

HHS Public Access

Author manuscript Schizophr Res. Author manuscript; available in PMC 2023 July 01.

Published in final edited form as:

Schizophr Res. 2022 July ; 245: 50-58. doi:10.1016/j.schres.2022.01.054.

Computational Mechanisms Underlying Illusion of Control in **Delusional Individuals**

Soojung Na^{1,2,*}, Sylvia Blackmore^{4,*}, Dongil Chung⁵, Madeline O'Brien^{1,2,3}, Sarah Banker^{1,2,3}, Matthew Heflin^{2,3}. Vincenzo G. Fiore^{2,3}, Xiaosi Gu^{1,2,3}

¹Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai

² Department of Psychiatry, Icahn School of Medicine at Mount Sinai

³.Center for Computational Psychiatry, Icahn School of Medicine at Mount Sinai

⁴ Queen Square Institute of Neurology, University College London

⁵ Department of Biomedical Engineering, UNIST

Abstract

Humans navigate complex situations that require the accurate estimation of the controllability of the environment. Aberrant controllability computation might lead to maladaptive behaviors and poor mental health outcomes. Illusion of control, which refers to a heightened sense of control while the environment is uncontrollable, is one such manifestation and has been conceptually associated with delusional ideation. Nevertheless, this association has not yet been formally characterized in a computational framework. To address this, we used a computational psychiatry approach to quantify illusion of control in human participants with high (n=125) or low (n=126)trait delusion. Participants played a two-party exchange game in which their choices either did ("Controllable condition") or did not ("Uncontrollable condition") influence the future monetary

Declaration of Competing Interest

Corresponding Author: Xiaosi Gu, PhD, Icahn School of Medicine at Mount Sinai, 1 Gustave Levy Place, New York, NY, Tel: +1) 212-585-4670, xiaosi.gu@mssm.edu. These authors contributed equally.

Contributors

Conceptualization: X.G.; Methodology: S.N. and X.G.; Investigation: S.E.B., S.N., D.C., M.O., S.B., M.H.; Formal Analysis: S.N. and S.E.B.; Writing - Original Draft: S.N. and S.E.B.; Writing - Review and Editing: All authors; Funding Acquisition: X.G.; Supervision: X.G.

The authors declare no competing interests.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

offers made by simulated partners. We found that the two groups behaved similarly in modelagnostic measures (i.e. offer size, rejection rate). However, computational modeling revealed that compared to the low trait delusion group, the high delusion group overestimated their influence ("expected influence" parameter) over the offers made by their partners under the Uncontrollable condition. Highly delusional individuals also reported a stronger sense of control than those with low trait delusion in the Uncontrollable condition. Furthermore, the expected influence parameter and self-reported beliefs about controllability were significantly correlated in the Controllable condition in individuals with low trait delusion, whereas this relationship was diminished in those with high trait delusion. Collectively, these findings demonstrate delusional ideation is associated with aberrant computation of and belief about environmental controllability, as well as a belief-behavior disconnect.

1. Introduction

Humans live in complex environments, in which controllability estimation plays a crucial role in behavior optimization and mental well-being (Huys and Dayan, 2009; Lachman and Weaver, 1998; Maier and Seligman, 1976; Maier and Watkins, 2005; Overmier, 1968). As our environments are often social, aberrant estimation of social controllability, i.e., one's ability to influence somebody else's behavior, may also have a negative impact on one's community standing and social relationships. Illusion of control is an example of altered controllability estimation that is widespread in the general population and is characterized by the incorrect perception of a strong sense of control when things are not controllable (Johnson and Fowler, 2011). However, illusion of control has also been considered a form of "positive illusion" with adaptive functions (Taylor and Brown, 1994, 1988), as it distorts information in a positive direction and may help promote optimism, improve psychological well-being, and allow people to remain confident, motivated, and hopeful. However, if such belief persists in the face of evidence to the contrary, illusion of control can lead to maladaptive behaviors and even psychopathological symptoms (Balzan et al., 2013; Makridakis and Moleskis, 2015). For example, illusion of control over stock markets or slot machines might encourage risky investments and compulsive gambling, resulting in financial loss and hardships.

Illusion of control has been found to be heightened in individuals with schizophrenia (Balzan et al., 2013; Moritz et al., 2014), and has been associated with the formation and maintenance of delusions (Balzan et al., 2013). Delusions are a hallmark characteristic of psychosis (American Psychiatric Association, 2013) and often center around themes of control (Appelbaum et al., 1999; Sadock and Sadock, 2003; Moritz et al., 2014). In these cases, delusions can range from an illusory sense of hypercontrol (as exhibited in delusions of grandeur) to extreme helplessness (Sadock and Sadock, 2007; Moritz et al., 2014). Although a link has been found between delusions and illusion of control, delusions have not been richly studied in direct relation to social controllability. Nevertheless, common types of delusions such as paranoia and suspicion are manifestations of inherently distorted social models, prompting one's fixed belief that they might be harmed or otherwise manipulated by another person, group, or organization (American Psychiatric Association, 2013; Griffin and Fletcher, 2017).

Furthermore, considering the dimensional approach to psychosis, faulty estimations of social controllability and their detrimental consequences might be shared along a continuum that ranges from a personality trait in community samples to a psychiatric symptom in patients with schizophrenia (Van Os et al., 2009).

