|  |
| --- |
|  |
| PS3200: Project |
| **The relationship between “jumping to conclusions” behaviour on the Beads task and the sampling of human faces via an Optimal Stopping Task** |
|  |
| **Word Count: 5935** |
|  |

|  |
| --- |
|  |

**The relationship between “jumping to conclusions” behaviour on the Beads task with the sampling of human faces via an Optimal Stopping Task**

Abstract

The Optimal Stopping Problem occurs when one has to decide when is the best time to stop sampling options in order to gain the best choice option. Previous research shows that in best choice optimal stopping tasks, people undersample and miss out on their best choice option. However, recent research has found that when sampling attractive faces participants oversample in comparison to a Bayesian probabilistic model. The current study aims to replicate the findings of Furl, Averbeck and Mckay’s (In Prep) study and see if Jumping to Conclusions (JTC) behaviour is associated with the sampling of faces. Participants sampled and chose faces in a Fiancé Task, and then completed the Beads Task, a measure of JTC, as well as a measure of delusional traits. Results showed that participants viewed significantly more faces and chose lower ranked faces than the model, but this was not influenced by JTC in the Beads task or by having delusional traits.

Introduction

Humans are constantly searching for their best choice in life that will reap the most benefits. The ability to decide when the optimal time to stop sampling is, can be seen in many everyday instances. These instances can range from deciding when is the optimal time to apply for a low interest mortgage; or deciding whether one should buy an item of clothing, or waiting until the item goes on sale at the risk of the item going out of stock (Costa & Averbeck 2015). The Optimal Stopping Problem occurs when one has to decide whether the future gain outweighs the potential loss due to stopping (van Moerbeke, 1974). In order to gain rewards, humans and animals must expend energy and time on collecting evidence to maximise rewards and minimise the costs of choosing an unfulfilling choice (Furl & Averbeck, 2011). However, collecting more evidence does not necessarily lead to a greater reward as one cannot go back to rejected choices, therefore too much sampling can lead to missing out on the best choice.

When the optimal stopping problem is tested via best choice tasks, it is thought people should be able to choose higher ranked options. However, in Costa and Averbeck’s (2015) study, participants did not choose their best choice. Participants completed a range of best choice tasks, one being a value comparison optimal task, where they were told that they had to choose the best value car based on its price and mileage out of 12 options. Choices were displayed on screen one by one and participants either accepted the choice or rejected it to move onto the next option. The rejected choices also appeared on screen in order for participants to compare values before making the next decision. Despite these value comparison parameters in place, participants undersampled, missing out on their best choice. These results seem to suggest that in haste or worry of missing out on their best choice, they stop sampling too early.

The optimal stopping problem has had many names, depending on the goal of the choosing behaviour, including the best choice problem, the optimal choosing strategy and the Secretary Problem, all comprising the same concept of sampling, stopping and choosing at an optimal time. The Secretary Problem is a scenario where the optimal choice paradigm is applied to recruitment. A company is looking to employ a new secretary undergoes interviewing candidates. The agents must decide when they have found the best candidate for the job and stop sampling, despite having the knowledge that they may miss out on a better candidate. The concept of the secretary scenario and optimal stopping can also be applied in the dating world. How does one know they have found the one and can stop looking for a better partner? This real world problem of finding a mate can be referred to as “The Fiancé Problem”. Especially with the boom in recent years of mobile dating apps such as Tinder which involves swiping left and right on potential dates, choosing a partner is becoming more about strategy than before. Therefore it is important to examine factors which influence our sampling and decision making abilities in making optimal choices.

To test this idea of “The Fiance problem”, Averbeck, Furl and Mckay (In Prep) applied the optimal stopping paradigm to the choosing of faces. Participants completed the computerised Fiance task, and rated the attractiveness of faces on a scale of 1-9 over three phases. The average ratings of the faces displayed created a rank of faces based on how attractive the participants rated them. Then similar to Costa and Averbeck (2015), a sequence task with these faces was created. In sequences of 12, faces from the rating tasks appeared one at a time on screen, which were either accepted as a “date” or rejected to view another face as a potential “date”. The results from the human participants were compared against the results of a Bayesian Ideal Observer Model (Furl & Averbeck, 2011; Costa & Averbeck, 2015). Contrary to previous research on optimal choosing, on multiple trials participants oversampled when sampling faces, viewing too many faces and then missing out on their best option in comparison to the model. Since this is the first study to examine how optimal choosing applies to the sampling of human faces, it is important to understand what factors cause there to be an oversampling effect, rather than the undersampling effect commonly found in other optimal choosing tasks. Are faces simply more interesting or addictive to sample than other items, or is there some other decision making factor that influences the Fiancé task?

There seems to be a human bias for looking at humans and animal faces over non-living items, such as food and scenery. Furl, Gallagher and Averbeck (2012) showed participants multiple affective of: human faces either smiling or angry faces; cute or threatening animals; and food that was either appetising or repulsive whilst completing a decision making task. The study found that participants showed bias for human, then animal faces, as well as a bias for images showing positive affect. This suggests that human and animal faces have a social and emotional valence that non-living items do not have. Therefore, this emotional and social information present in living faces may cause us to pay more attention to them, and explain the oversampling effect with faces.

