An evolutionary perspective on paranoia

Nichola J Raihani[1] & Vaughan Bell[2]

1. Department of Experimental Psychology, University College London, 26 Bedford Way, WC1H 0AP

2. Division of Psychiatry, University College London.

Abstract

Paranoia is the most common symptom of psychosis but paranoid concerns occur throughout the general population. Here, we argue for an evolutionary approach to paranoia across the spectrum of severity that accounts for its complex social phenomenology – including the perception of conspiracy and selective identification of perceived persecutors – and considers how it can be understood in light of our evolved social cognition. We argue that the presence of coalitions and coordination between groups in competitive situations could favour psychological mechanisms that detect, anticipate and avoid social threats. Our hypothesis makes testable predictions about the environments in which paranoia should be most common as well as the developmental trajectory of paranoia across the lifespan. We suggest that paranoia should not solely be viewed as a pathological symptom of a mental disorder but also as a part of a normally-functioning human psychology.

Humans evolved in complex and dynamic groups comprised of kin and non-kin. Life in complex social groups favours the evolution of specialized and sophisticated socio-cognitive abilities [1–3] including the ability to form and maintain coalitions and alliances (e.g. hyenas [4]; chimpanzees [5], corvids [6]), to recognise and categorise other individuals in terms of dominance (e.g. pinyon jays [7]) and alliance membership (e.g. Hamadryas baboons [8]), and - to varying degrees - to predict and manipulate the intentions and behaviour of others (e.g. anthropoid apes [9], western scrub jays [10]). In this article we argue that paranoia involves all of these socio-cognitive abilities and that the human ability for paranoid thinking evolved in response to these social selection pressures. Evolutionary accounts of paranoia have been proposed before [11,12] but have not fully accounted for the full phenomenological complexity of paranoia, nor shown how such a perspective has the potential to explain variation in paranoia both across contexts and over development. We explore why paranoid thinking is such a common human characteristic and why paranoia can become intense and disabling after many forms of social, psychological and neurological difficulties.

Current conceptualisation of paranoia

A persecutory belief is considered to be the central defining feature of paranoia and includes two essential elements: i) a belief that harm will occur, and ii) an attribution that others intend this harm [13]. In the general population, such persecutory ideas can be experienced with varying degrees of frequency and entertained to varying degrees of intensity. Paranoia can range from mild concerns about others' intentions to beliefs that are sufficiently unlikely, and inflexible to be classified as a psychiatric symptom, most notably, as a paranoid delusion [14]. One of the implicit assumptions about paranoia is that it represents an exaggerated or false attribution of harmful intent to others. However, given the continuum of paranoia, paranoid explanations can, and occasionally should, be accurate (e.g. see [15,16]) although these are likely to be increasingly inaccurate as paranoia becomes more becomes more disabling and a likely focus of clinical concern [16,17].

Epidemiological studies show that paranoia shows full taxometric continuity throughout the population, indicating that categorical distinctions used in psychiatric diagnosis are not reflected in a clear point of change of severity in the population [18–20]. Nevertheless, this continuous distribution in the population does not necessarily imply that underlying causes are fully continuous within individuals, over time, or between sub-groups [21]. Most current research has focused on paranoia in the context of mental health, typically in people without individually diagnosable neurological disorder, and has identified various risk factors and cognitive process that support paranoid thinking. Indeed, paranoia has now been reliably associated with living in areas of low social cohesion [22], worry [23], sleep deprivation [24,25], victimisation [25,26], and early life adversity, abuse and trauma [27]. Paranoia has also been found to co-occur with general cognitive biases relating to causal and probabilistic reasoning and belief flexibility [28,29]. However, diagnosable paranoid states can also be caused by a

wide range of direct disturbances to brain function. Paranoia is common in psychosis following epilepsy [30], brain injury [31,32] and dementia [33]. It is also one of the most frequent unwanted side-effects for several classes of recreational drugs [34–36]. Indeed, in terms of the causes and contexts in which it appears, paranoia is perhaps most remarkable for being associated with such a wide range of difficulties, impairments and stresses.

Given this diversity, the aim of this article is to ask whether paranoia might sometimes serve an adaptive (fitness-enhancing) function and how an evolutionary perspective can help us to predict where paranoia will be most common. To be clear, our aim is not an attempt to explain how frank paranoid delusions and – by extension – psychotic spectrum disorders, have been favoured by selection. Indeed, based on the lowered reproductive success of individuals with these disorders and the lack of evidence of benefits to kin [37], we think that this is highly unlikely. Our overarching hypothesis is that the existence of paranoia can generally be understood as a consequence of selection for detecting and evaluating coalitional threat. We first describe the phenomenology of paranoia and argue that current evolutionary theories do not fully account for the perception of conspiracy and selective identification of arbitrary persecutors that are so common in paranoia. We suggest that coalitionary competition, which can occur both within and between groups and which can be relatively stable in some contexts and yet highly flexible in others, can help to explain why paranoia takes the form it does. Our hypothesis predicts that within-individual variation in paranoid thinking should occur in response to immediate context-specific changes in the perception of coalitionary threat (as defined by [38]), whereas stable between-individual differences in paranoia are likely to emerge in response to chronic threat from others. Finally, we explore why impairments to brain function also commonly predispose individuals to paranoia, and whether this is likely to be an adaptive response to the environment or a maladaptive consequence of cognitive constraints.

Understanding the full social phenomenology of paranoia

Freeman and Garety's [13] definition has been useful in providing a clear operational definition of a central component of paranoia. However, existing approaches to paranoia have tended to conceptualise paranoia in terms of cognitive processes used to make sense of other *individuals* rather than *groups*. One limitation of this approach is that it fails to account for why the experience of more severe paranoia often involves the misperception of group boundaries and collective action. Indeed, paranoia is frequently accompanied by other features that are common enough to be included in phenomenological descriptions, both historical and modern, but are often neglected by more recent cognitive approaches. These are i) the perception of a conspiracy behind the intentional harm, and ii) social selectivity in terms of identifying the people perceived to be the source of intentional harm.

Conspiracy thinking is common in the general population [39,40] and is defined as a tendency to provide "explanations for important events that involve secret plots by powerful and malevolent groups" [41]. In paranoid delusions, however, conspiracy thinking often becomes self-focused, with delusions commonly involving the perception of organised attempts to harm the believer, rather than malign and impersonal explanations for public events. The perception of a self-focused conspiracy has been identified as a central characteristic of delusional paranoia from early in the history of psychiatry [42] and forms part of many modern phenomenological descriptions [43,44]. Cameron [45] conceptualised this aspect of paranoia as a belief in a persecutory 'pseudo-community' who are perceived to be united in a co-ordinated undertaking against the paranoid individual but who fail to correspond to any group in wider society who share the coordinated aims and actions attributed to them. Unlike public conspiracy theories, these concerns are more likely to focus on the history, intentions and day-to-day activities of the believer.

