequency and mean severity.

Since threat is key, we must study the perception of real and delusional unpleasant events, and the responses of both healthy control participants and of patients in the face of these perceptions. This approach implies that models of aversive processing in healthy humans, and indeed animals, might capture important aspects of paranoia. The psychological and neurobiological aspects particularly of animal models are experimentally and theoretically quite tractable. They therefore offer considerable opportunities to identify important components of the elementary normal and abnormal processes that might be involved in paranoid thinking.

In the following sections, we first highlight the central role played by threat and aversion in persecutory delusions. We then describe the Conditioned Avoidance Response (CAR) from both psychological and neural perspectives. Finally, we discuss three aspects of the CAR as a model of persecutory delusions, and then indicate directions for future research.

THREAT PERCEPTION AND THE ATTRIBUTIONAL MODEL OF PARANOIA

The perception of threat is a central feature of paranoia almost by definition. However, several studies have explored this issue empirically by asking paranoid patients to estimate the past frequency with which they had experienced positive, negative and neutral events, and also the probability

that they will experience these events in the future (Bentall et al., in press; Corcoran, Blackwood, Howard, Kinderman, & Bentall, 2006; Kaney, Bowen-Jones, Dewey, & Bentall, 1997). In these studies, patients have reported high estimates for both past and future negative events, a phenomenon that can be resolved into three separate components. First, there is considerable evidence that paranoid patients have indeed actually experienced an abnormal frequency of adverse events such as discrimination and victimisation (Fuchs, 1992; Janssen et al., 2003; Mirowsky & Ross, 1983). Not only does this affect evaluations of the past, but also, because there is a tendency to rely on recollection of past events when making predictions about the future (called the availability heuristic; Kahneman, Slovic, & Tversky, 1982), it also tends to inflate estimates of future negative events. Second, patients also preferentially recall threat-related information (Bentall, Kaney & Bowen-Jones, 1995; Kaney, Wolfenden, Dewey, & Bentall, 1992), thus further biasing future estimates via the availability heuristic. Third, it is found that paranoid patients make inflated estimates of future negative events even after controlling for the above effects, as well as for the effects of comorbid anxiety and depression (Bentall et al., in press). This third component suggests that there is a specific abnormality in the mechanism responsible for aversive processing; and we will argue that this makes a large contribution to the formation of paranoid delusions.

Given perceived and potential aversive outcomes, which are exaggerated in paranoid patients, a second question concerns appropriate cognitive responses. A universal human response when faced with salient events is to construct an explanation for them, and attribution theory is the field of psychology that deals with how individuals construct such explanations (or attributions); it has been estimated that ordinary people generate a statement that either includes or implies the word "because" in every few hundred words of speech (Zullow, Oettingen, Peterson, & Seligman, 1988). Building on early psychodynamic and social-psychological work, proponents of attributional models of psychopathology have suggested that people appeal to two main classes of explanation for negative events. One is to attribute these events to something they themselves did (an internal explanation). The other is to attribute them to factors external to the self (an external attribution), and this latter kind of explanation can be further subdivided into other-blaming (external-personal) and circumstance (external-situational) attributions (Kinderman & Bentall, 1997). Most people err towards attributing negative events to external causes, which is thought to buffer against self-esteem loss in the face of failure or other threats to the self (Campbell & Sedikides, 1999; Mezulis, Abramson, Hyde, & Hankin, 2004), a phenomenon known as the self-serving bias.

It is known that the kinds of attributions people make have important implications for psychopathology. Numerous studies have shown that

depressed patients tend to make abnormally internal attributions for negative events (Mezulis et al., 2004). However, a number of studies have shown that paranoid patients, by contrast, tend to attribute negative events to excessively external causes (e.g., Fear, Sharp, & Healy, 1996; Kaney & Bentall, 1989) and especially external personal causes (Kinderman & Bentall, 1997). These observations have led to attempts to explain paranoid delusions in terms of these attributional processes. According to an early model (Bentall, Kinderman, & Kaney, 1994), paranoid patients have implicit negative self-schemas, which would ordinarily be readily activated to provoke conscious discrepancies between the individual's ideal self-concept and actual perception of the self. In an attempt to avoid this discrepancy, the individual attributes the cause of the schema-activating event to an external-personal cause (the actions of other people) but this leads to the belief that other people have malevolent intentions towards the self. The model proposed that persecutory delusions arise as the consequence of the iterative use of this defensive strategy in the face of repeated threats.