“Jumping to conclusions” (JTC) is a decision making behaviour where people make hasty decisions based on very little information or evidence. JTC behaviour is commonly found amongst Schizophrenic patients. Schizophrenia is a condition characterised by cognitive dysfunctioning, hallucinations and delusions. Specifically, the delusional symptoms of schizophrenia are consistently linked to JTC behaviour (Evans, Averbeck, Furl, 2015). It has been found that the higher a person scored on delusional traits, the earlier they made a decision on a JTC reasoning task (Van der Leer, Hatig, Goldmanis & McKay, 2015)., JTC is predominantly measured in research using the Beads Task. The Beads task is a probabilistic reasoning task designed to measure individuals’ sampling and decision making behaviour. The participant is shown two jars with complementary ratios of coloured beads inside, for example, 30% green and 70% blue beads in one jar, and 30% blue and 70% green beads in the other. The jars are then hidden from view and the experimenter chooses one jar out of the sight of the participant. One by one, beads are drawn from one jar, and based on the beads the participant sees, they must either choose which jar they think they beads are coming from or draw another bead and repeat the process. Research has shown that delusion prone individuals consistently gather less data before making decisions, therefore making fewer draws before choosing a jar (Garety et al, 2005). Delusional patients tend to make a decision after 1 or 2 draws (Evans, Averbeck & Furl 2015; Garety et al, 2005), whilst the majority of healthy participants make a choice after 5-6 draws (Garety, Hemsley & Wessely, 1991), indicating how in this type of sampling task, those with delusional traits are more likely to show JTC by requiring less information to make decisions.

The Liberal Acceptance Account (Moritz & Woodward, 2005) has been proposed to explain the JTC phenomena in Schizophrenics. The theory proposes that Schizophrenics simply make decisions with less information. However JTC behaviour is only present when there are limited alternatives in the task at hand. The beads task usually only involves two jars, however in versions of the task when there are four jars, the JTC effect disappears (Moritz, Woodward & Lambert, 2007). This suggests that the element of choice removes the chance of JTC behaviour to occur, seeming to support the idea that Schizophrenics JTC when they are given less information. Moreover, due to their delusional beliefs, Schizophrenics have also been found to be less confident when reporting correct answers, and overconfident when reporting errors (Moritz & Woodward, 2002 cited in Evans, Averbeck & Furl, 2015). Specifically, those who score highly on delusions, had stronger convictions for implausible beliefs and also required fewer draws to make a decision on the beads task (McKay, Langdon & Coltheart, 2006). However, despite paying more attention to less valid information, Schizophrenics did not inspect fewer pieces of information compared to non-schizophrenics. Therefore, this could suggest that schizophrenics maybe sample a similar amount to non-schizophrenics, but are more likely to consider irrelevant or less valid evidence when making a decision (Moritz & Woodward, 2006 cited in Evans, Averbeck & Furl, 2015).

JTC is not only found within Schizophrenic patients, it is also found in 20% of the general population, so is also a relevant concept within non-clinical and clinical samples alike (Freeman, Pugh & Garety, 2008). Suboptimal decision making behaviour may not be solely due to JTC, since individuals who score high on delusion proneness perform suboptimally, making earlier decisions based on less evidence (Van der Leer, Hatig, Goldmanis & McKay, 2015). Moreover, using measures of delusion-proneness, predominantly The Peters’ Delusion Inventory (PDI) (Peters, Joseph & Garety, 1999), there is a direct link between PDI scores and JTC behaviour on the beads task. Specifically, high scores on the PDI are associated with participants drawing fewer beads on the beads task. Therefore, the link between delusional symptomology and behaviour on the beads task is prevalent within non-clinical populations as well as within Schizophrenics, indicating that JTC behaviour may be related to other decision making behaviours.

Patient’s decision making deficiencies have also been attributed to poor Theory of Mind. Theory of Mind is the ability to understand that other people have a separate mental state to one’s own mental state. Theory of Mind may be linked to decision making and optimal choosing, because faulty understanding about the motives of other people can lead to suboptimal decisions. Specifically, Schizophrenic patients with persecutory delusions have a poorer Theory of Mind and an impaired ability to infer others intentions (Mehl, Rief, Lüllmann, Ziegler, Kesting & Lincoln, 2010). Moreover, delusion proneness and Theory of Mind deficiencies have been associated with reduced performance on probabilistic reasoning tasks, including the beads task (Langdon, Ward & Coltheart, 2010). This suggests that impairments with Theory of Mind in Schizophrenics are linked to delusional thoughts and JTC behaviour. Although, as a comparison, Autistic individuals show the opposite effect on the beads task to schizophrenic and delusional patients, requiring more draws than controls in the beads task before choosing a jar (Brosnan, Chapman & Ashwin, 2014). Autistic individuals also have an impaired Theory of Mind and struggle to understand the intentions and mental states of others, but unlike Schizophrenia, the disorder is not characterised by delusions. The fact that Autistic probands show the opposite result to schizophrenics, despite both proband groups showing Theory of Mind deficiencies, suggests that the JTC behaviour Schizophrenics show in the beads task may be due to either the interaction between Theory of Mind and delusional traits, or solely delusional traits.