Although paranoia involves a belief that others intend harm to the believer, these concerns typically pertain to specific individuals or social groups and also commonly involve the misperception of group boundaries and coordinated group action. In increasingly severe paranoia, these concerns and misperceptions become increasingly exaggerated and may present as frank persecutory delusions. Studies of delusional patients indicate that the majority selectively identify specific groups as responsible for their maltreatment. In a study of delusions in Korean, Korean-Chinese, and Chinese patients conducted by Kim et al.[46], only 27.4%, 17.7% and 24.6% of persecutors, respectively, were unspecified, while the rest were variously identified as groups such as relatives, neighbours, the police, or medical personnel (see also[47]). Green et al.[48] reported that persecutory delusions could be classified as focusing on individuals (e.g. "my father"), groups with defined members ("[the patient's] neighbour, his neighbour's brother and mates"), established social groups ("the police"), undefined groups ("people", "spirits") and all others ("everyone") with perceived individual and multiple persecutors each consisting 50% of the total.

Evolutionary approaches to paranoia

Attempts to answer the question of why some people are more paranoid than others have typically appealed to proximate level explanations (Box 1) such as genetics, life history or cognitive biases [14]. Nevertheless, these approaches do not answer the issue of why we have a cognitive capacity for paranoid thinking (Box 2) and whether between-individual variation in paranoid thinking might, in some environments, be selectively advantageous in fitness terms. From a Darwinian perspective, a fearful response to danger, whether actual or potential, is likely to carry significant fitness benefits and to have been subject to strong selection in many species [49]. Nevertheless, not all individuals show an equivalent magnitude of response to the same threatening stimulus or context: levels of fearfulness differ markedly across individuals, even within a species. The question of how stable, between-

individual differences in fearful responses might arise and be stabilised by selection falls under a broader banner of research on the evolution of stable behavioural types. Research in this field has shown that the evolution of variation in behavioural types stems from trade-offs in pursuing different fitness-relevant activities. For example, investing in growth (e.g. via foraging) often comes with an attendant increased risk of predation [50,51] and so strategies aimed at increasing growth are likely to be traded-off against strategies that reduce predation risk. Organisms must therefore balance the rewards of investment in growth against the increased mortality risk; the optimal resolution of such trade-offs in different environments or for different individuals can therefore select for variation in fearfulness, aggression, risk appetite and so on, which broadly dictate individual life history strategies and associated behaviour.

In addition to balancing such trade-offs, organisms must also effectively manage costs from errors that occur due to perceptual uncertainty ('error management theory' [52], Box 3). Specifically, error management theory (also conceptualised as 'the smoke detector principle' in evolutionary medicine [53]) predicts that when there are asymmetries in the costs of false-positive and false-negative error types, selection will favour strategies that minimise the chance of making the costlier error, even if this produces many behavioural mistakes. Following the logic of error management theory, previous evolutionary accounts [11,52] have suggested that paranoia is an evolved psychological mechanism shaped by the selective pressures of catastrophic harm from others that is tuned to have a low threshold for detecting social threat. Individual variation in the relative asymmetry of error types is proposed to account for variation in paranoia across the full spectrum (see Box 3 for a critique).

Shortcomings of existing evolutionary theories

Nevertheless, existing evolutionary theories of paranoia based solely on social threat detection do not fully account for the complex phenomenology of paranoia. Specifically, we have to ask why a mechanism aimed at detecting and avoiding social threats does not solely result in variation in avoidance, submissive or appeasement behaviours (as is also observed in many non-human species, see [54] and also discussed elsewhere [55–59], but also incorporates more complex features that are not adequately explained by this approach. Namely, selective identification of a specific yet often seemingly arbitrary group of persecutors, the attribution of unobservable malign intentions and motives to these individuals, and the formulation of hypothetical narratives rendering these attributions subjectively plausible. Below, we focus on the first of these features but see Box 2 for a discussion of the evolution of inferential causal reasoning abilities (including mental state attribution) in humans.

An important feature of human social groups is the presence of coalitions: any situation where two or more individuals unite in competition against a third party or parties [60,61]. Coalitionary conflict in

human groups can manifest in the form of lethal aggression ('lethal raids' reviewed in [62]) but can also include non-lethal and non-aggressive conflict, such as stigmatization, ostracism, exclusion, and derogation. For example, witchcraft accusations have been (and still are) used to identify individuals or groups for ostracism, persecution or even death [63,64]. In modern industrialised societies, similar forms of indirect aggression are used by coalitions to damage the reputation of (often higher-ranking) rival, for example via gossip or derogation (see [65,66]).

This persistent risk of persecution selects for what others have called a 'coalitional psychology' that anticipates and deflects these threats by integrating oneself within a coalition or coalition(s), recognising and categorizing others as allies or potential competitors; and using these categorizations to predict how others might behave or react in specific social interactions [38,67,68]. One might expect social threat detection mechanisms to be sensitive to reliable indicators of coalitional threat, such as dominance hierarchies, signals of group membership and the cohesiveness of rival coalitions [38,67] and, accordingly, experimental evidence shows that exposing people to these different forms of coalitional threat does increase the tendency to make paranoid attributions [69,70].

Nevertheless, paranoia often involves the selective identification of a (seemingly arbitrary) group of persecutors, where malign intent is attributed to some individuals (or groups) but not others (e.g. 'I'm being persecuted by the CIA' [and not FBI] or 'I'm being persecuted by my family' [but not my neighbours]'). We suggest that this arbitrary selectivity might reflect the fact that coalition boundaries in human groups are themselves highly fluid and flexible and can be formed in the absence of any stable group identifiers [71]. The fact that coalitions can be formed on the basis of minimal cues or markers of similarity in turn selects for cognitive machinery that readily and flexibly categorizes people into groups on the basis of such 'minimal' cues[72,73]. Indeed, humans readily form and detect minimal groups, even from a young age[73] and the perception of these groups fundamentally alters expectations about the intentions and behaviour of individuals within them (reviewed in [74]). Assuming that paranoia builds on this existing cognitive machinery helps to explain the seemingly arbitrary selectivity in the identification of perceived persecutors. This raises an interesting theoretical question as to the extent to which increasingly severe paranoia reflects variation in cognitive processes involved in perceiving coalitions and alliances, as opposed to processes involved in the attribution of (harmful) intent to others. We suggest that disambiguating these processes and how they vary across the paranoia spectrum will be a fruitful avenue for further research.