A common criticism of this model is that self-esteem is often low in paranoid patients (Freeman et al., 1998). In fact, research on self-esteem in paranoid patients has revealed a complex picture, with some studies showing either a close association between negative self-esteem and paranoia (e.g., Bentall et al., in press; Freeman et al., 1998), relatively preserved self-esteem on explicit measures but low self-esteem on implicit measures (e.g., Lyon, Kaney, & Bentall, 1994; McKay, Langdon, & Coltheart, 2007; Moritz, Werner, & von Collani, 2006), or no relationship between self-esteem and paranoia (Drake et al., 2004). Partly in response to this criticism, and also in the light of evidence that attributional judgements are influenced by current self-esteem (e.g., Kinderman & Bentall, 2000) and are highly labile in paranoid patients (Bentall & Kaney, 2005), a more recent version of the attributional model was proposed, in which a cyclic relationship between attributions and self-esteem was hypothesised to lead to highly unstable self-esteem in paranoid patients (Bentall, Corcoran, Howard, Blackwood, & Kinderman, 2001).

TWO TYPES OF PARANOIA

A further complication so far not explicitly addressed by the attributional models is that the defensive function of paranoid attributions appears to predominate only for a specific type of paranoia, or perhaps at specific stages in the development of persecutory delusions. Trower and Chadwick (1995) have distinguished between two types of paranoid beliefs: poor-me (in which persecution is believed to be undeserved) and bad-me (in which it is believed to be deserved), and have argued that defensive processes operate

only in the first of these types. Chadwick, Trower, Juusti-Butler, and Maguire (2005) recently reported that self-esteem is relatively preserved in poor-me patients compared to bad-me patients., However, another recent study reported that acutely ill patients often switch between poor-me and bad-me beliefs, but that, consistent with Trower and Chadwick's predictions, abnormal attributions are only present when patients hold poor-me beliefs (Melo, Taylor, & Bentall, 2006).

Abnormal attributions also appear to be absent in nonpsychotic individuals with paranoid beliefs (Janssen et al., 2006; McKay, Langdon, & Coltheart, 2005). The implication of these observations is that defensive processes are evident only in the acutely ill poor-me phase, and that bad-me paranoia is probably more evident during the prodromal phase before an acute crisis. Consistent with this account, a recent study of prodromal patients reported that this group has very marked discrepancies between their ideal self and their perceived self, but that the presence of actual psychotic symptoms was associated with a lack of such discrepancies (Morrison et al., 2006).

In the following account, we will argue that the conditioned avoidance paradigm helps us to understand the defensive processes operating in the poor-me phase.

THE *CAR* PARADIGM

The Conditioned Avoidance Response (CAR) paradigm was designed to assess learning and performance of behaviours motivated by aversion. Since the discovery of chlorpromazine (Swazey, 1974), it has been known that antipsychotic drugs selectively suppress avoidance responding, leaving escape responding relatively intact; the CAR is thus routinely used to help assess if a new compound is likely to be active against psychosis (Wadenberg & Hicks, 1999). Despite some important modelling of the role of dopamine in the CAR and psychosis (Kapur, Mizrahi, & Li, 2005; Smith, Becker, & Kapur, 2005), the threat inherent in the CAR has not yet been directly linked to the role of threat perception in persecutory delusions.

In the animal CAR paradigm, the subject is placed in a shuttle box with two compartments, for example one black and one grey. An animal placed in one of the compartments learns that a neutral warning stimulus (WS: e.g., a light) is followed after some seconds by an unconditioned aversive stimulus (AvS: an electric shock). After the onset of the WS, the subject can avoid the AvS by moving (shuttling) to the other compartment of the apparatus. Shuttling before the onset of the AvS avoids the shock and also interrupts the WS (in typical experiments the WS and the AvS overlap). This is termed an avoidance response (AR). Shuttling after AvS onset (an escape response,

or ER) also aborts the shock. CAR tasks are also performed with human participants. In human experiments shock administration is typically by cutaneous electrodes while the AR/ER involves pulling a lever (Unger, Evans, Rourke, & Levis, 2003).

Naive animals typically first display freezing in response to the AvS and subsequently show increased locomotion. After a few shocks, they perform the ER, presumably by chance. From then on they quickly learn to shuttle before AvS onset (AR). One main theory of learning in the CAR (Mowrer, 1947; Schmajuk & Zanutto, 1997) is that subjects first learn to fear (i.e., to predict the aversive values of) the states leading from the WS to the AvS in the absence of avoidance. They also learn the neutral value of the "other" compartment, and then have the avoidance response reinforced by the appetitive affective change experienced in going from the aversive states to the neutral (safety) states. For a detailed review of experimental findings, see Schmajuk and Zanutto (1997).