Theory of mind can be seen as part of the link between JTC, delusions and optimal choosing in dating. Having faulty inferences about the motives of others can negatively affect the potential to form and maintain relationships. These faulty inferences are a part of an individual’s propensity to have delusional beliefs. Schizophrenics tend to have a higher propensity of experiencing delusions but delusional beliefs on a lesser scale are also prevalent in the non-clinical population, therefore the JTC behaviour found in high delusion prone individuals during the beads task and other probabilistic reasoning task may also be prevalent in optimal choosing tasks. If delusion prone individuals JTC during sampling in the beads task, then there is a possibility that they also JTC when sampling in other tasks, such as when sampling faces. Therefore, those who show an undersampling JTC effect in the beads task may not show the same oversampling effect when completing the Fiancé task.

The aim of the present study is to see if there is a link between the sampling of faces in an optimal choosing task and JTC behaviour. Consistent with research from Furl, Averbeck and McKay (In Prep), we hypothesise that the participants will oversample and view more faces in the sequence before making a decision, in comparison to the model. Participants will also choose lower ranked faces in the sequence compared to the model. Moreover, there will be a link between delusional traits and JTC, with participants with higher delusional traits displaying JTC behaviour in the beads task, and drawing fewer beads than those with less delusional traits. In order to show the JTC effect in the Fiance task, the number of views in the Fiance task should be related to the number of draws in the beads task. Those who make fewer draws in the beads task should view fewer faces before choosing in the fiancé task. Lastly, we hypothesise a positive association between the average rank of face chosen and the percentage of answers correct on the beads task.

Methods

Participants

Twenty-eight participants were recruited (19 female) from students at Royal Holloway, University of London. Their ages ranged from 18-23 years of age (Mean=20.39, SD=1.23). All participants were entered into a draw to win a £50 amazon voucher. The experiment was approved by the Royal Holloway, University of London ethics committee.

Materials

Peters’ Delusion Inventory (2004)

Peters’ Delusion Inventory (PDI) was designed by Peters and colleagues (1999) in order to measure delusional traits. The original inventory consisted of 40 items; whereas the present study uses the revised inventory which has been condensed down to 21 items (Peters, Joseph, Day & Garety, 2004). Each item is a statement which the participant can answer with a yes or no. If the participant answers yes, they must complete the three Likert subscales within the statement, which measures the degree of the subject’s identification with the statement, for example “Do you ever feel as if some people are not what they seem to be?” The subscales for all 21 items measure: distress, preoccupation and conviction. The Likert scales are scored from 1-5, with 1 as the lowest value meaning either “not at all distressing” (distress), “hardly ever think about it” (preoccupation) or “don’t believe it’s true” (conviction) and 5 as the highest value corresponding with “very distressing” (distress); “think about it all the time” (preoccupation), or “believe it is absolutely true” (conviction). From these scales, four scores are created. The PDI total score out of 21 is created from the initial yes or no answers, with each yes answer scored as 1 and no as 0. The scores from the distress, preoccupation and conviction scores are then totalled from the Likert subscales to create the remaining delusion scores. In order to create a single representative score for the entire questionnaire which can be compared against the other variables, the conviction, preoccupation and distress scores were totalled, creating a single PDI Total.

The beads task

Jumping to conclusions behaviour can be measured using the beads task. The beads task involves the subject being shown two jars, one containing 70% green beads and 30% blue beads, and the other containing 70% blue beads and 30% green beads. The jars are covered from the subjects view and then one jar is chosen by the experimenter. One bead is drawn from the jar and shown to the subject. The subject must then either guess which jar the bead came from or draw another bead. This process is repeated till the participant makes a decision. On average, participants tend to make 5 to 6 draws before making a decision (Garety, Hemsley, & Wessely, 1991) on the beads task, whilst schizophrenic participants with delusional traits tend to make a decision after 1 draw (Evans, Averbeck & Furl, 2015). Jumping to conclusions type behaviour is marked by having an average of two or less draws when performing the beads task.

In the present study, subjects complete the beads task using the computer programme MATLAB (The Mathworks, Natick, MA). The screen says there is a 30/70 ratio of green and blue beads within the jar and then either the word green or blue appeared on the screen. Participants must then press the appropriate numbers on the keyboard to either choose, green, blue or draw another. If they decide to draw another, the next bead colour in the sequence will appear and they must repeat the process again until they make a decision. The participants did 40 trials of the beads task. During analysis, the first 10 trials were omitted to account for any confusion in learning how to perform the task. Previous studies have found that there is a miscomprehension bias, with participants who struggle to understand the task performing 38% of premature decisions (Balzan, Delfabbro & Galletly, 2012). The last 10 trials were also omitted to account for any fatigue effects the subject was experiencing. Therefore 20 trials in total were analysed, accounting for the 20 trials that were removed.

