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

Anna Greenburgh[a], Joe M Barnby, [b,c], Raphaëlle Delpech [a], Adam Kenny [a], Vaughan Bell [d,e] , Nichola Raihani[a]

Affiliations:

[a] Department of Experimental Psychology, University College London, London, UK.

[b]Cultural and Social Neuroscience Group and [c]Neuropharmacology Group, Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

[d]South London and Maudsley NHS Foundation Trust, London, UK

[e] Research Department of Clinical, Educational, and Health Psychology, University College London, London, UK.

Corresponding author:

Email: a.greenburgh@ucl.ac.uk

Abstract

Believing that others intend to harm you (paranoia) is often accompanied by social withdrawal, avoidance and isolation. We investigated whether paranoia is related to betrayal aversion: the tendency to avoid potential harm caused by other people over and above an equivalent harm caused by a non-social mechanism. Across three large-N (N_{total}=2433) pre-registered online studies, we employed a game theoretic paradigm where participants engaged in interactions with real players. Studies 1 and 2 explored betrayal aversion by eliciting participants' willingness to enter interactions where monetary reward was either determined by another player or a lottery. Study 3 examined betrayal aversion in a context where choices were not financially-incentivised. Paranoia was not associated with betrayal aversion or risk aversion in any study. We consider two possibilities: that paranoia does not involve increased 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.

Key words

Paranoia, betrayal aversion, risk aversion

Introduction

Paranoia, or the exaggerated belief that others intend you harm, has been robustly associated with heightened social avoidance, isolation and social anxiety (Martin & Penn, 2001; Freeman & Garety, 1999; Gilbert et al., 2005; Freeman et al., 2008, 2007; Lim et al., 2018; Gayer-Anderson & Morgan, 2013). Paranoia has been conceptualised in terms of 'reduced trust' (Fett et al, 2016; Martinez et al, 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 and Bell, 2018; Raihani et al, 2021).

Increased social avoidance is a reliable and disabling feature of paranoia (Martin and Penn, 2001; Murphy et al., 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 and 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 et al, 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 perception of the likelihood of negative events and an increased expectation of harm (e.g. So et al, 2020; Freeman et al., 2013; Bennett and Corcoran, 2010) 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 (Bohnet et al., 2008; Aimone et al., 2015), and in non-economic behavioural contexts (Driscoll et al., 2017) although see Fetchenhauer et al (2020) for a recent null finding.

The established betrayal aversion paradigm (Bohnet & Zeckhauser, 2004; Aimone et al., 2012, 2015) 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. Saluvich et al., 2018; Gromann et al., 2013; Fett et al., 2016; Raihani & Bell, 2018, Greenburgh et al., 2019; Barnby et al., 2020).

Given the high level of interpersonal sensitivity in paranoia (Bebbington et al., 2013; Bell and O'Driscoll, 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.

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.

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.

Procedure – study 1 & 2

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., 2019). 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).

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.

Social framing

The modified trust games closely followed the design of the classic trust game (Berg et al., 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.

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_A). This probability was used to determine whether the participant (playing as the investor) entered or avoided the interaction with the receiver. If MAP_A was **below** the true percentage of reciprocating participants the pool of receivers (P_{-I}), then the “trust” option was selected, and the participant entered the interaction with the receiver. If $\mathrm{MAP}_{\mathrm{A}}$ was **above** the true percentage of reciprocating participants the pool of receivers (P_{-i}), 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_{-i}), where this receiver population were real responders selected from the pre-collected sample.

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 (MAP_B) that they would require, if they were to enter the lottery. As above, if $MAP_B < P_{-I}$ 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.

Procedure – Study 3

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

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 money. 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.

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).

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 participant 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.

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 et al., 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 chi squared tests, $p > 0.05$ in both studies). For study 3, paranoia was negatively associated with passing the manipulation check (*Kruskal-Wallis chi-squared* = 7.49, p=0.02).