Once successful avoidance has been reliably achieved, the AR becomes resistant to extinction. Factors that confer resistance to extinction include increased magnitude of the AvS, factors relating to the timing of the AvS, and others whose details are peripheral to the issues in question here. These factors may all strengthen the WS-AvS association (Schmajuk & Zanutto, 1997). In some cases the latency between the warning signal and the avoidance response continues decreasing long after a reliable avoidance CR has been achieved (Solomon, Kamin, & Wynne, 1953). In these cases not only is there no extinction, but learning seems to continue to occur in the absence of shocks. One important way in which this can be reversed is by blocking shuttling while a WS–No-AvS contingency is presented.

Some of the neurobiological substrates of the CAR are well-established. All known antipsychotics (unlike other psychotropic compounds) disrupt performance of the well-learnt avoidance response at doses much lower than needed to affect the escape response (Wadenberg & Hicks, 1999). Almost all antipsychotic compounds with selective action on the CAR block dopamine D2 receptors. D2 blockade disrupts performance of well-learnt avoidance responses, but also the acquisition of the AR (reviewed by Smith et al., 2005). It was realised at an early stage (Beninger, Mason, Phillips, & Fibiger, 1980) that D2 blockade does not disrupt the development of the association between warning and aversive stimuli but the development of the AR itself. This was shown by the fact that when both the AvS and the D2 blocker were eliminated, presentation of the WS on its own led to the *gradual* acquisition of the AR. Anatomically, one of the most important sites of action of D2 receptors with respect to the CAR is the shell of the nucleus accumbens septi (NAS-shell). Drugs that affect other neurotransmitters can also affect the CAR, but mostly in synergy with dopaminergic modulation. The role of serotonin is particularly interesting, as NAS-shell 5HT2 blockade greatly

potentiates the effect of D2 blockade on the CAR (Wadenberg & Hicks, 1999).

LINKING PARANOID DELUSIONS AND THE *CAR*

Besides the CAR's well-recognised predictive validity for antipsychotic effects, we suggest that it is also a valid and revealing model for fundamental aspects of paranoid delusions. Further biological links do exist. For example, drugs that enhance dopamine function often cause paranoid syndromes in humans (e.g., Satel, Southwick, & Gawin, 1991). Moreover, neuroimaging studies indicate that it is during the acute stage of psychosis, when poor-me delusions are most evident, that abnormal functioning of the midbrain dopamine system is most evident (Laruelle, Abi-Dargham, Gil, Kegeles, & Innis, 1999). In this paper, however, we shall concentrate on the psychological/functional links. We draw three key psychological/functional parallels. First, that both CAR and paranoia involve threat-perception mechanisms. We thus relate the taking of defensive action in the CAR to defensive avoidance in paranoia. Second, we note that avoidance responses in the CAR, like paranoid delusions, are markedly resistant to extinction. Finally, a more subtle point perhaps, we consider that the defensive function of poor-me beliefs may represent a form of covert avoidance.

According to this hypothesis, threat-perception mechanisms are linked not only to normal aversive learning in the CAR but also to unwarranted associations in paranoia. Modern models of affectively charged adaptive behaviour have made important inroads in understanding both reward- and threat-motivated learning. These models already include accounts of functional aspects of neuromodulators and especially of dopamine (Daw, Kakade, & Dayan, 2002; Montague, Dayan, & Sejnowski, 1996; Schultz, Dayan, & Montague, 1997; Seymour et al., 2003). If our hypothesis is valid, a whole new field of investigation opens up—relating models of specifically aversive learning to the abnormal psychology of paranoia.

CAR, PARANOIA, AND THREAT

The first parallel, namely that both the CAR and paranoia involve the perception of substantial threat seems straightforward. In the CAR the shock is of course quite real while paranoid delusions are, by definition, unrealistic. The initial establishment of threat perception in paranoia is thus clearly important. Indeed, one main aim of our research programme is to understand what might be going wrong in paranoia by considering what might be going on in the CAR. Our working hypothesis is that abnormalities of aversive processing, perhaps in a prodromal phase of psychosis create

fictitious, internal, aversive evaluation states, and then lead to bad-me and then poor-me delusions through evaluative and defensive mechanisms.

The content of persecutory delusions is specifically about social/interpersonal threat. This is not surprising given that, in highly social animals, emotional systems are naturally responsive to the harms (and benefits) that may come from conspecifics. In this context it is interesting to note that animal studies show that repeated exposure to social defeat leads to sensitisation of the mesolimbic dopamine system (Selten & Cantor-Graae, 2005; Tidey & Miczeck, 1996). This is analogous, perhaps, to the experiences of discrimination and victimisation that seem to confer a high risk of paranoia (Fuchs, 1992; Janssen et al., 2006; Mirowsky & Ross, 1983). Another interesting and relevant observation is that oversensitivity of the dopamine system is important in the emotional sensitivity of people at high risk of psychosis (Myin-Germeys, Marcelis, Krabbendam, Delespaul, & van Os, 2005). Increased expectation of socially mediated harm as well as low self-esteem are likely psychological sequelae of repeated social defeat accompanied by increased dopamine reactivity. Hence, these observations are consistent with the hypothesis that fragile self-esteem plays a role in the onset of paranoia, as proposed by attribution theorists.