The Fiancé task

The fiancé task was also completed using MatLab. Subjects underwent three rating phases, a distracter task and a sequence choosing task. Before starting the rating phases, the subjects specified the gender they chose to date. Then either male or female faces were displayed according to the individual’s preferences. Individual faces are displayed on the screen and subjects rate the faces for attractiveness from 1-9 (1 being least attractive and 9 the most attractive). Over the phases, faces are displayed three times and average ratings are created for these faces. Once the 3 phases have been completed, subjects complete a distracter task, solving mathematical equations within a 2 minute time limit.

The final phase of the fiancé task is the sequence task. The individual faces from the previous phases are displayed sequentially to the subject in series of 12 faces. A face is displayed on the screen and the subject can either choose that face as a “date” or move onto the next face. They cannot go back to any previous faces, and once they have selected a “date” they cannot see the remaining faces left in the sequence. Also, if no faces are selected, the last face is automatically chosen. The subject completes 6 sequences.

Whilst the subject is completing the fiancé task, the Bayesian model (Furl & Averbeck, 2011; Costa & Averbeck, 2015) is also completing the same task based upon the subject’s rating choices. During the sequence task, the model aims to choose the best choice option as a date.

Procedure

The study was counterbalanced to prevent any order effects from occurring. Therefore half of the total subjects completed the fiancé task first, and then the beads task; meanwhile the second half completed the beads task first and then the fiancé task. In all instances, the PDI was always completed last to prevent any potential social desirability effects from the questionnaire affecting the computerised data.

Design

Analysis was conducted via two one-way repeated measures ANOVAs and two hierarchical regression models. The first ANOVA examined the difference between the average numbers of human (subject) versus model views of faces before choosing a face in the sequence task. The second ANOVA examines the difference between the average ranked face chosen by the human (subject) versus the model. The viewing and ranking behaviour was then compared with the subjects’ performance on the beads task and PDI. The first regression model compared human views with number of draws on the beads task and PDI. The second regression model compared human ranks with percentage correct in the beads task along with PDI. Any significant correlations were analysed for potential interactions with the participant’s sex.

Results

Results from the ANOVA show that the model viewed significantly fewer faces than the subjects (F(1,26)=27.00, p<.001) (see Figure 1) and this was not influenced by the sex of the face (F (1, 26)=.075 , p=.787). The model also chose significantly higher ranked faces than the subjects (F (1, 26) =20.90, p<.001) (see Figure 2), and this also was not significantly affected by the sex of the face (F(1,26)= .001, p=.970).

Figure 2: Human versus Model mean rank of chosen faces

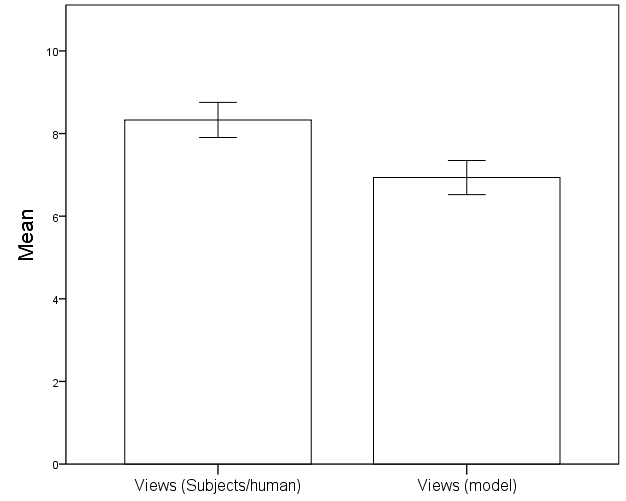
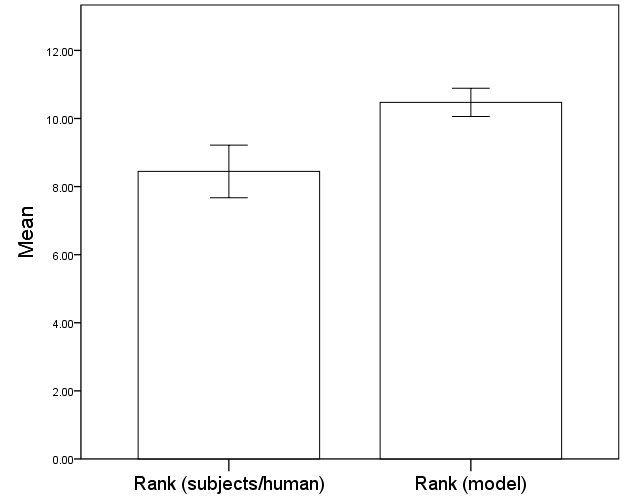