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:

$$BA_i = MAP_A – MAP_B$$

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 olsrr package in R (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, 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 et al., 2006 for review). This approach does not employ arbitrary significance levels as used in null hypothesis testing, but rather examines the AICc (Aikaike Information Criterion), where lower AICc values indicate a better fit (Grueber et al., 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 and 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 et al., 2014). We used package “MuMIn” (for information theoretic model averaging (Bartoń, 2018). Analyses were conducted in R 4.0.02

(Team R, 2016). 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/s2kvf/?view_only=09aa93d7163a4c6392b4151d4cf57011). 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.

Results

Paranoia

In each study, we recruited participants across a broad spectrum for paranoid thinking (Figure 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; $r_t = -0.12$, $p < .001$).

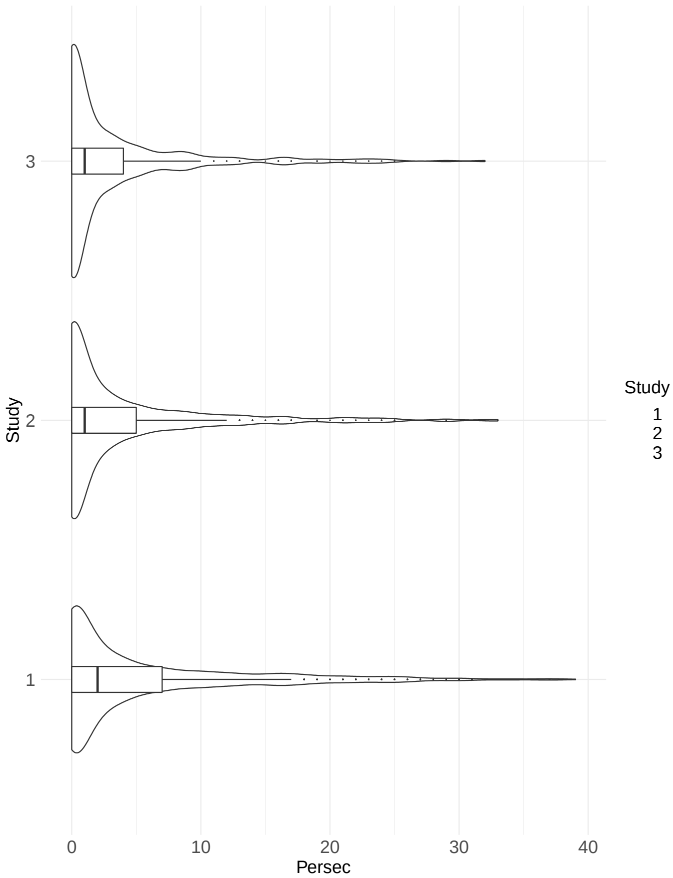

Figure 1. Distribution of persecution subscale R-GPTS score in each study. Violin plots, boxplots and raw data points plotted. Boxplots plotted with whiskers extending to +/- 1.5 IQR. Outliers plotted as black points beyond this range.

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 lotteries (from which betrayal aversion is calculated) are shown in Table 2.

	Betrayal aversion range	Mean Betrayal aversion ((± sd)	% betrayal averse	% betrayal neutral	% betrayal seeking
Study 1	-80 – 100	7.13 ± 25	55.6	13.7	30.8
Study 2	-70 – 73	6.64 ± 21	53.9	15.7	30.4
Study 3	-90 – 70	-1.23 ± 21.1	39.3	19.0	41.8

Table 1a. Summary of distribution of betrayal aversion across the three studies

Study	1		2		3	
	Above Clinical Mean	Below Clinical Mean	Above Clinical Mean	Below Clinical Mean	Above Clinical Mean	Below Clinical Mean
% Betrayal averse	56%	55%	48%	54%	32%	40%
% Betrayal Neutral	9.3%	16%	17%	15%	14%	19%
%Betrayal Seeking	33%	30%	33%	30%	55%	41%

Table 1b. Percentage of Betrayal Averse, Neutral and Seeking participants below and above mean on the persecutory subscale of individuals with psychosis (Freeman et al., 2021), for each study, reported to two significant figures.

