What motivates avoidance in paranoia? Three failures to find a betrayal aversion effect

Anna Greenburgh, Joe M. Barnby, Raphaëlle Delpech, Adam Kenny, Vaughan Bell, Nichola Raihani

1. Introduction

Paranoia, or the exaggerated belief that others intend you harm, has been robustly associated with heightened social avoidance, isolation and social anxiety (Freeman et al., 2007; Freeman et al., 2008; Freeman & Garety, 1999; Gayer-Anderson & Morgan, 2013; Gilbert, Boxall, Cheung, & Iorns, 2005; Lim, Gleeson, Alvarez-Jimenez, & Penn, 2018; Martin & Penn, 2001). Paranoia has been conceptualised in terms of ‘reduced trust’ (Fett et al., 2016; Martinez, Agostini, Al-Suhibani, & Bentall, 2020) but recent evidence has suggested that a reduced tendency to commit resources to others may be motivated by additional components including an increased concern about losses, low motivation, and altered subjective reward from social interactions (Gromann et al., 2013; Raihani & Bell, 2018; Raihani, Martinez-Gatell, Bell, & Foulkes, 2021).

Increased social avoidance is a reliable and disabling feature of paranoia (Martin & Penn, 2001; Murphy, Goodall, & Woodrow, 2020) but may be similarly underpinned by multiple components. The distinction between avoidance driven by a tendency to want to avoid taking risks (risk aversion) and the heightened sensitivity to losses once experienced (loss aversion) has been well-characterised in the cognitive psychology literature (Sokol-Hessner & Rutledge, 2019). In the clinical literature, these components seem to be separable in important ways. For example, people with anxiety disorders show avoidance driven by risk aversion rather than loss aversion (Charpentier, Aylward, Roiser, & Robinson, 2017; Ernst et al., 2014).

Taking a similar multi-component approach to social avoidance in paranoia, individuals may avoid situations because of an increased perception of the danger of material losses from social situations compared to non-social situations, but also because of the subjective experience of loss might be amplified when it is caused socially, compared to non-socially. Paranoia has been shown to involve an increased expectation of harm (e.g. Bennett & Corcoran, 2010; Freeman et al., 2013; So et al., 2020) but research in this area has focused on behavioural or inferential approaches to avoidance and risk that do not
distinguish between the potentially separable components that drive these concerns.

One challenge in testing whether social losses are experienced more negatively than non-social losses is that it requires a paradigm that controls the level of material risk across social and non-social conditions, to ensure that risk perception and subjective experience of loss are not confounded. One approach that is able to test this is the betrayal aversion paradigm from experimental economics (Bohnet & Zeckhauser, 2004).

Studies on participants from the general population have found that individuals are more averse to entering risky interactions when outcomes are determined by other people rather than non-social lottery mechanisms, even when the chance of a fair outcome is known to be the same across these two settings. This phenomenon is called “betrayal aversion”, and indicates that people have an intrinsic disutility to being harmed by other people rather than by random processes (Bohnet & Zeckhauser, 2004). Betrayal aversion has been found to varying degrees across cultures, in both between- and within-subjects designs (Aimone, Ball, & King-Casas, 2015; Bohnet, Greig, Herrmann, & Zeckhauser, 2008), and in non-economic behavioural contexts (Driacoll, Barclay, & Fenske, 2017) although see Petchenaker, Lang, Ehlebrachet, Schlöser, & Dunning, 2020 for a recent null finding.

The established betrayal aversion paradigm (Aimone et al., 2015; Aimone & Houser, 2012; Bohnet & Zeckhauser, 2004) isolates a specific cause of social avoidance. In the classic paradigm, participants have the option to either enter or avoid an interaction where a lottery or another participant will determine their outcome. They are asked to state what minimum probability of the interaction having a fair outcome they would require to enter the interaction. Betrayal aversion is the difference in the participant’s reported minimum acceptable probability of a fair outcome in the two conditions. Importantly, the participant is informed that the chance of a fair outcome is the same in both conditions. Betrayal aversion therefore cannot be attributed to altered risk perception as the outcomes are equal across social and non-social settings. Rather, betrayal aversion measures social avoidance that is purely attributable to social harm aversion (compared to non-social harm aversion).