SAFETY BEHAVIOURS PLAY AN IMPORTANT ROLE IN PARANOIA AND ARE DIFFICULT TO EXTINGUISH

The second parallel concerns overt avoidance behaviours. The notion that safety behaviours help maintain paranoid ideas was put forward by Morrison (1998) on the basis of case studies. Paranoid patients perceive serious threat in the social environment and take efficient action to neutralise the threat, mainly by avoiding circumstances in which the expected threat might be encountered. As a consequence, they miss opportunities to find out that their threat-beliefs are unfounded (a simple example is when a paranoid patient stays indoors to avoid meeting imagined persecutors, thereby failing to learn that people outside the home are benign).

Freeman, Garety, and Kuipers (2001) formally investigated safety behaviours in a sample of 25 psychotic patients and found that 92% of participants reported overt avoidance. As avoidance behaviours appear to reduce exposure to disconfirmatory evidence and hence prevent modification of threat-beliefs, cognitive-behaviour therapists often find it helpful to use behavioural experiments to help test persecutory beliefs (Morrison, Renton, Dunn, Williams, & Bentall, 2003). From this point of view, avoidance behaviours are extremely common in paranoia. They reduce the experience of perceived warning stimuli not leading to feared consequences,

thus reducing opportunities to modify inappropriate predictions. The CAR provides a close parallel in that response blocking is often required to achieve extinction of avoidance responses.

POOR-ME AND BAD-ME PARANOIA

The third parallel has to do with avoidance of internal aversive states in paranoia (i.e., experiential rather than overt avoidance). We suggest a mapping between avoidance behaviours in CAR and the possible defensive role of paranoid ideas. As already noted, there has been an intensive debate in the literature about whether persecutory beliefs enable the individual to avoid feelings of low self-esteem (Garety & Freeman, 1999) as initially suggested by attribution researchers (Bentall et al., 1994). This suggestion has been challenged on the basis of findings of low self-esteem in paranoid patients (Freeman et al., 1998). As we have already indicated, a possible resolution to this problem can be found in the distinction between poor-me and bad-me paranoia, and the observation that abnormal attributions and relatively preserved self-esteem are only found when acutely ill psychotic patients hold poor-me beliefs (Chadwick et al., 2005; Melo et al., 2006).

As we have also already seen, abnormal attributions are only found in acutely psychotic patients (Janssen et al., 2006; McKay et al., 2005) who are nearly always poor-me (Bentall et al., in press; Fornells-Ambrojo & Garety, 2005). Moreover, in prodromal patients scores on self-esteem-related measures appear to improve with increasing psychosis (Morrison et al., 2006). Together these observations suggest a developmental pathway leading to clinical paranoia, in which experiences of victimisation and social defeat lead to poor self-esteem and the growing conviction that others also hold negative views about the self, and hence to bad-me beliefs. These beliefs, maintained and amplified by safety behaviours, are eventually transformed into poor-me beliefs in acute psychosis, when avoidance is extended to attempts to avoid negative thoughts about the self.

Note that, in this hypothesised progression, the final defensive response (including the generation of an explanation for a negative event that implicates external-personal causes, a poor-me delusion) can be understood within the CAR framework as a form of covert avoidance behaviour, in which negative thoughts about the self are a covert WS which would elicit a strong internal AvS (a negative emotional state) which is avoided by means of an external-personal attribution. This account assumes that negative thoughts about the self have strong emotional effects, can be regarded as response-provoking stimuli, and that individuals sometimes respond so as to avoid them successfully. Clearly, the first two of these assumptions are concordant

with everyday observation (for example, of the surge of negative affect that follows the realisation that one has been seen to do something shameful) and the last is consistent with many recent accounts of psychopathology which have emphasised the role of experiential avoidance processes (e.g., Hayes, Strosahl, & Wilson, 1999; Rassin, Merckelbach, & Muris, 2000).

In this context, it is interesting to consider the possibility that the tendency to jump to conclusions when reasoning about probabilistic information found in many deluded individuals (Garety & Freeman, 1999) may also be related to the avoidance mechanisms involved in the CAR. Exaggerated avoidance of the discomfort associated with uncertainty could contribute to the cognitive biases of jumping-to-conclusions and increased need-for-closure found in delusions (McKay, Langdon, & Coltheart, 2006). This possibility is certainly worthy of further experimental investigation.