A coalitional psychology model of paranoia

A coalitional perspective suggests that variation in paranoia could function to protect individuals from coalitionary threat in specific contexts and therefore serve an adaptive function when either the probability and/or the costs of harm from others are high. A prediction of this hypothesis is therefore

that variation in paranoid thinking will reflect the background probability and/or costs of coalitional conflict. Epidemiological evidence supports this prediction: an increased tendency for paranoid thinking has been documented in general population groups that are involved in higher-than-average rates of coalitionary aggression, such as gang members [75] and army veterans [76,77]. The probability of inter-coalitionary violence is increased under conditions of resource scarcity[78] and, as expected, living in poverty is also associated with increased tendency for paranoid thinking[79].

Variation in paranoia should also be sensitive to the perceived costs of receiving inter-coalitionary aggression, which escalate with low coalitionary support, low social rank or increasing power imbalances between coalitions[80,81]. In support of this prediction, risk for psychosis (for which paranoia is the most common delusional theme) is higher among people who have small social networks [82] or who are socially isolated, both of which are proxies for low coalitionary support. Epidemiological evidence supports the idea that perceived power imbalances can raise the risk for psychosis and, by extension, can also increase the probability for paranoid thinking. For example, low social rank (both perceived and objective) is an important predictor for increased paranoia [83] – a finding that has recently been supported by experimental work where participants' social status relative to that of a partner was experimentally manipulated [69]. Similarly, being part of a marginalised social group (e.g. a low status immigrant, or an ethnic minority) is a risk factor for paranoia[84], which can be ameliorated by living in increased densities within the marginalised group[85]. A coalitional psychology perspective on paranoia would predict this otherwise paradoxical 'ethnic density effect' since living at higher ethnic densities with perceived coalition members should be associated with an increased perception of coalitionary support.

Paranoia also varies within individuals and is fine-tuned to the degree of coalitional threat in the current interaction. For example, experimental work where people interact with a political affiliate or with a political adversary shows that harmful intent attributions, the fundamental component of live paranoid ideation (Box 4) are stronger for the dissimilar than for the similar interaction partner, as expected[69]. Paranoid thinking should also respond flexibly to the cohesiveness of coalitions since cohesive coalitions are more able to work together to harm rivals[38]. As expected, recent work has shown that paranoid attributions increase when participants interact with a cohesive pair of opponents compared to a pair of non-cohesive opponents[70]. Thus, observational and experimental evidence suggests that paranoid thinking is flexible and responsive to social context in both the short and long-term, as would be expected if paranoia is the output of a mechanism for detecting and avoiding coalitionary threat.

Paranoia across the lifespan

Paranoia also varies widely across the lifespan, emerging in adolescence, being most pronounced in early adulthood [86] and declining as individuals age [22]. Indeed, if paranoia is an output of a coalitional psychology, then its emergence should coincide with onset of coalitional threat. Empirical evidence suggests that coalitional competition begins to emerge when individuals reach puberty and is most intense during late adolescence and early adulthood[87]. Competition during adolescence may play an important role in the formation of and integration into coalitions that ultimately determine individuals' status, access to resources (including mates) and reproductive success. In modern tribal societies, such as the *Nyangatom*, men form close alliances with same-age individuals during adolescence. It is also at this time that men begin to join lethal raiding excursions to neighbouring groups (usually with members of their coalition), continuing to participate in these raids until they end their reproductive careers (c. age 45 [88]). More generally, interaction with peers increases markedly during adolescence[89], leading also to an increase in social competition at this age. For example, bullying – which can be construed as a form of coalitional competition - is prevalent across all world cultures (and also in pre-industrialised societies) and increases in frequency as children enter adolescence[90], peaking around the age of 14 [91]. Other work has shown that adolescence is a period that is characterised by increased sensitivity to social threat, social risks and social exclusion[92–94], as well as being a common onset period for many mental health problems, including psychotic-spectrum disorders[86,95]. Thus, we suggest that the developmental trajectory of paranoia reflects a selective process that balances sensitivity to threat in line with fitness-relevant outcomes.

Individuals may also experience sensitive periods during development, where cues from the (social) environment exert exaggerated effects on subsequent development. Sensitive periods are expected to evolve whenever the early environment can reliably predict future conditions and when there are constraints on plasticity[96]. The conditions experienced during a sensitive period of development can act as a 'weather forecast', guiding subsequent development along different trajectories and generating adaptive matches between the environment and the individual's phenotype[96–99] . It has been suggested that adolescence could be one such sensitive period in development[96,100,101], with the evolutionary relevance being that individuals receive more reliable cues about the kind of social world they will inhabit and their place in it during adolescence than earlier in development (see[96]). One of the key outstanding questions with respect to paranoia will be to determine whether social threat shapes responses across the lifetime, or whether there are sensitive periods of development during which exposure to social threat exerts lasting consequences on social cognition and behaviour. If the latter, then identifying when these sensitive periods are and how they vary in response to the stochasticity of the social environment (e.g. [102,103]) will also be fruitful.

When does paranoia become pathological?

Having argued so far in favour of viewing variation in paranoia as part of a normally-functioning, naturally selected human psychology, we now address the question of when paranoia might be best viewed as a disorder and, therefore, under negative selection. The definition of mental disorder is historically controversial and beyond the scope of this article: here we adopt the 'harmful dysfunction' definition proposed by Wakefield[104] which states that a) mental disorders are conditions that cause harm to the person as judged by the standards of the person's culture, and b) that the condition results from the inability of some internal mechanisms (psychological or physiological) to perform its natural function, wherein a natural function is an effect that is part of the evolutionary explanation of the existence and structure of the mechanism. Importantly, as with many other biological continuities (e.g. weight), it may be difficult (if not impossible) to provide precise cut-offs that demarcate the boundary between ordered and disordered paranoia[105] without needing to deny clear pathology within this range.

An analogy may be helpful: fever helps the body fight off pathogens and can therefore be viewed as part of a normally-functioning body's evolved responses to infection. Nevertheless, the underlying mechanisms regulating temperature can become impaired or fail, leading to increasingly dysregulated fever that can sometimes be fatal. Clearly, in the latter case, fever would be viewed as pathological (i.e. disordered) despite that fact that, under normal circumstances, fever is an adaptive response to infection. Based on this logic, we suggest that as paranoia becomes increasingly severe and therefore less responsive to threat in the immediate environment, it is increasingly likely to stem from dysfunction in the underlying cognitive mechanisms that support threat evaluation and so is likely to fit the definition of disorder (being, by implication, maladaptive). We remain agnostic about the precise cut-off point for separating ordered from disordered paranoia, as well as about the magnitude and linearity / non-linearity of fitness costs involved.