Using the betrayal aversion paradigm in controlled experimental conditions, we can examine the extent to which social avoidance in paranoia is attributable to differences in social harm aversion rather than non-social harm aversion. That is, we can test if paranoia is associated with a bias toward avoiding harm caused by a social partner (compared to harm caused by a non-social process) when the material costs are the same. Economic paradigms have been used extensively to examine social cognition across the paranoia continuum (e.g. Barnby, Bell, Mehta, & Moutoussis, 2020; Fett et al., 2016; Greenburgh, Bell, & Raihani, 2021; Gromann et al., 2013; Raihani & Bell, 2018; Savulich et al., 2018).

Given the high level of interpersonal sensitivity in paranoia (Bebbington et al., 2013; Bell and O’Driscol, 2018) we predicted that betrayal aversion would increase with paranoia. Namely, the extent to which socially-mediated negative outcomes are experienced as aversive may increase with paranoia, leading more paranoid individuals to selectively avoid interactions involving social rather than non-social harm with equal risk of material losses in both conditions.

We ran three studies to test whether betrayal aversion was associated with paranoia. Two of these studies used classical betrayal aversion paradigms concerning economic choices, and the third used a modified paradigm with non-economic choices.

2. Method

This study was approved by the UCL Research Ethics Committee (project number 3720–002). All participants were recruited via the online platform, Prolific Academic (hereafter ‘Prolific’, http://www.prolific.ac) and took part on a voluntary basis. Data were collected in March (study 1), October (study 2), and November (study 3), 2020. We used Prolific’s screening tools to recruit participants from the UK who were fluent in English. In all studies, participants were compensated at least in line with minimum wage for their time. Sample size was determined and pre-registered before any data analysis. See SI for full study materials including game instructions for the three studies.

2.1. Participants

For study 1, we recruited 1743 participants (72% female; mean age = 37, sd = 12.5). For study 2, we recruited a new sample of 690 participants, (65% female; mean age = 37, sd = 13). For study 3, we recruited a sub-sample of the individuals who had taken part in study 2 on a first-come, first-served basis. We successfully recalled 400 of the 690 participants above (64% female; mean age = 37, sd = 14). In summary, study 1’s participants were entirely distinct from those in study 2 and 3, however participants in study 3 had all taken part in study 2 one month earlier.

2.2. Procedure – study 1 & 2

2.2.1. Paranoia

All participants were initially asked to complete a measure of trait paranoia: the Revised Green et al. Paranoid Thoughts Scale (R-GPTS, Freeman et al., 2021). We used the persecution subscale (Part B) of the R-GPTS in our main analyses. For study 2, participants also completed a measure of general cognitive function at this time point (International Cognitive Ability Resource, ICAR).

2.2.2. Betrayal aversion

Approximately one week after completing the paranoia survey, participants were recalled to take part in the betrayal aversion experiments. Both studies followed a within-subjects design, as described in Aimone et al. (2015). All participants took part in a modified trust game (social risk framing) and a lottery-based game (non-social risk framing). In each study, the order of the two games was counterbalanced between participants.

Study 2 was a replication study of study 1, but with task instructions made more explicit (to ensure comprehension that the probability of a fair outcome was the same in both tasks), and one additional manipulation check to measure comprehension of this probability structure.

2.2.2.1. Social framing. The modified trust games closely followed the design of the classic trust game (Berg, Dickhaut, & McCabe, 1995). Participants played as “investors” matched against a “receiver”. Receiver responses were pre-collected from a separate pool of participants who took part in this paradigm in February 2020, and their decisions were used to determine the investor payoffs, as described below. By using real participants as receivers in this game, participants playing as investors (i.e. participants of interest) were required to base their expectations on the behaviour of other real players: they were told that the players they interacted with in this game had already made their decisions.