FUTURE DIRECTIONS

In this paper we have outlined a number of parallels between the CAR and evidence regarding the psychological mechanisms in paranoia. The CAR has previously been studied as a model of anxiety disorders (Lovibond, 2006), hence our account suggests some overlap between the processes involved in paranoia and anxiety. Freeman and co-workers (Freeman & Garety, 2005) have previously argued for a close relationship between paranoia and anxiety. However, in their account it is assumed that the subjective experience of anxiety prompts paranoid verbalisations (i.e., the latter are *expressive* of the sense of impending danger that is central to anxiety). In our account, by contrast, anxiety is the *consequence* of the perception of either external or internal threat.

The potential relationship between CAR and paranoia opens up a number of lines of enquiry. From a neuroscience perspective, the role of dopamine is both central and enigmatic. Dopamine is substantially implicated in many aspects of learning predictions (Montague et al., 1996) and optimising actions. These aspects have been the target of substantial computational modelling that links neural, psychological, and ethological ideas. Such modelling is relevant to the CAR, but is unlikely to apply directly as these roles of dopamine have been established in the context of reward stimuli, not threats. Analysis of dopamine's role in signalling with respect to aversive events is rather less clear (Ungless, Magill, & Bolam, 2004). It may have little role in the signalling of the aversive events themselves, as evidenced by the intact formation of WS-AvS associations under dopamine blockade (Beninger et al., 1980). This finding is also a challenge to models that require a dopaminergic

"teacher" signal for the formation of internal representations of the environment (cf. Smith, Li, Becker & Kapur, 2006). Dopamine may instead be involved in the reward (relief) brought on by evasive actions. Opponency between dopamine and other neuromodulators, particularly serotonin, may turn out to be key in this case (Daw et al., 2002; Ungless, 2004). The psychological effects of antipsychotic drugs are yet to be integrated with these potential roles of neuromodulators, especially with respect to threat-based learning. However, if our account is correct, it follows that these drugs may function psychologically by reducing patients' experiential avoidance, which may help maintain their symptoms. An interesting corollary is that bad-me delusions, which we hypothesise not to involve this avoidance, should be less responsive to antipsychotic drugs than poor-me delusions. Although this possibility has never been tested empirically, it is interesting to note that a recent Cochrane review of the treatment of psychotic depression found no evidence that antipsychotics are effective in this condition (Wijkstra, Lijmer, Balk, Geddes, & Nolen, 2005).

If our suggestion is valid, and flight from states of high negative self-esteem plays an important contributory factor in poor-me persecutory delusions, then individuals prone to such delusions must differ from healthy individuals in the related parameters. That is, our account suggests that individuals vulnerable to paranoia may make excessive estimations of internal threat, or use excessively avoidant cognitive strategies to deal with it. This might then be a dynamic contributing to the transition from subclinical to clinical paranoia. This progression needs to be verified in a longitudinal investigation of prodromal patients. It would also be profitable to directly compare avoidance mechanisms in healthy individuals and those prone to delusions. This could take place in CAR-like experimental learning and the extent to which antipsychotic drugs blunt such learning, as they do in animals, would be important to investigate.

Most importantly, the suggested relationship between the CAR and paranoia permits theory-based psychological and neurobiological investigation, guided by quantitative computational models. Such investigations should include neuroimaging studies of both healthy individuals and paranoid patients using paradigms analogous to the CAR paradigms used in animal studies, extended to socially threatening stimuli and threats to self-esteem. Further research into the CAR and its relation to human psychological processes would thus help develop an integrated psychobiological understanding of the threat beliefs that are one of the most common symptoms of severe mental illness.

Manuscript received 17 July 2006
Revised manuscript received 20 June 2007

REFERENCES

Beninger, R., Mason, S., Phillips, A., & Fibiger, H. (1980). The use of extinction to investigate the nature of neuroleptic-induced avoidance deficits. *Psychopharmacology*, *69*, 11–18.

Bentall, R. P. (1994). Cognitive biases and abnormal beliefs: Towards a model of persecutory delusions. In A. David & J. Cutting (Eds.), *The neuropsychology of schizophrenia* (pp. 337–360). Hove, UK: Lawrence Erlbaum Associates Ltd.

Bentall, R. P. (2003). *Madness explained: Psychosis and human nature*. London: Penguin.

Bentall, R. P., Corcoran, R., Howard, R., Blackwood, N., & Kinderman, P. (2001). Persecutory delusions: A review and theoretical integration. *Clinical Psychology Review*, *21*, 1143–1192.