At this point however, it is also instructive to raise another question. Paranoia is increased by a wide range of brain injuries and impairments, including substance use, sleep deprivation, illness, traumatic head injury, and dementia: do these impairments imply that the resulting paranoia is necessarily disordered? We argue that it need not be the case. Rather, we suggest that it is possible that increased paranoia in response to brain impairment reflects the correct functioning of a 'cognitive failsafe' because cognitive impairment renders people at higher risk of being exploited by others whom were previously allies or makes them less able to incur the costs of being exploited (e.g. see[106,107]) and therefore a bias toward developing paranoia, rather than other socio-affective states, after impairment may have a protective effect. We note that an important disadvantage of this bias may be that it makes the person less likely to trust others who may provide help but we hypothesise that, on average, this could be protective given the potential catastrophic consequences of exploitation, historically high rates of exploitation of impaired individuals, and the fact that many acute stage impairments and

consequent periods of paranoia often improve naturally over time. Therefore, such a cognitive failsafe might constitute an adaptive response rather than a disorder, although theoretical and empirical data are needed to disambiguate these possibilities. Nevertheless, following the fever analogy above, this hypothesis allows that in some individual contexts, impairments to the mechanisms of the cognitive failsafe can lead to increasingly severe and disordered paranoia, resulting in worse or even catastrophic outcomes for an individual.

To conclude, we argue that an evolutionary approach can help make sense of otherwise puzzling features of paranoia. These include a population continuum of paranoia that includes both context-sensitive paranoid thinking and inflexible, unlikely paranoid delusions, as well as the tendency to selectively identify seemingly arbitrary groups of persecutors, and to perceive that one is the target of conspiracy. We also note that our approach highlights some key areas of future research. The first is on the phenomenology of paranoia and we suggest that the content of delusions in severe paranoia should often reflect common sources of coalitionary threat (e.g. coordinated groups and cliques, higher status individuals, physical harm, threats to reputation). For some individuals, different threats may be more salient or more likely and this might well be reflected in the content of delusions across individuals (e.g. see[108]). Secondly, we suggest additional focus is needed on how people perceive social groups, including processes relating to identification with in-group and categorising others as out-group, and how these processes may be altered in people experiencing severe paranoia. We also note that paranoia has received surprisingly little attention from evolutionary scientists in comparison to other psychiatric difficulties and we hope it becomes of further interest in the field, given its clear relevant to fitness concerns, its diverse presentation and ubiquity in human history.

Box 1. Proximate and Ultimate level explanations

It is worth clearly delineating between proximate and ultimate levels of explanation. In evolutionary biology, an answer to the question of 'why' an individual behaves in a certain way can take two broad, non-mutually exclusive forms: proximate and ultimate level explanations[109–112]. Ultimate level explanations provide the answer to 'why' the behaviour exists: they describe the function of the behaviour in question and show how such behaviour, on average, is associated with fitness increases. Proximate level explanations, on the other hand, are concerned with 'how' the behaviour is implemented. For example, proximate level explanations could describe the psychological mechanisms that support or constrain the behaviour but could also include the hormonal or physiological basis of behaviour. For example, one might answer the question of why a lioness chases a zebra by saying that the lioness needs to eat and is motivated by hunger, or that she has babies to feed, or that she is joining the other lionesses in the pride in the hunt – these would all be valid proximate-level explanations. An ultimate level explanation for hunting behaviour is that lionesses who attempt to hunt and kill prey have more surviving offspring than those who do not partake in

hunting and so this behaviour has been selected for in lion populations over evolutionary time. Clearly, the two explanations are not mutually exclusive. However, a proximate level answer cannot be posed as the solution to an ultimate question of why behaviour exists.

Box 2. Which features of paranoia are unique to humans and why?

Evidence for the sort of inter-coalition competition that we propose results in selective pressure for variation in paranoia is also present for other species, raising the question of to what extent features of paranoia may be present in non-humans animals. For example, lethal intergroup competition in the form of lethal raiding occurs also in chimpanzees[62], and more subtle forms of coalitional competition have also been observed in many other social non-human species (see[61] for a review). There is also convincing evidence for variation in social anxiety in non-human species[54]. However, we would argue that the key cognitive mechanism that underlies the ability for paranoid thinking: namely the ability to reason about unobservable causal mechanisms to explain why events have occurred in the past or might occur in the future seems to be, for the most part, unique to humans[113]. Additionally, the most complex forms of coordination and conspiracy are likely to rely on capacities for language and communication that are not present in any non-human species. It is possible that the ability to attribute intentions to others (also key in paranoia and arguably absent in non-human species[114]) might represent an instantiation of this ability for inferential causal reasoning, albeit one that is specific to the social domain[115]. The question of what selective pressures are most likely to have favoured the human-specific propensity to seek diagnostic causal explanations for phenomena humans is hotly debated (see[115,116]) and a full discussion is beyond the scope of this article. Specifically, it remains an open question whether the human tendency to seek and draw causal inferences evolved in response to social selection pressures, or whether this is more likely to have evolved in response to ecological selection pressures, being subsequently co-opted and used in the social domain.

Box 3. Error-management theory

Error management theory[117] also conceptualised in evolutionary medicine as the 'smoke detector principle'[53] states that the existence of asymmetric error costs can favour the evolution of strategies that err on the side of caution, thereby protecting individuals from catastrophic errors, and may be presented as cognitive biases – that is, psychological mechanisms that result in inaccurate perceptions of the true environment but that can shape behaviour in on-average beneficial ways (see[118–121] for discussion). For example, it may be better to mistake a stick for a snake, than a snake for a stick, because the latter mistake is more likely to be fatal. False alarms of this sort are abundant in nature, in humans and non-human species[37,52]. Crucially, selection is not expected to produce perfectly optimal behaviour under all circumstances but rather to produce strategies that are on average successful over the lifetime and within a population. From an evolutionary perspective, many behavioural 'mistakes' (mistaking sticks for snakes) would be permitted under a broadly adaptive strategy of 'all snake-

shaped things should be initially treated as if they could be snakes'. The strength of such biases (whether behavioural or cognitive) should therefore reflect the asymmetry in error costs: the greater the risk that one error type will produce a catastrophic outcome in comparison to the other, the more likely individuals are to be biased towards making the least costly of the error types. Nevertheless, it is worth noting a shortcoming in the typical application of error management theory to paranoia: in social groups, the asymmetric costs in terms of misperceiving social motivations may depend on context[52]. The costs of wrongly treating someone as trustworthy who actually wants to do you harm may be severe. However, the costs of wrongly treating a coalition member as untrustworthy may also be severe due to the fact non-cooperation often results in reciprocal defection[122], punishment[123,124], or exclusion[125,126]. Indeed, mistakenly treating others as if they might harm you can jeopardize the future of potentially mutually-beneficial partnerships, to the extent that the costs associated with such errors have been posited as the basis for the extraordinarily high levels of human trust and cooperation in seemingly anonymous, one-shot interactions (when the potential for cheating and being exploited is rife)[127] (but see[128]). So, while it may be adaptive to consistently err on the side of misperceiving a snake for a stick – as in the traditional formulation of error management theory – the costs are highly asymmetric in comparison to human threat examples in large part because you cannot form a coalition with a snake or incorrectly reject it as an ally. Importantly, the exact distribution of cost asymmetry that drives selection in these situations is an empirical question and it is possible that the costs of under-perceiving hostile intent in others is still on overage higher than the costs of over-perceiving hostile intent in allies. However, the fact that the latter is well-established as having costs in human social groups suggests that cost asymmetry will not mirror contexts that are most commonly cited as selective pressures that drive the evolution of cognitive biases (sticks, snakes etc).