Participants playing as “investors” could choose whether to trust the receiver and enter the social interaction, or not to trust the receiver and therefore avoid the social interaction. If the investor trusted the receiver, the receiver’s pre-collected decision to either betray or reciprocate the investor’s trust was enacted. If the receiver betrayed the investor, then the investor received £0.15 and the receiver received £0.85. If the receiver reciprocated the investor’s trust, then the investor and the receiver received £0.50 each. If the investor did not trust the receiver and therefore avoided the interaction, both players received £0.25. Therefore, investors could potentially earn more money by trusting their partner, but only if the partner was trustworthy. Interacting with an untrustworthy partner yielded lower payoffs than avoiding the social interaction.
interaction.

Participants were asked to give the minimum probability of being paired with a reciprocating partner that they would require if they were to trust this receiver (minimum acceptable probability, MAP). This probability was used to determine whether the participant (playing as the investor) entered or avoided the interaction with the receiver. If MAP was below or equal to the true percentage of reciprocating participants the pool of receivers (P), then the “trust” option was selected, and the participant entered the interaction with the receiver. If MAP was above the true percentage of reciprocating participants the pool of receivers (P), then the “do not trust” option was selected, and the interaction was avoided. The true percentage of receivers who chose to reciprocate was 50% (P), where this receiver population were real responders selected from the pre-collected sample.

2.2.2.2. Non-social framing. In addition to the social task described above, participants also took part in a non-social ‘lottery’ task. The lottery game had an identical risk-profile to the trust game described above, but the outcomes were determined by a lottery rather than by the decision of a receiver. Specifically, participants could enter or avoid a lottery, which allocated either a fair or unfair outcome to themselves and a new player they were paired with. Participants were asked to give the minimum acceptable probability of the lottery having a fair outcome (MAPL) that they would require, if they were to enter the lottery. As above, if MAPL ≤ P, then the participant entered the lottery, if not they avoided the lottery. The lottery therefore determined the allocation of monetary payoffs between a participant and a partner. Participants were aware that the chance of a good outcome was the same in both the social and the lottery tasks.

2.3. Procedure – study 3

2.3.1. Paranoia

As study 3 re-recruited participants from study 2, R-GPTS data was already available and was not re-collected.

2.3.2. Betrayal aversion

Study 3 was designed by a student team for their undergraduate research project and used slightly different stimuli. The betrayal aversion tasks in Study 3 had similar framing to Studies 1 and 2 but participants’ decisions did not have financial consequences. Instead of financial decisions, the task was framed using a vignette about planting apple trees. As such, study 3 acted as a replication study in a non-incentivised scenario. This was of interest as it may have been that paranoia is associated with sensitivity to being betrayed by others at a relational level that does not involve monetary outcomes. Both social and non-social tasks were designed to closely mirror the structure of the tasks in studies 1 and 2. As with the other studies, study 3 followed a within-subjects design where participants took part in the social and non-social tasks, order counterbalanced between participants.

2.3.2.1. Social framing. The participants began with 25 apples and could choose whether to engage in or avoid a social interaction with another player. Avoiding the interaction meant that both players kept 25 apples each. Trusting the partner meant that the participant trusted the partner with their apples. As above, the partner could reciprocate the participant’s trust by sharing their harvest (both players receive 50 apples overall), or the partner could betray the participant and only return 15 apples to the participant (keeping 85 apples for themselves).

2.3.2.2. Non-social framing. In this task, participants were told they could only eat red apples, whereas their partner could eat blue and red apples. As above, both players started with 25 red apples. The participant could choose whether to plant or keep their own apples. If they kept their own apples, each player would keep 25 apples. If the participant decided to plant their apples their outcome depended on “nature”: they could receive a good outcome (100 red apples grow and each player receives 50 apples each) or a bad outcome (85 blue apples and 15 red apples grow, so the partner only receives 15 whereas the partner receives 85 apples). As in studies 1 and 2, participants made their decision by giving their Minimum Acceptable Probability of a good outcome in each task.