Figure 1: Mean numbers of Human versus Model views before choosing a date

The hierarchical regression model for views, with number of draws, PDI and sex as predictors, explained 14% of the variance in the number of human views, but was not significant ((F(5,22)=1.90 , p=.140) (see Figure 3C). The regression model shows multicollinearity. The overall model is not a significant predictor of human views (F (5,22)=0.4, p=0.32). The remaining predictor variables: number of draws, PDI, and sex, were also not significant. At the next hierarchical level, an interaction of PDI with sex was added to the base model, but this was not significant; thereby indicating that sex had no effect upon the number of views before choosing a face (see Figure 3D). Additionally, an interaction of number of draws and sex were added to the base model, which was also found to be not significant. Therefore, there was no effect of sex of the participant upon their number of draws. As neither additional interactions improved the model, we can conclude the base model of: number of draws, PDI and sex to be the best model to predict the results. If we examine individual correlations, the closest to significance correlation is with average number of bead draws and PDI (r=-.256 , p=.188); all other correlations between predictors had a higher significance.

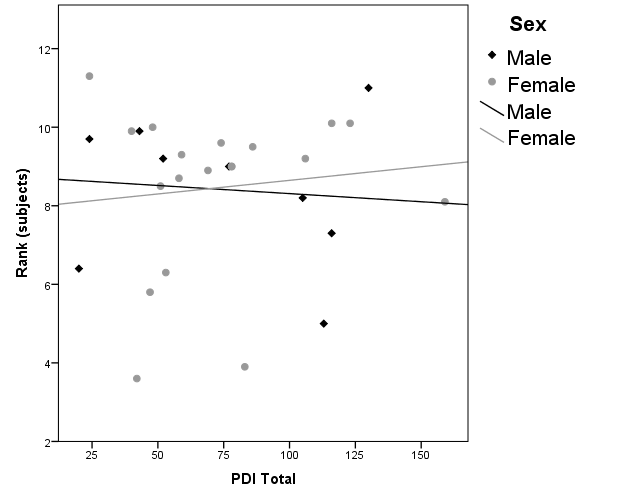
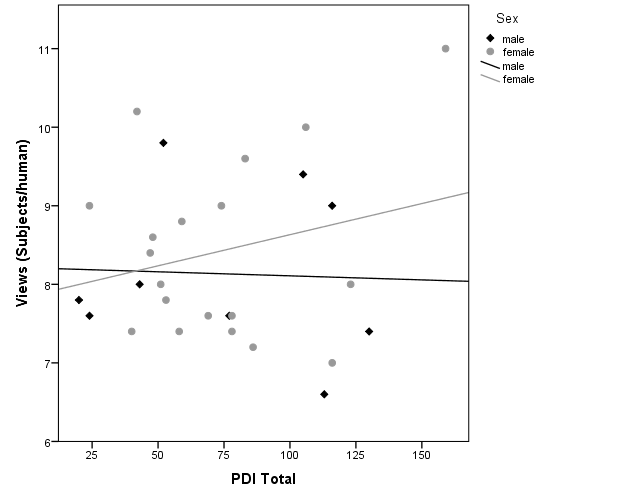
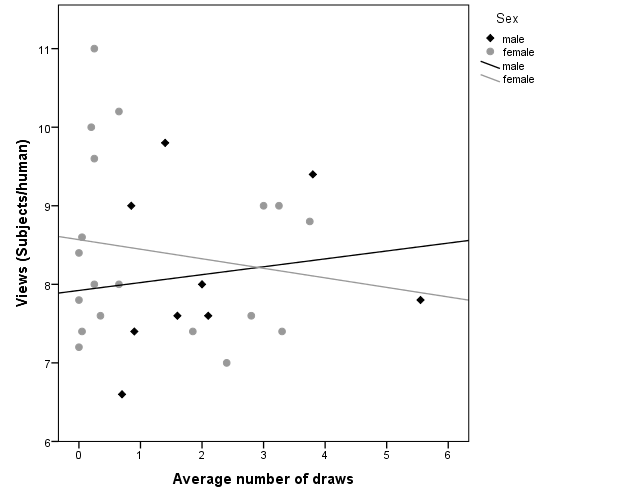
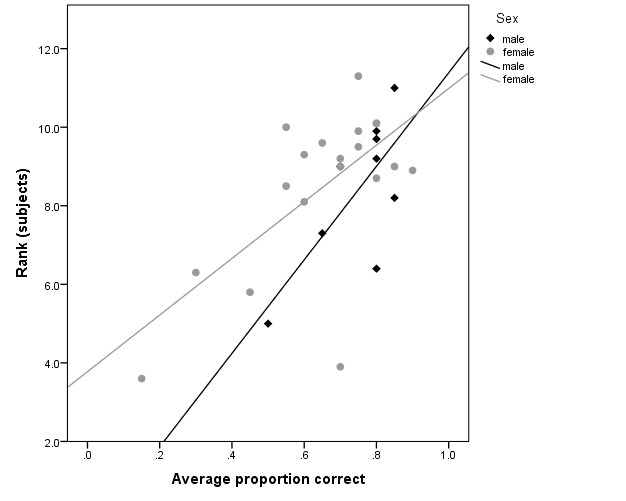
The hierarchical regression model for ranking with proportion correct on beads task, PDI, and sex as predictors, explained 50.2% of the variance in human ranking behaviour and significantly predicted the ranks of the faces chosen by the human participants (F(6,21)=3.53, p=.014). The individual predictors of PDI total (B=-.004 , t=-.353 , p=.728), sex (B=1.594 , t=1.174 , p=.254), age (B=.380, t=1.479 , p=.154), and ethnicity (B=-.240 , t=-.530 , p=.602) were not significant predictors of the subjects’ rank scores. However, average proportion of correct beads task answers were significant as a predictor (B= 7.781 , t=3.983 , p=.001), with every average correct answer, giving a 7.781 increase in rank of face chosen (see Figure 3A). In order to examine the cause of this interaction in further detail, an interaction of PDI and sex was added, but this was not significant (see Figure 3B). Thus, there was no effect of sex on ranking behaviour. In addition, an interaction of average proportion correct on the beads task and sex were added, however this interaction was also not significant. Therefore, both the interactions of PDI and sex, as well as average proportion correct and sex did not improve the base model. Consequently, the base model explaining the significant effect of average proportion correct on the beads task on ranking behaviour is the best model. Examining individual correlations for ranking behaviour, there is a significant correlation between human rank and average proportion correct on the beads task (r=.629 , p<.001). The next closest correlation to significance was between Face sex and average proportion correct (r=.275 , p=.156), and all other correlations for ranking were of a higher significance.