	MAP_A range	Mean MAP_A (± sd)	MAP_B range	Mean MAP_B (± sd)
Study 1	0-100	57.9 ± 20.7	0-100	50.8 ± 20.8
Study 2	2-100	59.2 ± 19.4	1-100	52.6 ± 16.9
Study 3	0-100	58.1 ± 19.7	0-100	59.3 ± 19.6

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

In studies 1 and 2, participants were significantly less willing to enter risky interactions where outcomes were determined by another human (MAP_A), compared to those where outcomes were determined by a lottery (MAP_B) (study 1: paired t test, t (1742) = 11.88, p<.001; study 2: paired t test, $t(689)$ = 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_A was significantly correlated between the two tasks (r_t = 0.17, p < .001), but MAP_B and betrayal aversion were not (MAP_B: r_t = 0.06, p = 0.10; BA: r_t = 0.02, p = 0.59).

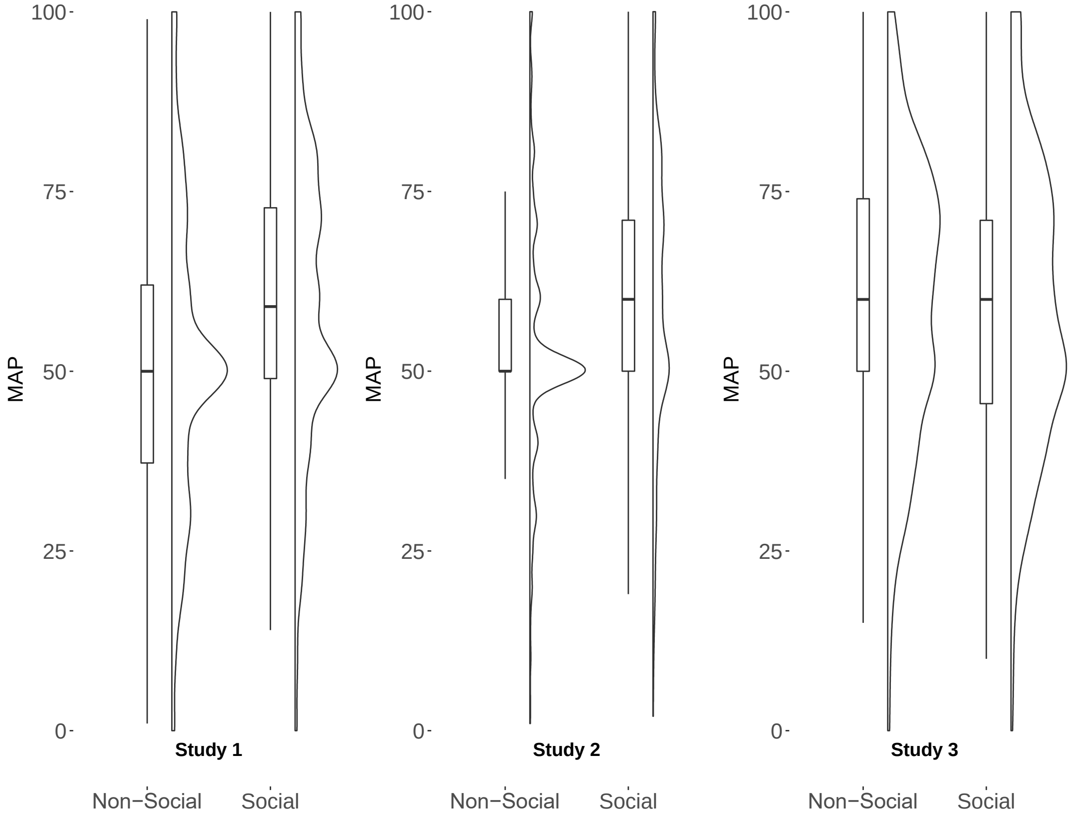

Figure 2. Distribution of Minimum Acceptable Probability (MAP) of a fair outcome in either non-social (red, MAP_B) or social (blue, MAP_A) conditions in all three studies.

Betrayal aversion and paranoia

We found no association between betrayal aversion and paranoia in any study (Figure 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.