2.4. Manipulation checks

In all three studies, the participants were told that the probability of receiving a fair outcome was the same in both the social risk task and the non-social risk task. By stating that the chance was the same across both tasks, any difference in willingness to accept risk across the two tasks can be attributed to an individual’s expectation of psychological rather than financial harm (Bohnet & Zeckhauser, 2004).

In all three studies, after completing both social and non-social games, all participants were asked whether they thought the probability of a fair outcome was higher in the non-social or the social task, or the same across the two tasks. This manipulation check serves as a check as to whether participants understood the instructions.

As comprehension of the manipulation was lower than expected in study 1, the instructions were made more explicit in studies 2 and 3. To further check comprehension in Study 2, participants were asked to answer the manipulation check both before and after taking part in the tasks.

In study 1, 37% of participants passed the manipulation check after taking part in the tasks. In study 2, 83% of participants passed the manipulation check before taking part in the experiment, and 64% of participants passed afterwards. This increased comprehension rate was unsurprising as the instructions in study 2 were designed to make the manipulation clearer. In study 3, 60% of the sample passed the manipulation check after taking part in the tasks.

We also included a number of other comprehension checks across the three studies. We checked that all results for our main analyses were robust to the exclusion of all non-comprehenders, and report any qualitative differences in results when non-comprehenders were excluded.

For studies 1 and 2, we detected no association between participants’ paranoia score and the tendency to pass the manipulation check (Kruskal-Wallis tests, study 1: H(2) = 0.97, p = 0.62; study 2: H(2) = 5.43, p = 0.07). For study 3, paranoia was negatively associated with passing the manipulation check (Kruskal-Wallis test, H(2) = 7.49, p = 0.02).

2.5. Analyses

Betrayal aversion is indicated by the difference in the risk participants will accept in order to enter the social compared to the non-social interaction. Betrayal aversion for each participant was calculated as follows:

\[ BA_i = MAP_s - MAP_n \]

According to our pre-registration, our analyses varied depending on the skew of our data. In studies 1 and 3, Shapiro-Wilk analyses using the olsr package in R (Hebbali & Hebbali, 2020) indicated violation of normality assumption. Therefore, in these two studies we converted the variable of Betrayal Aversion into a categorical variable (5 levels in study 1, 4 levels in study 3, with >10 observations per level). Our data did not violate assumption of normality in study 2, so we kept Betrayal Aversion as a continuous variable. Consequently, in studies 1 and 3 we conducted two cumulative link models (Christensen & Christensen, 2015); and in study 2 we conducted a generalized linear model (simple linear regression). In all three models, betrayal aversion was the output variable and paranoia, task order, age and gender were model inputs. All continuous input variables were standardized and binary input variables were centred.
We used an information-theoretic approach with multi-model selection and model averaging for all confirmatory regression analyses. This approach is popular in ecology research and is recognised to have many advantages (see Whittingham, Stephens, Bradbury, & Freckleton, 2006 for review). This approach does not employ arbitrary significance levels as used in null hypothesis testing, but rather examines the AICc (Aikake Information Criterion), where lower AICc values indicate a better fit (Grueber, Nakagawa, Laws, & Jamieson, 2011). Analysis using this method proceeds in four steps: 1) a full global model is specified containing all terms of interest, 2) all possible combinations of terms in this model forming all possible subsets of this model are compared, 3) a ‘top model set’ is obtained containing all models within 2 AICc units of the best model, and 4) models in the top model set are averaged to generate model-averaged effect sizes and confidence intervals (Burnham & Anderson, 2004). This approach acknowledges the uncertainty over which model is the ‘best’ model when many models have similar AICc values. Parameter estimates and confidence intervals are reported with the full global model (Galipaud, Gillingham, David, & Dechaume-Moncharmont, 2014). We used package “MuMIn” (for information theoretic model averaging (Barto, 2020). Analyses were conducted in R 4.0.3 (R Core Team, 2020). Model statistics reported are beta coefficients. Visualisations were created with the package ‘ggplot2’ (Wickham, 2016).