Box 4. Measuring paranoia in experiments involving genuine social interactions

Paranoia by definition affects how we form and update impressions of others in social interactions. It is therefore instructive to attempt to measure paranoia in settings where participants experience genuine social interactions with others. Game theory tasks – typically used in experimental and behavioural economics - provide many paradigmatic examples of stylized social interactions that can be used to infer or measure social behaviour and preferences and these tasks are now being used to great effect to better understand how social cognition and behaviour vary in paranoia[69,129–132]. Many game theoretic tasks operationalise pro-social behaviour as the willingness to forego financial earnings in the task in order to benefit the partner(s) in the interaction. Games can be one-shot or repeated, occur among pairs or groups of individuals and allow for various forms of social behaviour, including cooperation and punishment. In particular, many game theoretic tasks allow us to measure paranoid attributions since the motives underpinning the decisions to cooperate or not in these tasks are often murky. Consider, for example, the Dictator Game[133]. In this two-player game, one person (the 'dictator') is given a sum of money and can choose whether to send some to the partner (the

429 'receiver') or to keep all the money for themselves. The receiver has no active role in this game and
430 must accept whatever share the dictator offers. Importantly, the motives underpinning a dictator's
431 decision to keep all the money are ambiguous. One might infer that the dictator is motivated by greed
432 (or self-interest). Alternatively, one might also infer that the dictator is motivated by a desire to deny
433 the receiver any money (i.e. intent to harm). Inferring harmful intent in such an interaction is a
434 reliable proxy for paranoid thinking and, in a series of studies using participants from the general
435 population[69,70,131], it has been shown that people who have higher tendency for paranoid thinking
436 make stronger harmful intent attributions in these tasks. The degree to which individuals attribute
437 harmful intent to others in turn predicts their willingness to punish their interaction partners[132].
438

Acknowledgements

NR is funded by a Royal Society University Research Fellowship; VB is supported by a Wellcome Trust Seed Award in Science [200589/Z/16/Z]. Thanks to Louise Barrett and Willem Frankenhuis for helpful comments on an earlier draft. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

447 **References**

448 1. Dunbar, RIM & Shultz, S. Why are there so many explanations for primate brain
449 evolution? *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20160244 (2017).
450 2. Holekamp, KE & Benson-Amram, S. The evolution of intelligence in mammalian
451 carnivores. *Interface Focus* **7**, 20160108 (2017).
452 3. Emery, NJ., Seed, AM., von Bayern, AMP & Clayton, NS. Cognitive adaptations of social
453 bonding in birds. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 489–505 (2007).
454 4. Holekamp, KE., Dantzer, B., Stricker, G., Yoshida, KCS & Benson-Amram, S. Brains,
455 brawn and sociality: a hyaena's tale. *Anim. Behav.* **103**, 237–248 (2015).
456 5. De Waal, FBM. *Chimpanzee Politics*. (2007).
457 6. Clayton, NS & Emery, NJ. The social life of corvids. *Curr. Biol.* **17**, R652–R656 (2007).
458 7. Paz-y-Mino, G., Bond, AB., Kamil, AC & Balda, RP. Pinyon jays use transitive inference
459 to predict social dominance. *Nature* **430**, 778–781 (2004).
460 8. Bergman, TJ., Beehner, JC., Cheney, DL & Seyfarth, RM. Hierarchical classification by
461 rank and kinship in baboons. *Science* **302**, 1234–1236 (2003).
462 9. Krupenye, C., Kano, F., Hirata, S., Call, J. & Tomasello, M. Great apes anticipate that
463 other individuals will act according to false beliefs. *Science* **354**, 110–114 (2016).
464 10. Clayton, NS. Social cognition by food caching corvids: the western scrub-jay as a
465 natural psychologist. *Philos. Trans. R. Soc. B Biol. Sci.* **362,** 507-522 (2007).
466 11. Green, MJ & Phillips, ML. Social threat perception and the evolution of paranoia.
467 *Neurosci. Biobehav. Rev.* **28**, 333–342 (2004).
468 12. Veras, AB. *et al.* Paranoid delusional disorder follows social anxiety disorder in a long-
469 term case series: evolutionary perspective. *J. Nerv. Ment. Dis.* **203**, 477–479 (2015).

13. Freeman, D. & Garety, PA. Comments on the content of persecutory delusions: Does the definition need clarification? *Br. J. Clin. Psychol.* **39**, 407–414 (2000).
14. Freeman, PD. Persecutory delusions: a cognitive perspective on understanding and treatment. *Lancet Psychiatry* **3**, 685–692 (2016).
15. Jack, AH. & Egan, V. Childhood Bullying, Paranoid Thinking and the Misappraisal of Social Threat: Trouble at School. *School Ment. Health* **10**, 26–34 (2017).
16. Bebbington, PE. *et al.* The structure of paranoia in the general population. *Br. J. Psychiatry* **202**, 419–427 (2013).
17. Bell, V. & O'Driscoll, C. The network structure of paranoia in the general population. *Soc. Psychiatry Psychiatr. Epidemiol.* **53**, 737–744 (2018).
18. Elahi, A., Algorta, GP., Varese, F., McIntyre, JC & Bentall, RP. Do paranoid delusions exist on a continuum with subclinical paranoia? A multi-method taxometric study. *Schizophr. Res.* **190**, 77–81 (2017).
19. van Os, J., Linscott, RJ., Myin-Germeys, I., Delespaul, P. & Krabbendam, L. A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness–persistence–impairment model of psychotic disorder. *Psychol. Med.* **39**, 179–195 (2009).
20. Taylor, MJ., Freeman, D. & Ronald, A. Dimensional psychotic experiences in adolescence: Evidence from a taxometric study of a community-based sample. *Psychiatry Res.* **241**, 35–42 (2016).
21. David, AS. Why we need more debate on whether psychotic symptoms lie on a continuum with normality. *Psychol. Med.* **40**, 1935–1942 (2010).
22. Freeman, D. *et al.* Concomitants of paranoia in the general population. *Psychol. Med.* **41**, 923–936 (2011).

23. Startup, H., Freeman, D. & Garety, PA. Persecutory delusions and catastrophic worry in psychosis: Developing the understanding of delusion distress and persistence. *Behav. Res. Ther.* **45**, 523–537 (2007).

24. Kahn-Greene, ET., Killgore, DB., Kamimori, GH., Balkin, TJ. & Killgore, WDS. The effects of sleep deprivation on symptoms of psychopathology in healthy adults. *Sleep Med.* **8**, 215–221 (2007).