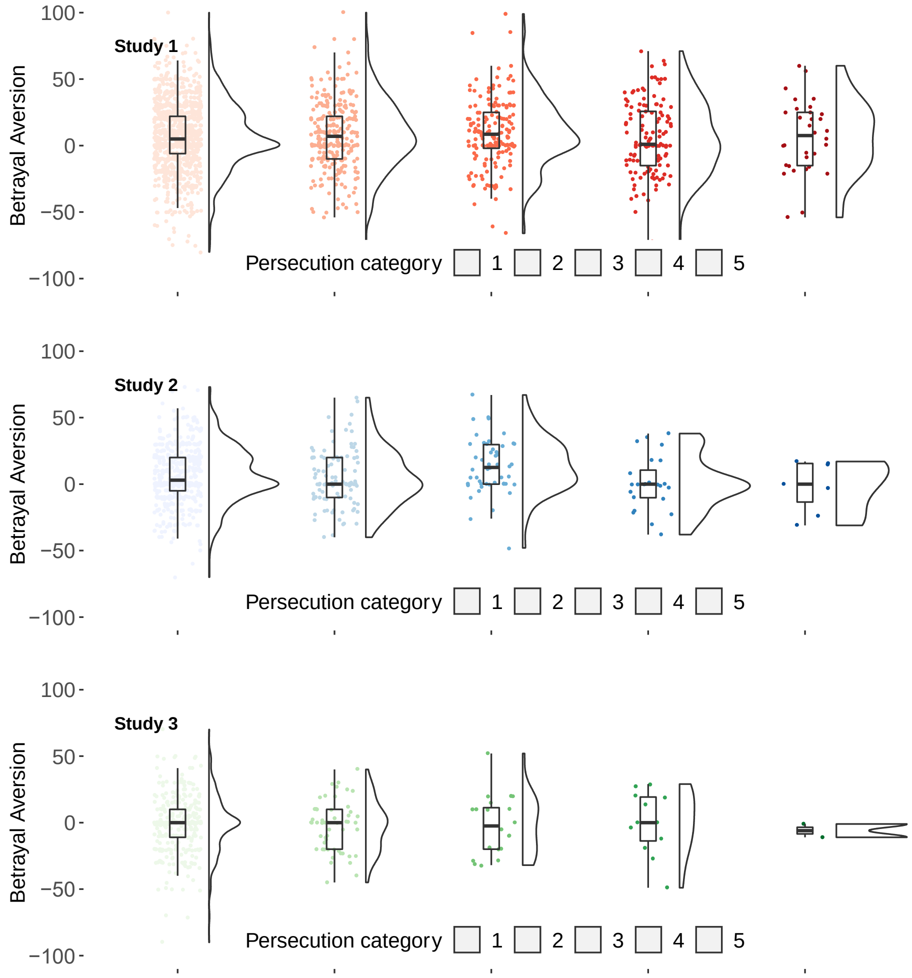

Figure 3. No association between paranoia and betrayal aversion across three studies. Paranoia (measured by the persecution subscale of the R-GPTS) is divided into 5 subgroups, according to thresholds defined by Freeman et al., 2019 (1=average ideation, 5 = very severe ideation). Betrayal aversion is indicated by more positive betrayal aversion scores, betrayal seeking is indicated by negative betrayal aversion scores.

Parameter	Estimate	Unconditional SE	Confidence Interval
Task Order	*-0.61*	*0.09*	*(-0.79, -0.42)*
Paranoia (persecution)	*0.004*	*0.02*	*(-0.04, 0.05)*
Age	*0.002*	*0.02*	*(-0.04, 0.04)*
Gender (Female=1)	*0.004*	*0.04*	*(-0.08, 0.09)*

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

Parameter	Estimate	Unconditional SE	Confidence Interval
Intercept	*6.28*	*1.05*	*(4.21, 8.35)*
Task order	*-6.78*	*2.10*	*(-10.91, -2.64)*
ICAR	*1.88*	*1.19*	*(-0.46, 4.22)*
Paranoia	*0.56*	*1.04*	*(-1.48, 2.59)*
Age	*0.38*	*0.82*	*(-1.23, 1.98)*
Gender (Female=1)	*-0.16*	*0.89*	*(-1.90, 1.58)*

Table 4. Information of the main analysis for study 2 including a measure of cognitive reasoning (ICAR) as a predictor.