All three studies were separately pre-registered and have open code and data (https://osf.io/s2ktv/?view_only=09aa93d7163a4e6392b4151d4cf57011). Analyses conform to those outlined in our preregistration (either in the main hypotheses sections or in the exploratory analyses sections of the pre-registration), unless stated otherwise.

3. Results

3.1. Paranoia

In each study, we recruited participants across a broad spectrum for paranoid thinking (Fig. 1). For study 1, mean persecution subscale score (±sd) was 5.34 ± 7.47 (range: 0–39). For study 2, mean persecution subscale score was 3.90 ± 6.15 (range: 0–33). For study 3, mean persecution subscale score was 3.09 ± 5.33 (range: 0–32). Mean persecution subscale score reported by the authors of the R-GPTS (Freeman et al., 2021) was 15.8 for participants with a diagnosis of psychotic disorder. Unregistered Kendall rank correlations revealed that paranoia was negatively associated with the measure of general cognitive function in study 2 (ICAR; r = −0.12, p < 0.001).

![Fig. 1. Distribution of persecution subscale R-GPTS score in each study. Violin plots, boxplots and raw data points plotted. Boxplots plotted with whiskers extending to +/− 1.5 IQR. Outliers plotted as black points beyond this range.](image-url)

<table>
<thead>
<tr>
<th>Study</th>
<th>Betrayal aversion range</th>
<th>Mean Betrayal aversion (±sd)</th>
<th>% Betrayal avere</th>
<th>% Betrayal neutral</th>
<th>% Betrayal seeking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80–100</td>
<td>7.13 ± 25</td>
<td>55.6</td>
<td>13.7</td>
<td>30.8</td>
</tr>
<tr>
<td>2</td>
<td>70–73</td>
<td>6.64 ± 21</td>
<td>53.9</td>
<td>15.7</td>
<td>30.4</td>
</tr>
<tr>
<td>3</td>
<td>90–70</td>
<td>−1.23 ± 21.1</td>
<td>39.3</td>
<td>19.0</td>
<td>41.8</td>
</tr>
</tbody>
</table>
Table 1b
Percentage of Betrayal averse, Betrayal neutral and Betrayal seeking participants below and above the mean paranoia (R-GPTS, Part B) score of individuals with psychosis (Freeman et al., 2021). Percentages reported to two significant figures for each study.

<table>
<thead>
<tr>
<th>Study</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Above Clinical Mean</td>
<td>Below Clinical Mean</td>
<td>Above Clinical Mean</td>
</tr>
<tr>
<td>% Betrayal averse</td>
<td>56%</td>
<td>55%</td>
<td>48%</td>
</tr>
<tr>
<td>% Betrayal neutral</td>
<td>9.3%</td>
<td>16%</td>
<td>17%</td>
</tr>
<tr>
<td>% Betrayal seeking</td>
<td>33%</td>
<td>30%</td>
<td>33%</td>
</tr>
</tbody>
</table>

Table 2
Summary of distribution of minimum acceptable probabilities in social (MAP_s) and lottery (MAP_l) conditions across the three studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>MAP_s (± sd)</th>
<th>MAP_l (± sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Range</td>
</tr>
<tr>
<td>Study 1</td>
<td>0–100</td>
<td>57.9 ± 20.7</td>
</tr>
<tr>
<td>Study 2</td>
<td>2–100</td>
<td>59.2 ± 19.4</td>
</tr>
<tr>
<td>Study 3</td>
<td>0–100</td>
<td>58.1 ± 19.7</td>
</tr>
</tbody>
</table>