25. Catone, G., Marwaha, S., Kuipers, E. & Lennox, B. Bullying victimisation and risk of psychotic phenomena: analyses of British national survey data. *Lancet Psychiatry* **2**, 618–624 (2015).

26. Bird, JC., Waite, F., Rowsell, E., Fergusson, EC. & Freeman, D. Cognitive, affective, and social factors maintaining paranoia in adolescents with mental health problems: A longitudinal study. *Psychiatry Res.* **257**, 34–39 (2017).

27. Bentall, RP., Wickham, S., Shevlin, M. & Varese, F. Do Specific Early-Life Adversities Lead to Specific Symptoms of Psychosis? A Study from the 2007 The Adult Psychiatric Morbidity Survey. *Schizophr. Bull.* **38**, 734–740 (2012).

28. McLean, BF., Mattiske, JK. & Balzan, RP. Association of the Jumping to Conclusions and Evidence Integration Biases With Delusions in Psychosis: A Detailed Meta-analysis. *Schizophr. Bull.* **43**, 344–354 (2017).

29. Buchy, L., Woodward, T. & Liotti, M. A cognitive bias against disconfirmatory evidence (BADE) is associated with schizotypy. *Schizophr. Res.* **90**, 334–337 (2007).

30. Elliott, B., Joyce, E. & Shorvon, S. Delusions, illusions and hallucinations in epilepsy: 2. Complex phenomena and psychosis. *Epilepsy Res.* **85**, 172–186 (2009).

31. Fujii, D. & Ahmed, I. Characteristics of Psychotic Disorder Due to Traumatic Brain Injury. *J. Neuropsychiatry Clin. Neurosci.* **14**, 130–140 (2002).

32. Koponen, S. *et al.* Axis I and II Psychiatric Disorders After Traumatic Brain Injury: A 30-Year Follow-Up Study. *Am. J. Psychiatry* **159**, 1315–1321 (2002).
33. Van Assche, L. *et al.* The Neuropsychological Profile and Phenomenology of Late Onset Psychosis: A Cross-sectional Study on the Differential Diagnosis of Very-Late-Onset Schizophrenia-Like Psychosis, Dementia with Lewy Bodies and Alzheimer's Type Dementia with Psychosis. *Arch. Clin. Neuropsychol.* **10**, 229 (2018).
34. Bersani, G. & Prevete, E. Novel psychoactive substances (NPS) use in severe mental illness (SMI) patients: Potential changes in the phenomenology of psychiatric diseases. *Hum. Psychopharmacol. Clin. Exp.* **32**, e2591 (2017).
35. McKetin, R., Baker, AL., Dawe, S., Voce, A. & Lubman, DI. Differences in the symptom profile of methamphetamine-related psychosis and primary psychotic disorders. *Psychiatry Res.* **251**, 349–354 (2017).
36. Quinn, CA., Wilson, H., Cockshaw, W., Barkus, E. & Hides, L. Development and validation of the cannabis experiences questionnaire – Intoxication effects checklist (CEQ-I) short form. *Schizophr. Res.* **189**, 91–96 (2017).
37. Nesse, RM. *Evolutionary Psychology and Mental Health*. (John Wiley & Sons, Inc., 2015). doi:10.1002/9781119125563.evpsych243
38. Boyer, P., Firat, R. & van Leeuwen, F. Safety, Threat, and Stress in Intergroup Relations. *Perspect. Psychol. Sci.* **10**, 434–450 (2015).
39. Oliver, JE. & Wood, TJ. Conspiracy Theories and the Paranoid Style(s) of Mass Opinion. *Am. J. Polit. Sci.* **58**, 952–966 (2014).
40. van Prooijen, JW. & van Vugt, M. Conspiracy Theories: Evolved Functions and Psychological Mechanisms. *Perspect. Psychol. Sci.* 1745691618774270 (2018). doi:10.1177/1745691618774270

41. Douglas, KM., Sutton, RM. & Cichocka, A. The Psychology of Conspiracy Theories. *Curr. Dir. Psychol. Sci.* **26**, 538–542 (2017).
42. Harper, DJ. Histories of suspicion in a time of conspiracy: a reflection on Aubrey Lewis's history of paranoia. *Hist. Hum. Sci.* **7**, 89–109 (1994).
43. Andreasen, NC. *SAPS - Scale for the Assessment of Positive Symptoms*. (University of Iowa, 1984).
44. Oyebode, F. *Sims' Symptoms in the Mind*. (Saunders Elsevier, 2008).
45. Cameron, N. The Paranoid Pseudo-Community Revisited. *Am. J. Sociol.* **65**, 52–58 (1959).
46. Kim, KI. *et al.* Schizophrenic Delusions Among Koreans, Korean-Chinese and Chinese: a Transcultural Study. *Int. J. Soc. Psychiatry* **39**, 190–199 (1993).
47. Stompe, T. *et al.* Comparison of Delusions among Schizophrenics in Austria and in Pakistan. *Psychopathology* **32**, 225–234 (1999).
48. Green, C. *et al.* Content and affect in persecutory delusions. *Br. J. Clin. Psychol.* **45**, 561–577 (2010).
49. Boissy, A. Fear and Fearfulness in Animals. *Q. Rev. Biol.* **70**, 165–191 (1995).
50. Smith, BR. & Blumstein, DT. Fitness consequences of personality: a meta-analysis. *Behav. Ecol.* **19**, 448–455 (2008).
51. Sih, A. & Del Giudice, M. Linking behavioural syndromes and cognition: a behavioural ecology perspective. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 2762–2772 (2012).
52. Haselton, MG. & Nettle, D. The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases. *Personal. Soc. Psychol. Rev.* **10**, 47–66 (2006).
53. Nesse, RM. The Smoke Detector Principle. *Ann. N. Y. Acad. Sci.* **935**, 75–85 (2001).

54. Brosnan, SF., Tone, EB. & Williams, L. The Evolution of Social Anxiety. in *The Evolution of Psychopathology* 93–116 (Springer International Publishing, 2017). doi:10.1007/978-3-319-60576-0_4

55. Miloyan, B., Bulley, A. & Suddendorf, T. Episodic foresight and anxiety: Proximate and ultimate perspectives. *Br. J. Clin. Psychol.* **55**, 4–22 (2016).

56. Miloyan, B., Bulley, A. & Suddendorf, T. Anxiety: Here and Beyond. *Emot. Rev.* **10**, 175407391773857 (2018).

57. Rodebaugh, TL., Klein, SR., Yarkoni, T. & Langer, JK. Measuring social anxiety related interpersonal constraint with the flexible iterated prisoner's dilemma. *J. Anxiety Disord.* **25**, 427–436 (2011).