Bentall, R. P., & Kaney, S. (2005). Attributional lability in depression and paranoia. *British Journal of Clinical Psychology*, *44*, 475–488.

Bentall, R. P., Kaney, S., & Bowen-Jones, K. (1995). Persecutory delusions and recall of threat-related, depression-related and neutral words. *Cognitive Therapy and Research*, *19*, 331–343.

Bentall, R. P., Kinderman, P., Howard, R., Blackwood, N., Cummins, S., Rowse, G., et al. (in press). Paranoid delusions in schizophrenia and depression: The transdiagnostic role of expectations of negative events and negative self-esteem. *Journal of Nervous and Mental Disease*.

Bentall, R. P., Kinderman, P., & Kaney, S. (1994). The self, attributional processes and abnormal beliefs: Towards a model of persecutory delusions. *Behaviour Research and Therapy*, *32*, 331–341.

Berrios, G. (1991). Delusions as "wrong beliefs": A conceptual history. *British Journal of Psychiatry*, *159*, 6–13.

Campbell, W. K., & Sedikides, C. (1999). Self-threat magnifies the self-serving bias: A meta-analytic integration. *Review of General Psychology*, *3*, 23–43.

Chadwick, P., Trower, P., Juusti-Butler, T.-M., & Maguire, N. (2005). Phenomenological evidence for two types of paranoia. *Psychopathology*, *38*, 327–333.

Corcoran, R., Blackwood, N., Howard, R. J., Kinderman, P., & Bentall, R. P. (2006). Heuristic reasoning in deluded and depressed patients. *Psychological Medicine*, *36*, 1109–1118.

David, A. S. (1999). On the impossibility of defining delusions. Philosophy. *Psychiatry and Psychology*, *6*, 17–20.

Daw, N., Kakade, S., & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks*, *15*, 603–616.

Drake, R. J., Pickles, A., Bentall, R. P., Kinderman, P., Haddock, G., Tarrier, N., et al. (2004). The evolution of insight, paranoia and depression during early schizophrenia. *Psychological Medicine*, *34*, 285–292.

Fear, C. F., Sharp, H., & Healy, D. (1996). Cognitive processes in delusional disorder. *British Journal of Psychiatry*, *168*, 61–67.

Fornells-Ambrojo, M., & Garety, P. (2005). Bad me paranoia in early psychosis: A relatively rare phenomenon. *British Journal of Clinical Psychology*, *44*, 521–528.

Freeman, D., & Garety, P. A. (2005). Connecting neurosis to psychosis: The direct influence of emotion on delusions and hallucinations. *Behaviour Research and Therapy*, *41*, 923–947.

Freeman, D., Garety, P. A., Bebbington, P. E., Smith, B., Rollinson, R., Fowler, D., et al. (2005). Psychological investigation of the structure of paranoia in a non-clinical population. *British Journal of Psychiatry*, *186*, 427–435.

Freeman, D., Garety, P., Fowler, D., Kuipers, E., Dunn, G., Bebbington, P., & Hadley, C. (1998). The London-East Anglia randomized controlled trial of cognitive-behaviour therapy for psychosis IV: Self-esteem and persecutory delusions. *British Journal of Clinical Psychology*, *37*, 415–430.

Freeman, D., Garety, P., & Kuipers, E. (2001). Persecutory delusions: Developing the understanding of belief maintenance and emotional distress. *Psychological Medicine*, *31*, 1293–1306.

Fuchs, T. (1992). Life events in late paraphrenia and depression. *Psychopathology*, *32*, 60–69.

Garety, P., & Freeman, D. (1999). Cognitive approaches to delusions: A critical review of theories and evidence. *British Journal of Clinical Psychology*.

Garety, P. A., & Hemsley, D. R. (1987). The characteristics of delusional experience. *European Archives of Psychiatry and Neurological Sciences*, *236*, 294–298.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. New York: Guilford Press.

Hoenig, J. (1982). Kurt Schneider and anglophone psychiatry. *Comprehensive Psychiatry*, *23*, 391–400.

Janssen, I., Hanssen, M., Bak, M., Bijl, R. V., de Graaf, R., Vollenberg, W., et al. (2003). Discrimination and delusional ideation. *British Journal of Psychiatry*, *182*, 71–76.

Janssen, I., Versmissen, D., Campo, J. A., Myin-Germeys, I., van Os, J., & Krabbendam, L. (2006). Attributional style and psychosis: Evidence for externalizing bias in patients but not individuals at high risk. *Psychological Medicine*, *27*, 1–8.