Table 3
Information for the CLM investigating predictors of Betrayal Aversion in study 1. Model averaged estimates, unconditional standard errors, confidence intervals are reported for terms included in the top model set. Reference levels are shown in parentheses.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Unconditional SE</th>
<th>Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task Order (Non-social first − 1)</td>
<td>−0.61</td>
<td>0.09</td>
<td>(−0.79, −0.42)</td>
</tr>
<tr>
<td>Paranoia</td>
<td>0.004</td>
<td>0.02</td>
<td>(−0.04, 0.05)</td>
</tr>
<tr>
<td>Age</td>
<td>0.002</td>
<td>0.02</td>
<td>(−0.04, 0.04)</td>
</tr>
<tr>
<td>Gender (Female − 1)</td>
<td>0.004</td>
<td>0.04</td>
<td>(−0.08, 0.09)</td>
</tr>
</tbody>
</table>

3.2. Betrayal aversion

Betrayal aversion scores can range from −100 to 100. A betrayal aversion score of 100 implies a participant is maximally betrayal averse: they require a 100% probability of a fair outcome before engaging in the social interaction but a 0% probability of a fair outcome before entering the lottery. A betrayal aversion score of −100 implies a participant is maximally betrayal-seeking: they require a 0% chance of a fair outcome in the social interaction and a 100% chance of a fair outcome in the lottery. Means and ranges for betrayal aversion scores, as well as the proportion of betrayal averse, neutral and seeking participants per study are shown in Table 1a; and these proportions as a function of paranoia is reported in Table 1b. The distributions of minimum acceptable probability (MAP) scores for accepting interactions with social partners and lottery (from which betrayal aversion is calculated) are shown in Table 2.

In studies 1 and 2, participants were significantly less willing to enter risky interactions where outcomes were determined by another human (MAP_h), compared to those where outcomes were determined by a lottery (MAP_l) (study 1: paired t-test, t(1742) = 11.88, p < 0.001; study 2: paired t-test, t(687) = 8.29, p < 0.001; both unregistered). Conversely, participants in study 3 were no more willing to enter a risky interaction with a lottery than with another person (paired t-test, t(399) = −1.17; p = 0.24; unregistered) (see Table 2).

Therefore, participants in studies 1 and 2 were betrayal averse but in study 3 they were not betrayal averse (Table 1a). Kendall’s rank correlation statistics to determine the consistency of betrayal aversion and MAP scores across tasks in the sample who took part in study 2 and 3 revealed that MAP_h was significantly correlated between the two tasks ($r = 0.17, p < 0.001$), but MAP_l and betrayal aversion were not (MAP_l: $r = 0.06, p = 0.10$; BA: $r = 0.02, p = 0.59$).

3.3. Betrayal aversion and paranoia

We found no association between betrayal aversion and paranoia in any study (Fig. 3, Tables 3–5. See SI for top model sets and coefficients when re-run excluding non-comprehenders). This main finding is robust to the exclusion of people who failed at least one comprehension check and the manipulation checks in each case. Post hoc regression analyses including a quadratic term for paranoia revealed that there was no non-linear relationship between betrayal aversion and paranoia: the quadratic term did not predict betrayal aversion in any study.

3.4. Betrayal aversion and task order

In each study, we found an effect of task order on betrayal aversion. Specifically, participants who took part in the non-social condition first were more likely to accept risk in the social condition in each study (Tables 3–5).
3.5. Minimum acceptable probability and paranoia

Neither social (MAP_A) nor non-social risk aversion (MAP_B) were associated with paranoia in any study (Table 6; analyses unregistered).

4. Discussion

Across three studies, we explored whether betrayal aversion was more pronounced among people who scored higher for paranoid thinking. Although we detected evidence for betrayal aversion across the sample as a whole in studies 1 and 2, participants were not betrayal averse in study 3. Despite detecting betrayal aversion in the full sample in two studies, our main prediction was not supported: paranoia was not associated with betrayal aversion any of these studies. These results suggest that people scoring high in paranoia do not avoid social interactions due to a greater aversion to being betrayed. However, we also found no evidence for increased risk aversion in paranoia, either in the social or non-social contexts. This null result raises questions about the paradigm used in this study.

We consider two possibilities for these results. Firstly, a lack of betrayal and risk aversion motivating avoidance in paranoia, and secondly, limitations in the capacity of the paradigm used in this study to adequately measure these effects.