Figure 3: Scatterplots of regression models and interactions.   
A: is the significant base model representing the regressiomn between Average Proportion Correct on the beads task with Subjects Average ranked face choice.

B: Depicts the addition of the interaction between PDI and Sex to the base model A, which was not significant.

C: The not significant base model correlating the Average number of views/faces participants viewed before choosing a face, with their average number of draws on the beads task.

D : Displays the addition of the interaction of PDI and Sex to the base model C. This interaction did not explain more of the variance of the model, and was therefore rejected.



**A**

**B**

**C**

**D**

Discussion

The present study aimed to examine how the optimal choosing paradigm can be applied to the sampling of faces and whether JTC behaviour is linked to decision making when sampling and choosing faces. Consistent with the findings from Furl, Averbeck and McKay (In Prep) the study found that subjects viewed more faces before choosing a face, and choose lower ranked faces in comparison to the model. This corroborates the finding that humans oversample when sampling faces, causing them to miss out on their best choice option. The average number of faces viewed before choosing was not related with the number of beads drawn before choosing, suggesting that viewing behaviour is unrelated to JTC. However, average ranked face was associated with the percentage correct answers on the beads task, with those who got a higher percentage correct also choosing higher ranked faces on the Fiancé task. All other associations measured within the hierarchical regression models were not significant. Contradictory to previous research, there was no association between the PDI and the beads task, suggesting that delusional thoughts are not related to the sampling of beads or to JTC. There was also no relation between PDI and the sampling of faces in the Fiancé task. As the relationship between optimal choosing and stopping with facial sampling is a novel area, it is important to examine many of the potential factors which contribute to this oversampling effect, as well as understand why the JTC effect was not found in facial sampling.

The main significant link found between the Fiancé task and the beads task was that ranking was linked with performance on beads task. Those who chose higher ranked faces during the sequence task also got more answers correct on the beads task. This relationship does not show JTC behaviour because these participants who performed well on both tasks would have shown an appropriate level of sampling behaviour, similar to the sampling of the model, and therefore did not JTC. However, since we did not find an association between the number of face views and number of bead draws, we can infer that the oversampling effect prevalent in the Fiance task is not affected by JTC behaviour. This lack of association between the types of sampling could link back to Furl, Gallagher and Averbeck’s (2012) research, where participants simply preferred to look at faces than non-living objects. Especially in this study, the beads task was not image based in the same way that the Fiance task was, and the words Green and Blue appeared on the screen, rather than physically being able to see the beads, or even an image of the jars, the participants would have also had to learn to visualise the task in order to understand how to perform the task. Consequently, participants may have found the task more difficult to visualise and understand in comparison to the Fiance task because on computerised multimedia based tasks, it is easier to learn via pictures and words, rather than solely words (Mayer & Moreno, 2002). Moreover, despite the omissions of both the first 10 and last 10 trials out of the 40 trials in order to combat miscomprehension and fatigue effects, there still may be some miscomprehension and fatigue effects present in the 20 middle trials that were recorded. For example, a participant may still have found the task difficult and made fewer draws out of confusion (Balzan, Delfabbro & Galletly, 2012); or the fatigue effects from performing 40 trials may have induced participants to make fewer draws because they want to complete the study quicker, even after the initial 10 trials. Therefore, the lack of relationship between face views and bead draws may be due to the disparities in how interesting the participants found the stimuli.