58. Rodebaugh, TL. *et al.* The behavioral economics of social anxiety disorder reveal a robust effect for interpersonal traits. *Behav. Res. Ther.* **95**, 139–147 (2017).

59. Tone, EB. *et al.* Social Anxiety and Social Behavior: A Test of Predictions From an Evolutionary Model. *Clin. Psychol. Sci.* 2167702618794923 (2018). doi:10.1177/2167702618794923

60. Harcourt, AH. & de Waal, F. *Coalitions and Alliances in Humans and Other Animals*. (Oxford University Press, 1992).

61. Bissonnette, A. *et al.* Coalitions in theory and reality: a review of pertinent variables and processes. *Behaviour* **152**, 1–56 (2015).

62. Wrangham, RW. & Glowacki, L. Intergroup Aggression in Chimpanzees and War in Nomadic Hunter-Gatherers. *Hum. Nat.* **23**, 5–29 (2012).

63. Gershman, B. Witchcraft beliefs and the erosion of social capital: Evidence from Sub-Saharan Africa and beyond. *J. Dev. Econ.* **120**, 182–208 (2016).

64. Mace, R. *et al.* Population structured by witchcraft beliefs. *Nat. Hum. Behav.* **2**, 39–44 (2018).
65. Vaillancourt, T. Do human females use indirect aggression as an intrasexual competition strategy? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20130080–20130080 (2013).
66. Hess, NH. & Hagen, EH. Sex differences in indirect aggression: Psychological evidence from young adults. *Evol. Hum. Behav.* **27**, 231–245 (2006).
67. Tooby, J. & Cosmides, L. *Groups in Mind: The Coalitional Roots of War and Morality*. (Macmillan Education UK, 2010). doi:10.1007/978-1-137-05001-4_8
68. Pietraszewski, D. How the mind sees coalitional and group conflict: the evolutionary invariances of n -person conflict dynamics. *Evol. Hum. Behav.* **37**, 470–480 (2016).
69. Saalfeld, V., Ramadan, Z., Bell, V. & Raihani, NJ. Experimentally induced social threat increases paranoid thinking. *R. Soc. Open Sci.* **5**, 180569 (2018).
70. Greenburgh, A., Bell, V. & Raihani, NJ. PsyArXiv Preprints | Paranoia and conspiracy: group cohesion increases harmful intent attribution in the Trust Game. *psyarxiv.com*
71. Tajfel, H. & Turner, J. An integrative theory of intergroup conflict. in *The social Psychology of intergroup relations* (eds. Austin, W. & Worchel, S.) 33–48 (1979).
72. Liberman, Z., Woodward, AL. & Kinzler, KD. The Origins of Social Categorization. *Trends Cogn. Sci.* **21**, 556–568 (2017).
73. Dunham, Y. Mere Membership. *Trends Cogn. Sci.* **22**, 780–793 (2018).
74. Otten, S. The Minimal Group Paradigm and its maximal impact in research on social categorization. *Curr. Opin. Psychol.* **11**, 85–89 (2016).
75. Wood, J. & Dennard, S. Gang Membership: Links to Violence Exposure, Paranoia, PTSD, Anxiety, and Forced Control of Behavior in Prison. *Psychiatry* **80**, 30–41 (2017).

612 76. Pizarro, J., Silver, RC. & Prause, J. Physical and Mental Health Costs of Traumatic War
613 Experiences Among Civil War Veterans. *Arch. Gen. Psychiatry* **63**, 193–200 (2006).
614 77. Kaštelan, A. *et al.* Psychotic Symptoms in Combat-Related Post-Traumatic Stress
615 Disorder. *Mil. Med.* **172**, 273–277 (2007).
616 78. Ember, CR., Adem, TA. & Skoggard, I. Risk, Uncertainty, and Violence in Eastern Africa.
617 *Hum. Nat.* **24**, 33–58 (2012).
618 79. Anderson, F. & Freeman, D. Socioeconomic Status and Paranoia. *J. Nerv. Ment. Dis.*
619 **201**, 698–702 (2013).
620 80. Wrangham, RW. Evolution of coalitionary killing. *Am. J. Phys. Anthropol.* **110**, 1–30
621 (1999).
622 81. Johnson, DDP. & MacKay, NJ. Fight the power: Lanchester's laws of combat in human
623 evolution. *Evol. Hum. Behav.* **36**, 152–163 (2015).
624 82. Gayer-Anderson, C. & Morgan, C. Social networks, support and early psychosis: a
625 systematic review. *Epidemiol. Psychiatr. Sci.* **22**, 131–146 (2013).
626 83. Wickham, S., Taylor, P., Shevlin, M. & Bentall, RP. The Impact of Social Deprivation on
627 Paranoia, Hallucinations, Mania and Depression: The Role of Discrimination Social
628 Support, Stress and Trust. *PLoS ONE* **9**, e105140 (2014).
629 84. Shaikh, M. *et al.* Perceived ethnic discrimination and persecutory paranoia in
630 individuals at ultra-high risk for psychosis. *Psychiatry Res.* **241**, 309–314 (2016).
631 85. Bosqui, TJ., Hoy, K. & Shannon, C. A systematic review and meta-analysis of the ethnic
632 density effect in psychotic disorders. *Soc. Psychiatry Psychiatr. Epidemiol.* **49**, 519–529
633 (2014).
634 86. Kessler, RC. *et al.* Age of onset of mental disorders: a review of recent literature. *Curr.*
635 *Opin. Psychiatry* **20**, 359–364 (2007).

87. Geary, D. Evolution and development of boys' social behavior. *Dev. Rev.* **23**, 444–470 (2003).
88. Glowacki, L. *et al.* Formation of raiding parties for intergroup violence is mediated by social network structure. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12114–12119 (2016).
89. Del Giudice, M., Angeleri, R. & Manera, V. The juvenile transition: A developmental switch point in human life history. *Dev. Rev.* **29**, 1–31 (2009).
90. Cook, CR., Williams, KR., Guerra, NG., Kim, TE. & Sadek, S. Predictors of bullying and victimization in childhood and adolescence: A meta-analytic investigation. *Sch. Psychol. Q.* **25**, 65–83 (2010).
91. Volk, AA., Camilleri, JA., Dane, AV. & Marini, ZA. Is Adolescent Bullying an Evolutionary Adaptation? *Aggress. Behav.* **38**, 222–238 (2012).
92. Blakemore, SJ. Avoiding Social Risk in Adolescence. *Curr. Dir. Psychol. Sci.* 096372141773814 (2018). doi:10.1177/0963721417738144
93. Spielberg, JM., Olino, TM., Forbes, EE. & Dahl, RE. Exciting fear in adolescence: Does pubertal development alter threat processing? *Dev. Cogn. Neurosci.* **8**, 86–95 (2014).
94. Silk, JS. *et al.* Increased neural response to peer rejection associated with adolescent depression and pubertal development. *Soc. Cogn. Affect. Neurosci.* **9**, 1798–1807 (2014).
95. Paus, T., Keshavan, M. & Giedd, JN. Why do many psychiatric disorders emerge during adolescence? *Nat. Rev. Neurosci.* **9**, 947–957 (2008).
96. Fawcett, TW. & Frankenhuis, WE. Adaptive explanations for sensitive windows in development. *Front. Zool.* **12 Suppl 1**, S3 (2015).
97. Frankenhuis, WE. & de Weerth, C. Does Early-Life Exposure to Stress Shape or Impair Cognition? *Curr. Dir. Psychol. Sci.* **22**, 407–412 (2013).