There are two possible ways in which betrayal aversion may have shown the null association with paranoia as seen in our results assuming the validity of the paradigm. The first would characterise a “shift” in aversion: aversion to both social (MAP_A) and non-social (MAP_B) situations would be higher (or lower) overall in individuals scoring high in paranoia, but the difference in aversion to these two situations would be the same as those scoring lower in paranoia. Secondly, no shift would be witnessed at all: aversion to social risk is the same across the paranoia spectrum, and aversion to non-social risk is the same across the paranoia spectrum. We note that previous research shows that MAP_B rather than MAP_A is associated with traditional measures of risk preferences, as measured by gambling decisions in the Eckel and Grossman (2002) risk preference task (Aimone et al., 2015). Given that our results show evidence that neither MAP_A nor MAP_B were associated with paranoia in any study, we not only find no relation between paranoia and betrayal aversion, but no relation whatsoever between paranoia and aversion to risk using this paradigm.

We note here that several studies have reported evidence for increased levels of risk perception in clinical and non-clinical paranoia (Bennett & Corcoran, 2010; Bentall et al., 2008; Corcoran et al., 2006; Freeman et al., 2013; Kaney, Bowen-Jones, Dewey, & Bentall, 1997; So et al., 2020). However, increased risk perception (a tendency to perceive risks as larger than they are) is distinct from both risk aversion (a tendency to avoid risk even when the level of risk is perceived to be the same) and betrayal aversion (a tendency to avoid harm caused by people rather than non-social mechanisms, even when the risk of harm is the same). It is therefore possible that avoidance in paranoia may be driven by increased risk perception rather than risk aversion, however we note that other recent work (also using a game-theoretic paradigm) found no
association between paranoia and the expectation that harmful outcomes would occur (Barnby, Greenburgh, Bell, & Raihani, 2021).

We also consider potential limitations of the paradigm in measuring paranoia-relevant motivations for avoidance. Indeed, it is surprising that paranoia was not associated with risk aversion given that increased risk aversion has been reported in anxiety (Admon et al., 2012; Charpentier et al., 2017; Lorian & Grisham, 2011; Maner et al., 2007), schizophrenia (Reddy et al., 2014; Sabater-Grande et al., 2020; although see Yu et al., 2017), delusion proneness (Van der Leer, Hartig, Goldmanis, & McKay, 2015), autism (Gosling & Moutier, 2018), and a personality measure of suspiciousness (Johnson, Rustichini, & MacDonald III, 2009) – all of which commonly co-occur with high levels of paranoia. Further, it was surprising that a social-specific bias in avoidance wasn’t found in corroboration with previous self-report studies (e.g. Martin & Penn, 2001) as well as recent computational results suggesting a hypersensitivity to social information in psychiatric disorders where paranoia is a common feature (Henco et al., 2020).

We note other failures to find a relationship between betrayal aversion and psychopathological traits employing the same betrayal aversion paradigm. Aimone, Ball, and King-Casas (2014) found no correlation between anxiety and betrayal aversion, and no correlation between anxiety and risk preferences. These null results mirror ours, in a sample with a similar level of betrayal aversion to ours (44.6%), although their sample was significantly smaller (n = 55) and laboratory based. It may be that the potential ‘harms’ (in terms of small monetary losses) both in their study and ours may not have been substantial enough to trigger anxiety- or paranoia-relevant avoidance. Equally, the single round nature of the games may not have been sensitive enough compared to multi-round tasks used in previous studies where the stability of preferences can be determined over a greater number of choices (e.g. Charpentier et al., 2017; Gosling & Moutier, 2018; Sabater-Grande et al., 2020). Indeed in a 12 round iterated trust game, Aimone et al. (2014) found that anxiety was associated with a lower growth rate of

Fig. 3. No association between paranoia and betrayal aversion across three studies. Paranoia (measured by the persecution subscale of the R-GPTS) is divided into 5 subgroups, according to thresholds defined by Freeman et al., 2021 (1 = average ideation, 5 = very severe ideation). Betrayal aversion is indicated by more positive betrayal aversion scores, betrayal seeking is indicated by negative betrayal aversion scores.
trust where, when in the role of investor, low anxiety participants increase investments between early and late rounds whereas high anxiety participant do not.