The study could be criticised for the lack of diversity in ethnicity or age of the faces in the Fiancé task, with all images showing young Caucasian faces. Some participants mentioned not finding the faces displayed as physically attractive because they have a preference for another ethnicity, so gave faces lower rating. Since people tend to show an own race bias when looking at faces (Bar-Haim, Ziv, Lamy, & Hodes, 2006), participants may naturally give lower ratings for all faces. However this criticism does not apply in the present study because the Fiancé task is looking at the comparative rankings between the faces, not the attractiveness of the faces, with one face rank always being higher than another, regardless of its individual rating. To examine whether race may play a part in individual ratings of facial attractiveness, it may be interesting to create Fiancé tasks with faces of different ethnicities, and compare ratings between tasks.

Another contributing factor that could show variations in the Fiancé task is sexual orientation. There was no effect of the participants’ sex upon any of the data, with neither males nor females performing better or worse on a task. However, one female participant chose to sample female faces in the Fiancé task rather than sample the opposite sex like the remaining participants. Although one participant was not enough to sway results, it may also be interesting to examine whether there is a difference in sampling behaviour when sampling faces of the same versus opposite sex.

The lack of association between JTC and PDI scores in the present study is interesting because it conflicts with the results from previous research which has shown a strong link between high delusion scores on the PDI and JTC behaviour (Ross, McKay, Coltheart & Langdon, 2015; Van der Leer, Hatig, Goldmanis & McKay, 2015). This could have been a result of a number of factors, including sample size. The present study had a sample size of 28 participants whereas results showing a significant association between PDI scores and JTC have a sample size of upwards of 100 participants (Freeman, Pugh & Garety, 2008; Ross, Mckay, Coltheart & Langdon, 2015; Van der Leer, Hatig, Goldmanis & McKay, 2015). McKay, Langdon and Coltheart (2006) studied 58 participants and also did not find a significant association between JTC and the PDI. However, when they looked at each three PDI dimensions separately, they found that high conviction scores were associated with fewer draws. Delusional conviction is the extent to which individuals believe their delusional and paranoid thoughts, and many studies find this specific link between high delusional conviction scores on the PDI and fewer bead draws (Freeman, Pugh, Garety, 2008). Moreover, individuals with high delusional conviction show reasoning biases, including JTC behaviour, belief inflexibility and extreme responses, notably with belief inflexibility mediating the relationship between delusional conviction and JTC (Garety et al, 2005). Therefore, the lack of association between PDI and JTC in the present study may be a result of sample size or may be because JTC has a stronger and more specific link to delusional convictions rather than delusions in general.

Alternatively, the lack of association between delusional traits and JTC may be a result of not examining a clinical sample, therefore any delusional traits that could be linked to bead draws may not be severe enough to show JTC. Especially since JTC is only prevalent in 20% of the general population, there may have just not been enough participants that do show JTC in both the beads and Fiancé task. Therefore, perhaps using more clinical samples, for example schizophrenics showing high delusional traits who draw fewer beads, or perhaps even Autistic probands with impaired Theory of Mind (Brosnan, Chapman & Ashwin, 2014), would show a stronger link between delusions and JTC, as well as JTC in both the beads and Fiancé tasks.

There are also issues with motivation and incentive with the beads task. Van der Leer and McKay (2014), highlighted multiple confounds prevalent in the beads task, the main one relevant to the present study is motivation. In order for participants to have motivation to perform well on the task, the task should be incentivised, because there must be a real cost to collecting more information, rather than just the amount of extra time taken to sample. Without an incentive there is no objective reason to collect the data. In order to combat this effect, Van der Leer and McKay (2014) remodelled the beads task into a task involving distributing money across lakes containing different coloured fish. In the control condition, where there was no incentive, high delusion prone participants gave higher estimates than the low delusion group, with the normal JTC conclusion being reached. However, in the incentivised condition high delusion-prone participants gave lower estimates than the low delusion group, who were more likely to perform accurately, and similar to the Bayesian model. The researchers suggest that this difference may be that, unlike the high delusion group, the low delusion group may be more sensitive to reward therefore will provide more accurate results as they have an incentive to perform well. In comparison to the Fiancé task where participants had the incentive of trying to choose a highly ranked “date”, our beads task did not have that incentive. Moreover, the incentive is an extremely important variable in optimal choosing, in both optimal choosing tasks and in real world examples. For example, the reward of choosing the best employee to benefit the company should act as an incentive to combat the costs of sampling. In this version of the beads task, the only incentive was getting the correct answer which may not have been a big enough incentive for the participant, therefore if both tasks were incentivised, there may have been a more significant association between viewing behaviour and bead draws. The incentive may also contribute to the consistency of the optimal choosing behaviour effect in research. Since the addition of the incentive altered the over and undersampling effects in high and low delusion participants (Van der Leer & McKay, 2014), this may also explain why the Fiancé task has been found to be so consistent across its preliminary measures. In the Fiancé task, the same oversampling effect of faces was found even in sample sizes as low as 16 (Furl, Averbeck & McKay, 2014), which may be influenced by participants being routinely influenced and interested by the sampling of faces for an incentive with more emotional valence, rather than the beads task which has less emotional valence.