Jorgensen, P., & Jensen, J. (1994). Delusional beliefs in first admitters. *Psychopathology*, *27*, 100–112.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Kaney, S., & Bentall, R. P. (1989). Persecutory delusions and attributional style. *British Journal of Medical Psychology*, *62*, 191–198.

Kaney, S., Bowen-Jones, K., Dewey, M. E., & Bentall, R. P. (1997). Frequency and consensus judgements of paranoid, paranoid-depressed and depressed psychiatric patients: Subjective estimates for positive, negative and neutral events. *British Journal of Clinical Psychology*, *36*, 349–364.

Kaney, S., Wolfenden, M., Dewey, M. E., & Bentall, R. P. (1992). Persecutory delusions and the recall of threatening and non-threatening propositions. *British Journal of Clinical Psychology*, *32*, 85–87.

Kapur, S., Mizrahi, R., & Li, M. (2005). From dopamine to salience to psychosis—linking biology, pharmacology and phenomenology of psychosis. *Schizophrenia Research*, *79*, 59–68.

Kay, S. R., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, *13*, 507–518.

Kendler, K. S., Glazer, W., & Morgenstern, H. (1983). Dimensions of delusional experience. *American Journal of Psychiatry*, *140*, 466–469.

Kinderman, P., & Bentall, R. P. (1997). Causal attributions in paranoia: Internal, personal and situational attributions for negative events. *Journal of Abnormal Psychology*, *106*, 341–345.

Kinderman, P., & Bentall, R. P. (2000). Self-discrepancies and causal attributions: Studies of hypothesized relationships. *British Journal of Clinical Psychology*, *39*, 255–273.

Laruelle, M., Abi-Dargham, A., Gil, R., Kegeles, L., & Innis, R. (1999). Increased dopamine transmission in schizophrenia: Relationship to illness phases. *Biological Psychiatry*, *46*, 56–72.

Lovibond, P. (2006). Fear and avoidance: An integrated expectancy model. In M. G. Craske, D. Hermans, & D. Vansteenwegen (Eds.), *Fear and learning: From basic processes to clinical implications* (pp. 117–132). Washington, DC: American Psychological Association.

Lyon, H. M., Kaney, S., & Bentall, R. P. (1994). The defensive function of persecutory delusions: Evidence from attribution tasks. *British Journal of Psychiatry*, *164*, 637–646.

McKay, R., Langdon, R., & Coltheart, M. (2005). Paranoia, persecutory delusions and attributional biases. *Psychiatry Research*, *136*, 233–245.

McKay, R., Langdon, R., & Coltheart, M. (2006). Need for closure, jumping to conclusions, and decisiveness in delusion-prone individuals. *Journal of Nervous and Mental Disease*, *194*(6), 422–426.

McKay, R., Langdon, R., & Coltheart, M. (2007). The defensive function of persecutory delusions: An investigation using the Implicit Association Test. *Cognitive Neuropsychiatry*, *12*, 1–24.

Melo, S., Taylor, J., & Bentall, R. P. (2006). Poor me'' versus "bad me'': Paranoia and the instability of persecutory ideation. Psychology and Psychotherapy—Theory, Research. *Practice*, *79*, 271–287.

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, *130*, 711–747.

Mirowsky, J., & Ross, C. E. (1983). Paranoia and the structure of powerlessness. *American Sociological Review*, *48*, 228–239.

Montague, R., Dayan, P., & Sejnowski, T. (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience*, *76*, 1936–1947.

Moritz, S., Werner, R., & von Collani, G. (2006). The inferiority complex in paranoia readdressed: A study with the Implicit Association Test. *Cognitive Neuropsychiatry*, *11*, 402–415.

Morrison, A. (1998). Treating complex cases: The cognitive behavioural therapy approach. In N. Tarrier, A. Wells, & G. Haddock (Eds.), *Cognitive behaviour therapy for psychotic symptoms in schizophrenia* (pp. 195–216). Chichester, UK: Wiley.

Morrison, A. P., French, P., Lewis, S. W., Roberts, M., Raja, S., Parker, S., et al. (2006). Psychological factors in people at ultra-high risk of psychosis: Comparisons with non-patients and associations with symptoms. *Psychological Medicine*, *36*, 1395–1404.

Morrison, A. P., Renton, J., Dunn, H., Williams, S., & Bentall, R. P. (2003). *Cognitive therapy for psychosis: A formulation-based approach*. Hove, UK: Brunner-Routledge.

Mowrer, O. (1947). On the dual nature of learning: A reinterpretation of conditioning and problem solving. *Harvard Educational Review*, *17*, 102–148.