Parameter	**Estimate**	**Unconditional SE**	**Confidence Interval**
Task Order	-0.49	0.20	(-0.89, -0.10)
Gender (Female=1)	0.08	0.17	(-0.25, 0.41)

***Table 5*.** Information for the CLM investigating predictors of Betrayal Aversion in study 3. Model averaged estimates, unconditional standard errors, confidence intervals and relative importance for the terms included in the top model set. Reference levels are shown in parentheses.

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).

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).

	Paranoia ~ MAP_A	Paranoia ~ MAP_B
Study 1	$r_S = 0.03$, $p = 0.29$	$r_S = 0.02$, $p = 0.42$
Study 2	$r_s = -0.01$, $p = 0.81$	$r_S = -.002$, $p = 0.96$
Study 3	$r_s = -0.04$, $p = 0.45$	$r_s = -0.02$, $p = 0.67$

Table 6. Results of Spearman correlations between paranoia and minimum acceptable probability for social (MAP_A) and lottery (MAP_B) conditions across the three studies.

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 (Kaney et al., 1997; Corcoran et al., 2006; So et al., 2020; Freeman et al., 2013; Bennett and Corcoran, 2010; Bentall et al., 2008). 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 et al., *in prep).*

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 (Lorian and Grisham, 2011; Charpentier et al, 2017; Maner et al, 2007; Admon et al, 2012), schizophrenia (Reddy et al, 2014; Sabater-Grande et al, 2020; although see Yu et al, 2017), delusion proneness (van der Leer et al, 2015), autism (Gosling and Moutier, 2018), and a personality measure of suspiciousness (Johnson et al, 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 and 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 this computational study tested probabilistic reward learning and therefore employed a vastly different task design to the present study, however another computational study employing a game theoretical paradigm has similarly suggested that paranoia in the general population involves a greater sensitivity to current social context (Barnby et al., 2020).

We note other failures to find a relationship between betrayal aversion and psychopathological traits employing the same betrayal aversion paradigm. Aimone et al., (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, Sabater-Grande et al, 2020; Gosling and Moutier, 2018). Indeed in a 12 round iterated trust game, Aimone et al (2014) found that anxiety was associated with a lower growth rate of 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, replicate the negative relationship between paranoia and general cognitive function found in other general population studies (Freeman et al., 2011; Ibanez-Casas et al., 2021).

On average, we expected participants across the whole sample to be betrayal averse – as evidenced by previous studies (Bohnet and Zeckhauser, 2004; Bohnet et al., 2008; Aimone et al., 2015). 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 figure 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 et al., 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 (McIntyre et al., 2018; Greenaway et al., 2019), and that social threat from familiar others is particularly strongly associated with paranoia (Greenburgh et al., in prep), paranoid individuals may show higher betrayal aversion when interacting with familiar (but not unfamiliar) individuals. Additionally, more interactive paradigms could be employed in future research, such as Cyberball, which has been effectively used to investigate the affective and behavioural consequences of both social and non-social rejection (Driscoll et al., 2017).

While our sample included individuals scoring at clinically-relevant levels of persecutory ideation, we did not collect any information on psychiatric history or diagnoses. It therefore remains a question for future research to determine whether betrayal aversion is higher in individuals from the clinical population, and whether this may relate to high levels of social avoidance. We note that data collection was undertaken during the coronavirus pandemic, where lockdown regulations varied over each data collection time point, as well between regions within the UK where participants were based. It is possible that risk 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/s2kvf/?view_only=09aa93d7163a4c6392b4151d4cf57011.

References

Admon, R., Bleich-Cohen, M., Weizmant, R., Poyurovsky, M., Faragian, S., & Hendler, T. (2012). Functional and structural neural indices of risk aversion in obsessive–compulsive disorder (OCD). *Psychiatry Research: Neuroimaging*, *203*(2-3), 207-213.