To conclude, when applying the optimal choosing paradigm to the sampling of faces in this novel area, participants oversample compared to undersampling when sampling in other areas. However this oversampling tendency is unrelated to JTC in the beads task. Further research to examine the reasoning behind facial oversampling behaviour should look into the importance of incentivised measures and perhaps also examine different sample type, for example clinical groups such as Schizophrenics and individuals with Autism, or compare the sampling behaviour of same versus opposite sex faces.

**References**

Balzan, R., Delfabbro, P., & Galletly, C. (2012). Delusion‐proneness or miscomprehension? A re‐examination of the jumping‐to‐conclusions bias. *Australian journal of psychology*, *64*(2), 100-107.

Balzan, R., Delfabbro, P., & Galletly, C. (2012). Delusion-proneness or miscomprehension? A re-examination of the jumping-to-conclusions bias. *Australian Journal of Psychology*, *64*, 100–107.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in own-race face processing. *Psychological science*, *17*(2), 159-163.

Brosnan, M., Chapman, E., & Ashwin, C. (2014). Adolescents with autism spectrum disorder show a circumspect reasoning bias rather than ‘jumping-to-conclusions’. *Journal of autism and developmental disorders*, *44*(3), 513-520.

Costa, V. D., & Averbeck, B. B. (2015). Frontal–Parietal and Limbic-Striatal Activity Underlies Information Sampling in the Best Choice Problem. *Cerebral Cortex (New York, NY)*, *25*(4), 972–982.

Evans, S. L., Averbeck, B. B., & Furl, N. (2015). Jumping to conclusions in schizophrenia. *Neuropsychiatric Disease and Treatment*, *11*, 1615–1624.

Freeman, D., Pugh, K., & Garety, P. (2008). Jumping to conclusions and paranoid ideation in the general population. *Schizophrenia research*, *102*(1), 254-260.

Furl, N., & Averbeck, B. B. (2011). Parietal cortex and insula relate to evidence seeking relevant to reward-related decisions. *Journal of Neuroscience*, *31*(48), 17572-17582.

Furl, N., Gallagher, S., & Averbeck, B. B. (2012). A selective emotional decision-making bias elicited by facial expressions. *PLoS One*, *7*(3), e33461.

Garety, P. A., Freeman, D., Jolley, S., Dunn, G., Bebbington, P. E., Fowler, D. G., ... & Dudley, R. (2005). Reasoning, emotions, and delusional conviction in psychosis. *Journal of abnormal psychology*, *114*(3), 373.

Garety, P. A., Freeman, D., Jolley, S., Dunn, G., Bebbington, P. E., Fowler, D. G., ... & Dudley, R. (2005). Reasoning, emotions, and delusional conviction in psychosis. *Journal of abnormal psychology*, *114*(3), 373.

Langdon, R., Ward, P. B., & Coltheart, M. (2010). Reasoning anomalies associated with delusions in schizophrenia. *Schizophrenia bulletin*, *36*(2), 321-330.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and instruction*, *12*(1), 107-119.

McKay, R., Langdon, R., & Coltheart, M. (2006). Need for closure, jumping to conclusions, and decisiveness in delusion-prone individuals. *The Journal of nervous and mental disease*, *194*(6), 422-426.

Mehl, S., Rief, W., Lüllmann, E., Ziegler, M., Kesting, M. L., & Lincoln, T. M. (2010). Are theory of mind deficits in understanding intentions of others associated with persecutory delusions?. *The Journal of nervous and mental disease*, *198*(7), 516-519.

Moritz, S., Woodward, T. S., & Lambert, M. (2007). Under what circumstances do patients with schizophrenia jump to conclusions? A liberal acceptance account. *British Journal of Clinical Psychology*, *46*, 127–137.

Peters, E. R., Joseph, S. A., & Garety, P. A. (1999). Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al. Delusions Inventory). *Schizophrenia bulletin*, *25*(3), 553-576.

Peters, E., Joseph, S., Day, S., & Garety, P. (2004). Measuring delusional ideation: the 21-item Peters et al. Delusions Inventory (PDI). *Schizophrenia bulletin*, *30*(4), 1005.

Ross, R. M., McKay, R., Coltheart, M., & Langdon, R. (2015). Jumping to conclusions about the beads task? A meta-analysis of delusional ideation and data-gathering. *Schizophrenia bulletin*, sbu187.

Van der Leer, L., Hartig, B., Goldmanis, M., & McKay, R. (2015). Delusion proneness and ‘jumping to conclusions’: relative and absolute effects. *Psychological medicine*, *45*(06), 1253-1262.

van Moerbeke, P. (1974). An optimal stopping problem with linear reward. *Acta Mathematica*, *132*(1), 111-151.