Myin-Germeys, I., Marcelis, M., Krabbendam, L., Delespaul, P., & van Os, J. (2005). Subtle fluctuations in psychotic phenomena as functional states of abnormal dopamine reactivity in individuals at risk. *Biological Psychiatry*, *58*(2), 105–110.

Ndetei, D. M., & Vadher, A. (1984). Frequency and clinical significance of delusions across cultures. *Acta Psychiatrica Scandinavica*, *70*, 73–76.

Rassin, E., Merckelbach, H., & Muris, P. (2000). Paradoxical and less paradoxical effects of thought suppression: A critical review. *Clinical Psychology Review*, *20*, 973–995.

Raune, D., Bebbington, P., Dunn, G., & Kuipers, E. (2006). Event attributes and the content of psychotic experiences in first-episode psychosis. *Psychological Medicine*, *36*, 221–230.

Satel, S., Southwick, S., & Gawin, F. (1991). Clinical features of cocaine-induced paranoia. *American Journal of Psychiatry*, *148*, 495–498.

Schmajuk, N., & Zanutto, B. (1997). Escape, avoidance and imitation: A neural network approach. *Adaptive Behavior*, *6*, 63–129.

Schneider, K. (1974). Themes and variations in European psychiatry. In S. Hirsch & M. Shepherd (Eds.), *The concept of delusion* (pp. 33–39). Bristol, UK: John Wright & Sons. (Original work published 1949)

Schultz, W., Dayan, P., & Montague, P. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Selten, J.-P., & Cantor-Graae, E. (2005). Social defeat: Risk factor for schizophrenia? *British Journal of Psychiatry*, *187*, 101–102.

Seymour, B., O'Doherty, J., Dayan, P., Koltzenburg, M., Jones, A., Dolan, R., et al. (2003). Temporal difference models describe higher-order learning in humans. *Nature*, *429*, 664–667.

Smith, A., Becker, S., & Kapur, S. (2005). A computational model for the functional role of ventral-striatal d2 receptor in the expression of previously acquired behaviors. *Neural Computation*, *17*, 361–395.

Smith, A., Li, M., Becker, S., & Kapur, S. (2006). Dopamine, prediction error and associative learning: A model-based account. *Network: Computation in Neural Systems*, *17*, 61–84.

Solomon, R., Kamin, L., & Wynne, L. (1953). Traumatic avoidance learning: The outcomes of several extinction procedures in dogs. *Journal of Abnormal and Social Psychology*, *48*, 291–302.

Stompe, T., Friedman, A., Ortwein, G., Strobl, R., Chaudhry, H. R., Najam, N., et al. (1999). Comparisons of delusions among schizophrenics in Austria and Pakistan. *Psychopathology*, *32*, 225–234.

Swazey, J. P. (1974). *Chlorpromazine in psychiatry: A study of therapeutic innovation*. Cambridge, MA: MIT Press.

Tarrier, N., Lewis, S., Haddock, G., Bentall, R. P., Drake, R., Dunn, G., et al. (2004). 18 month follow-up of a randomized controlled trial of cognitive-behaviour therapy in first episode and early schizophrenia. *British Journal of Psychiatry*, *184*, 231–239.

Tidey, J., & Miczeck, K. (1996). Social defeat stress selectively alters mesocorticolimbic dopamine release: An in vivo microdialysis study. *Brain Research*, *721*, 140–149.

Trower, P., & Chadwick, P. (1995). Pathways to defence of the self: A theory of two types of paranoia. *Clinical Psychology: Science and Practice*, *2*, 263–278.

Unger, W., Evans, I., Rourke, P., & Levis, D. (2003). The s-s construct of expectancy versus the s-r construct of fear: Which motivates the acquisition of avoidance behaviour? *Journal of General Psychology*, *130*, 131–147.

Ungless, M. (2004). Dopamine: The salient issue. *Trends in the Neurosciences*, *27*, 702–706.

Ungless, M., Magill, P., & Bolam, J. (2004). Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science*, *303*, 2040–2042.

Wadenberg, M., & Hicks, P. (1999). The conditioned avoidance response test re-evaluated: Is it a sensitive test for the detection of potentially atypical antipsychotics? *Neuroscience and Biobehavioral Reviews*, *23*, 851–862.

Wijkstra, L., Lijmer, J., Balk, F., Geddes, J., & Nolen, W. A. (2005). *Pharmacological treatment for psychotic depression*. Retrieved from www.chochrane.org

Zullow, H., Oettingen, G., Peterson, C., & Seligman, M. (1988). Pessimistic explanatory style in the historical record: CA Ving LBJ, Presidential Candidates, and East versus West Berlin. *The American Psychologist*, *43*, 673–682.