Aimone, J. A., & Houser, D. (2012). What you don't know won't hurt you: a laboratory analysis of betrayal aversion. *Experimental Economics*, *15*(4), 571-588.

Aimone, Jason Anthony and Ball, Sheryl B. and King-Casas, Brooks, Anxiety, Risk Preferences, Betrayal Aversion, and the Growth of Interpersonal Trust (February 27, 2014). Available at SSRN: https://ssrn.com/abstract=2402413 or http://dx.doi.org/10.2139/ssrn.2402413

Aimone, J., Ball, S., & King-Casas, B. (2015). The betrayal aversion elicitation task: an individual level betrayal aversion measure. *PloS one*, *10*(9), e0137491.

Bartoń, K. (2020). MuMIn: Multi-Model Inference. R package version 1.43.17. https://CRAN.R-project.org/package=MuMIn

Barnby, J. M., Bell, V., Mehta, M. A., & Moutoussis, M. (2020). Reduction in social learning and increased policy uncertainty about harmful intent is associated with pre-existing paranoid beliefs: Evidence from modelling a modified serial dictator game. *PLoS computational biology*, *16*(10), e1008372.

Barnby, J, Greenburgh, A., Bell, V & Raihani, N (*in prep)*. Cognitive and metacognitive differences in expecting versus explaining harm in paranoia.

Bebbington, P. E., McBride, O., Steel, C., Kuipers, E., Radovanoviĉ, M., Brugha, T., ... & Freeman, D. (2013). The structure of paranoia in the general population. *The British Journal of Psychiatry*, *202*(6), 419-427.

Bennett, K., & Corcoran, R. (2010). Biases in everyday reasoning: Associations with subclinical anxiety, depression and paranoia. *Psychosis*, *2*(3), 227-237.

Bentall, R. P., Rouse, G., Kinderman, P., Blackwood, N., Howard, R., Moore, R., ... & Corcoran, R. (2008). Paranoid delusions in schizophrenia spectrum disorders and depression: the transdiagnostic role of expectations of negative events and negative self-esteem. *The Journal of nervous and mental disease*, *196*(5), 375-383.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, *10*(1), 122-142.

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, *55*(4), 467-484.

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey, and the united states. *American Economic Review*, *98*(1), 294-310.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304.

Charpentier, C. J., Aylward, J., Roiser, J. P., & Robinson, O. J. (2017). Enhanced risk aversion, but not loss aversion, in unmedicated pathological anxiety. *Biological Psychiatry*, *81*(12), 1014-1022.

Christensen, R. H. B., & Christensen, M. R. H. B. (2015). Package ‘ordinal’. *Stand*, *19*, 2016.

Corcoran, R., Cummins, S., Rowse, G., Moore, R., Blackwood, N., Howard, R., ... & Bentall, R. P. (2006). Reasoning under uncertainty: heuristic judgments in patients with persecutory delusions or depression. *Psychological Medicine*, *36*(8), 1109-1118.

Diaconescu, A. O., Wellstein, K. V., Kasper, L., Mathys, C., & Stephan, K. E. (2020). Hierarchical Bayesian models of social inference for probing persecutory delusional ideation. *Journal of Abnormal Psychology*, *129*(6), 556.

Driscoll, R. L., Barclay, P., & Fenske, M. J. (2017). To be spurned no more: The affective and behavioral consequences of social and nonsocial rejection. *Psychonomic bulletin & review*, *24*(2), 566-573.

Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and human behavior*, *23*(4), 281-295.

Ernst, M., Plate, R. C., Carlisi, C. O., Gorodetsky, E., Goldman, D., & Pine, D. S. (2014). Loss aversion and 5HTT gene variants in adolescent anxiety. *Developmental cognitive neuroscience*, *8*, 77-85.

Fetchenhauer, D., Lang, A. S., Ehlebracht, D., Schlösser, T., & Dunning, D. (2020). Does betrayal aversion really guide trust decisions towards strangers?. *Journal of Behavioral Decision Making*.