In support of the validity of the paradigm, we replicated an overall betrayal aversion effect in two out of the three studies. Although we recorded a relatively high number of participants who failed the manipulation checks (study 1, 63%; study 2, 17%; study 3, 40%), manipulation check status was included in the analyses and had no effect on outcome. Similarly, although we saw clear order effects – in that betrayal aversion was lower for participants who completed the non-social risk task before the social risk task – the analyses fully controlled for these. Additionally, we replicate the negative relationship between paranoia and general cognitive function found in other general population studies (Freeman et al., 2011; Ibáñez-Casas, Maura, Gutiérrez, & Cervilla, 2021).

On average, we expected participants across the whole sample to be betrayal averse – as evidenced by previous studies (Aimone et al., 2015; Bohnet et al., 2008; Bohnet & Zeckhauser, 2004). Our data partially supported this prediction: in studies 1 and 2 where a monetary incentive was at stake participants were betrayal averse but in study 3 where participants could only gain ‘points’ with no additional value, participants were neutral with respect to whether risk was socially or non-socially determined. The level of betrayal aversion detected in studies 1 and 2 was smaller than in Bohnet and Zeckhauser (2004) and in Bohnet et al. (2008). The difference with these studies depended on the samples investigated, where the greatest contrast was that betrayal aversion in our sample was one-thirds that of the sample from Oman in Bohnet et al. (2008). However, the distribution of betrayal-averse, betrayal-neutral and betrayal-seeking participants in the current study was similar to that in Aimone et al. (2015). This slight discrepancy is likely because our study more closely mirrored that of Aimone et al. than Bohnet et al.: we used a within–subjects instead of a between-subjects design. The null result in study 3 may have stemmed from risk aversion being higher in non-social conditions, such that participants were risk-averse both in social and non-social interactions (see Fig. 2). A recent study involving German participants similarly did not find betrayal aversion in two financially incentivised one-shot paradigms, albeit with smaller sample sizes than our studies 1 and 2 and with similar sample sizes to study 3 (Fetchenhauer, Lang, Ehlebracht, Schlösser, & Dunning, 2020).

Future research could investigate the effect in offline samples and in more affectively engaged situations, for example, with known partners. We note that in the three studies we report participants were matched with anonymous strangers. In line with emerging evidence that social identification varies with paranoia (Greenaway, Haslam, & Bingley, 2019; McIntyre, Wickham, Barr, & Bentall, 2018), and that social threat from familiar others is particularly strongly associated with paranoia (Greenburgh, Bell, & Raihani, 2021), paranoid individuals may show higher betrayal aversion when interacting with familiar (but not unfamiliar) individuals. Additionally, more interactive paradigms could be familiar (but not unfamiliar) individuals. Additionally, more interactive paradigms could be

aversion during the pandemic was heightened, however this pattern does not seem to be immediately identifiable in our data.

To conclude, paranoia was not associated with betrayal aversion across three studies. Further, paranoia was not associated with general risk aversion in either social or non-social interactions. We consider two possibilities: that paranoia is largely motivated by increased risk perception rather than risk aversion or betrayal aversion, or that the paradigm was limited in terms of its ability to trigger betrayal and risk aversion behaviour in paranoia.

Open science practices
All materials data, and code are available at https://osf.io/52kvf/?view_only=09a9a9d7163a4c6392b4151d4cf57011.

Funding
AG is supported by The Royal Society.
JMB is supported by the UK Medical Research Council (MR/N013700/1) and King’s College London member of the MRC Doctoral Training Partnership in Biomedical Sciences.
NR is supported a Royal Society University Research Fellowship and the Leverhulme Trust.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of Competing Interest
Nothing to declare.

References