98. Frankenhuis, WE., Nettle, D. & McNamara, JM. Echoes of Early Life: Recent Insights From Mathematical Modeling. *Child Dev.* **6**, 769–15 (2018).
99. English, S., Browning, LE. & Raihani, NJ. Developmental plasticity and social specialization in cooperative societies. *Anim. Behav.* **106**, 37–42 (2015).
100. Blakemore, SJ. Development of the social brain during adolescence. *Q. J. Exp. Psychol.* **61**, 40–49 (2008).
101. Fuhrmann, D., Knoll, LJ. & Blakemore, SJ. Adolescence as a Sensitive Period of Brain Development. *Trends Cogn. Sci.* **19**, 558–566 (2015).
102. Panchanathan, K. & Frankenhuis, WE. The evolution of sensitive periods in a model of incremental development. *Proc. R. Soc. Lond. B Biol. Sci.* **283**, 20152439 (2016).
103. Frankenhuis, WE. & Panchanathan, K. Individual Differences in Developmental Plasticity May Result From Stochastic Sampling. *Perspect. Psychol. Sci.* **6**, 336–347 (2011).
104. Wakefield, JC. The concept of Mental Disorder. *calstatela.edu* **42**, 373–388 (1992).
105. Wakefield, JC. Evolutionary versus prototype analyses of the concept of disorder. *J. Abnorm. Psychol.* **108**, 374–399 (1999).
106. Bateson, M., Brilot, B. & Nettle, D. Anxiety: An Evolutionary Approach. *Can. J. Psychiatry* **56**, 707–715 (2011).
107. Nettle, D. & Bateson, M. The Evolutionary Origins of Mood and Its Disorders. *Curr. Biol.* **22**, R712–R721 (2012).
108. Campbell, MM. *et al.* The content of delusions in a sample of South African Xhosa people with schizophrenia. *BMC Psychiatry* **17**, 41 (2017).
109. Tinbergen, N. On aims and methods of Ethology. *Ethology* **20**, 410–433 (1963).
110. Mayr, E. Cause and Effect in Biology. *Science* **134**, 1501–1506 (1961).

111. Mayr, E. Proximate and ultimate causations. *Biol. Philos.* **8**, 93–94 (1993).
112. Scott-Phillips, TC., Dickins, TE. & West, SA. Evolutionary Theory and the Ultimate-Proximate Distinction in the Human Behavioral Sciences. *Perspect. Psychol. Sci.* **6**, 38–47 (2011).
113. Penn, DC. & Povinelli, DJ. Causal cognition in human and nonhuman animals: A comparative, critical review. *Annu. Rev. Psychol.* **58**, 97–118 (2007).
114. Penn, DC. & Povinelli, DJ. On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 731–744 (2007).
115. Penn, DC., Holyoak, KJ. & Povinelli, DJ. Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* **31**, 109–130 (2008).
116. Stuart-Fox, M. The origins of causal cognition in early hominins. *Biol. Philos.* **30**, 247–266 (2014).
117. Johnson, DDP., Blumstein, DT., Fowler, JH. & Haselton, MG. The evolution of error: error management, cognitive constraints, and adaptive decision-making biases. *Trends Ecol. Evol.* **28**, 474–481 (2013).
118. McKay, R. & Efferson, C. The subtleties of error management. *Evol. Hum. Behav.* **31**, 309–319 (2010).
119. McNamara, JM., Trimmer, PC., Eriksson, A., Marshall, JAR. & Houston, AI. Environmental variability can select for optimism or pessimism. *Ecol. Lett.* **14**, 58–62 (2011).
120. Marshall, JAR., Trimmer, PC., Houston, AI. & McNamara, JM. On evolutionary explanations of cognitive biases. *Trends Ecol. Evol.* **28**, 469–473 (2013).

121. Trimmer, PC. Optimistic and realistic perspectives on cognitive biases. *Curr. Opin. Behav. Sci.* **12**, 37–43 (2016).
122. McCullough, ME., Kurzban, R. & Tabak, BA. Cognitive systems for revenge and forgiveness. *Behav. Brain Sci.* **36**, 1–15 (2013).
123. Clutton-Brock, TH. & Parker, GA. Punishment in animal societies. *Nature* **373**, 209–216 (1995).
124. Raihani, NJ., Thornton, A. & Bshary, R. Punishment and cooperation in nature. *Trends Ecol. Evol.* **27**, 288–295 (2012).
125. Robertson, TE., Delton, AW., Klein, SB., Cosmides, L. & Tooby, J. Keeping the benefits of group cooperation: domain-specific responses to distinct causes of social exclusion. *Evol. Hum. Behav.* **35**, 472–480 (2014).
126. Feinberg, M., Willer, R. & Schultz, M. Gossip and Ostracism Promote Cooperation in Groups. *Psychol. Sci.* **25**, 656–664 (2014).
127. Delton, AW., Krasnow, M., Cosmides, L. & Tooby, J. Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13335–13340 (2011).
128. Zimmermann, J. & Efferson, C. One-shot reciprocity under error management is unbiased and fragile. *Evol. Hum. Behav.* **38**, 39–47 (2017).
129. Ellett, L., Allen-Crooks, R., Stevens, A. & Wildschut, T. A paradigm for the study of paranoia in the general population: The Prisoner's Dilemma Game. *Cogn. Ldots* **27**, 53–62 (2013).
130. Fett, A., Shergill, SS., Joyce, DW., Riedl, A. & Strobel, M. To trust or not to trust: the dynamics of social interaction in psychosis. *Brain* **135**, 976–984 (2012).

131. Raihani, NJ. & Bell, V. Paranoia and the social representation of others: a large-scale game theory approach. *Sci. Rep.* **7**, 4544 (2017).

132. Raihani, NJ. & Bell, V. Conflict and cooperation in paranoia: a large-scale behavioural experiment. *Psychol. Med.* **76**, 1–11 (2017).

133. Kahneman, D., Knetsch, JL. & Thaler, R. Fairness as a Constraint on Profit Seeking: Entitlements in the Market. *Am. Econ. Rev.* **76**, 728–741 (1986).

Competing Interests statement

The authors declare no competing interests.