Fett, A. K., Shergill, S. S., Korver-Nieberg, N., Yakub, F., Gromann, P. M., & Krabbendam, L. (2016). Learning to trust: trust and attachment in early psychosis. *Psychological Medicine*, *46*(7), 1437-1447.

Freeman, D., & Garety, P. A. (1999). Worry, worry processes and dimensions of delusions: an exploratory investigation of a role for anxiety processes in the maintenance of delusional distress. *Behavioural and Cognitive Psychotherapy*, *27*(1), 47-62.

Freeman, D., Garety, P. A., Kuipers, E., Fowler, D., Bebbington, P. E., & Dunn, G. (2007). Acting on persecutory delusions: the importance of safety seeking. *Behaviour research and therapy*, *45*(1), 89-99.

Freeman, D., Gittins, M., Pugh, K., Antley, A., Slater, M., & Dunn, G. (2008). What makes one person paranoid and another person anxious? The differential prediction of social anxiety and persecutory ideation in an experimental situation. *Psychological Medicine*, *38*(8), 1121-1132.

Freeman, D., McManus, S., Brugha, T., Meltzer, H., Jenkins, R., & Bebbington, P. (2011). Concomitants of paranoia in the general population. *Psychological medicine*, *41*(5), 923-936.

Freeman, D., Dunn, G., Fowler, D., Bebbington, P., Kuipers, E., Emsley, R., ... & Garety, P. (2013). Current paranoid thinking in patients with delusions: the presence of cognitive-affective biases. *Schizophrenia bulletin*, *39*(6), 1281-1287.

Freeman, D., Loe, B. S., Kingdon, D., Startup, H., Molodynski, A., Rosebrock, L., ... & Bird, J. C. (2019). The revised Green et al., Paranoid Thoughts Scale (R-GPTS): psychometric properties, severity ranges, and clinical cut-offs. *Psychological Medicine*, 1-10.

Galipaud, M., Gillingham, M. A., David, M., & Dechaume-Moncharmont, F. X. (2014). Ecologists overestimate the importance of predictor variables in model averaging: a plea for cautious interpretations. *Methods in Ecology and Evolution*, *5*(10), 983-991.

Gayer-Anderson, C., & Morgan, C. (2013). Social networks, support and early psychosis: a systematic review. *Epidemiology and psychiatric sciences*, *22*(2), 131.

Gilbert, P., Boxall, M., Cheung, M., & Irons, C. (2005). The relation of paranoid ideation and social anxiety in a mixed clinical population. *Clinical Psychology & Psychotherapy*, *12*(2), 124-133.

Gosling, C. J., & Moutier, S. (2018). Brief report: Risk-aversion and rationality in autism spectrum disorders. *Journal of autism and developmental disorders*, *48*(10), 3623-3628.

Greenaway, K. H., Haslam, S. A., & Bingley, W. (2019). Are “they” out to get me? A social identity model of paranoia. *Group Processes & Intergroup Relations*, *22*(7), 984-1001.

Greenburgh, A., Bell, V. & Raihani, N. (*In prep)*. Paranoia and coalitional psychology: a network study.

Gromann, P. M., Heslenfeld, D. J., Fett, A. K., Joyce, D. W., Shergill, S. S., & Krabbendam, L. (2013). Trust versus paranoia: abnormal response to social reward in psychotic illness. *Brain*, *136*(6), 1968-1975.

Grueber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: challenges and solutions. *Journal of evolutionary biology*, *24*(4), 699-711.

Hebbali, A., & Hebbali, M. A. (2020). Package ‘olsrr’.

Henco, L., Diaconescu, A. O., Lahnakoski, J. M., Brandi, M. L., Hörmann, S., Hennings, J., ... & Mathys, C. (2020). Aberrant computational mechanisms of social learning and decision-making in schizophrenia and borderline personality disorder. *PLoS computational biology*, *16*(9), e1008162.

Ibanez-Casas, I., Maura, C. D. A. C. C., Gutiérrez, B., & Cervilla, J. A. (2021). A population-based cross-sectional study of cognitive deficits in paranoia. *Psychiatry Research*, *299*, 113820.

Johnson, M. K., Rustichini, A., & MacDonald III, A. W. (2009). Suspicious personality predicts behavior on a social decision-making task. *Personality and individual Differences*, *47*(1), 30-35.

Kaney, S., Bowen-Jones, K., Dewey, M. E., & Bentall, R. P. (1997). Two predictions about paranoid ideation: Deluded, depressed and normal participants' subjective frequency and consensus judgments for positive, neutral and negative events. *British Journal of Clinical Psychology*, *36*(3), 349-364.

Lorian, C. N., & Grisham, J. R. (2011). Clinical implications of risk aversion: an online study of risk-avoidance and treatment utilization in pathological anxiety. *Journal of Anxiety Disorders*, *25*(6), 840-848.

Lim, M. H., Gleeson, J. F., Alvarez-Jimenez, M., & Penn, D. L. (2018). Loneliness in psychosis: a systematic review. *Social psychiatry and psychiatric epidemiology*, *53*(3), 221-238.

Maner, J. K., Richey, J. A., Cromer, K., Mallott, M., Lejuez, C. W., Joiner, T. E., & Schmidt, N. B. (2007). Dispositional anxiety and risk-avoidant decision-making. *Personality and Individual Differences*, *42*(4), 665-675.

Martin, J. A., & Penn, D. L. (2001). Social cognition and subclinical paranoid ideation. *British Journal of Clinical Psychology*, *40*(3), 261-265.

Martinez, A. P., Agostini, M., Al-Suhibani, A., & Bentall, R. P. (2020). Mistrust and negative self-esteem: Two paths from attachment styles to paranoia. *Psychology and Psychotherapy: Theory, Research and Practice*.

McIntyre, J. C., Wickham, S., Barr, B., & Bentall, R. P. (2018). Social identity and psychosis: Associations and psychological mechanisms. *Schizophrenia bulletin*, *44*(3), 681-690.

Murphy, R., Goodall, K., & Woodrow, A. (2020). The relationship between attachment insecurity and experiences on the paranoia continuum: A meta-analysis. *British Journal of Clinical Psychology*, *59*(3), 290-318.

Raihani, N. J., & Bell, V. (2018). Conflict and cooperation in paranoia: a large-scale behavioural experiment. *Psychological Medicine*, *48*(9), 1523-1531.

Raihani, N., Martinez-Gatell, D., Bell, V., & Foulkes, L. (2021). Social reward, punishment, and prosociality in paranoia. *Journal of abnormal psychology*.

Reddy, L. F., Lee, J., Davis, M. C., Altshuler, L., Glahn, D. C., Miklowitz, D. J., & Green, M. F. (2014). Impulsivity and risk taking in bipolar disorder and schizophrenia. *Neuropsychopharmacology*, *39*(2), 456-463.

Sabater-Grande, G., Haro, G., García-Gallego, A., Georgantzís, N., Herranz-Zarzoso, N., & Baquero, A. (2020). Risk-taking and fairness among cocaine-dependent patients in dual diagnoses: Schizophrenia and Anti-Social Personality Disorder. *Scientific reports*, *10*(1), 1-13.

Savulich, G., Jeanes, H., Rossides, N., Kaur, S., Zacharia, A., Robbins, T. W., & Sahakian, B. J. (2018). Moral emotions and social economic games in paranoia. *Frontiers in psychiatry*, *9*, 615.

So, S. H. W., Sun, X., Chan, G. H. K., Chan, I. H. H., De Chiu, C., Chan, S. K. W., ... & Chen, E. Y. H. (2020). Risk perception in paranoia and anxiety: Two investigations across clinical and non-clinical populations. *Schizophrenia Research: Cognition*, 100176.

Sokol-Hessner, P., & Rutledge, R. B. (2019). The psychological and neural basis of loss aversion. *Current Directions in Psychological Science*, *28*(1), 20-27.

Van der Leer, L., Hartig, B., Goldmanis, M., & McKay, R. (2015). Delusion proneness and 'jumping to conclusions': relative and absolute effects. *Psychological medicine*, *45*(6), 1253.

Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour?. *Journal of animal ecology*, *75*(5), 1182-1189

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. springer.