Previous work has attempted to understand what prompts delusional ideation by studying the cognitive distortions both in patients diagnosed with schizophrenia (Baker et al., 2019; Balzan et al., 2013; Eisenacher and Zink, 2017; So and Kwok, 2015; Ward and Garety, 2019) and in the general population (Balzan et al., 2013; Eisenacher and Zink, 2017; Larøi and Van der Linden, 2005; McLean et al., 2020; So and Kwok, 2015; Stainsby and Lovell, 2014; Ward and Garety, 2019; Woodward et al., 2007; Zawadzki et al., 2012). One prevalent finding suggests that delusional individuals prioritize evidence that matches their beliefs (Balzan et al., 2013) and tend to discount evidence that contradicts their beliefs (Woodward et al., 2007). An inclination to "jump to conclusions" has also been associated with delusion (Ward and Garety, 2019), although subsequent work has not consistently replicated this finding (McLean et al., 2020; So and Kwok, 2015) and this theoretical argument is challenged by findings that suggest a greater propensity for information-seeking in severe schizophrenia (Baker et al., 2019). A bias against disconfirmatory evidence has also been revealed to exist across the psychosis continuum (Eisenacher and Zink, 2017). Interestingly, metacognitive beliefs of need to control thoughts have been associated with delusion-proneness in non-clinical populations (Larøi and Van der Linden, 2005; Stainsby and Lovell, 2014), suggesting that both cognitive and metacognitive deficits might contribute to illusions of control.

Based on this literature, we hypothesize that individuals with delusions could have both overestimation of their influence over social situations (cognitive) and unrealistically strong beliefs about their influence (metacognitive). First, it is possible that delusional individuals may not be able to correctly compute environmental statistics and thus fail to estimate the correct level of influence that they have. Second, it is also possible that delusional individuals may fail to infer this objective estimate at a subjective level. In other words, there may be no problem with inferring controllability – but a metacognitive failure. Indeed, the neural substrates of metacognitive beliefs have also been shown to dissociate from those subserving cognitive and executive functions (Fotopoulou et al., 2009; Marsh, 2017). Nonetheless, cognitive (controllability computation) and metacognitive (subjective belief) processes are not mutually exclusive and may jointly contribute to aberrant estimation of controllability. To test this hypothesis, we examined individuals with either low or high trait delusion from a community sample (measured by the Peters et al. Delusion Inventory, PDI (Peters et al., 2004)) using a two-party economic exchange game (Na et al., 2021) in which participants could ("Controllable condition") or could not ("Uncontrollable condition") influence their simulated partners' future monetary offers. A set of computational models were used to capture each individual's estimation of controllability (termed "expected influence"); we also examined the relationship between these model-derived parameters and self-reported perceived controllability.

2. Materials and methods

2.1. Participants

This study was part of a larger online project examining social cognition and mental health. Participants were recruited from Prolific (www.prolific.co) with the eligibility criteria of (1) age between 18 and 64, (2) currently living in the United States, and (3) >90% approval rating on the Prolific behavioral research platform. Online consent was obtained for the 1,499 participants who participated and completed the social controllability task (Na et al., 2021; see 2.2. Social controllability task) and demographic surveys. 14 participants were excluded due to duplicate data files for their responses, and 143 because they accepted or rejected all offers within at least one condition in the social controllability task. Among the remaining 1,342 participants, 897 (male/female/other = 436/358/3, age = 36.14 ± 13.29) completed the 21-item Peters Delusions Inventory (PDI). Individuals who scored above and below 1SD from the mean (47.99) were subsampled from these 897 participants (high PDI group n = 125, low PDI group n = 126; PDI score mean = 47.99, SD = 37.58, low PDI group cutoff score = 10.41, high PDI group cutoff score = 85.57; see Table 1, sections 2.2. Psychometric assessment and 2.5. Statistical analysis below for details). Participants received a scaled bonus selected from a random trial in the social controllability task in addition to the base compensation. The study was approved by the Institutional Review Board at the Icahn School of Medicine at Mount Sinai. All participants provided informed consent online prior to participation.

2.2. Social controllability task

To investigate how delusional ideation affects illusion of control, we used a social controllability task (Na et al., 2021) in which individuals could estimate, and exploit whenever possible, the controllability of simulated social interactions. The basic rule of the game is constructed based on the ultimatum game, in which a proposer suggests a two-way split of some amount of money, and a responder may accept or reject the offer. If the responder accepts the offer, both the proposer and the responder receive the proposed portion. If the responder rejects the offer, both receive nothing. It is known that in this fairness game, individuals reject offers and sacrifice monetary rewards to account for considerations such as reputation building and aversion to unfairness (Fehr, 2004; Fehr and Schmidt, 1999).

The present task adds a social controllability manipulation to the classic ultimatum game (Na et al., 2021). In this version of the task, participants played the role of responder and were paired with proposers from two teams, each consisting of 30 different simulated players. Participants played with each team member in succession (Figure 1a). The teams were assigned a name (a team from "Aldertown" and "Banyan Bay") and a color (represented in the screen background and on team members' shirts; Figure 1a) to convey each proposer's team membership. Importantly, participants could influence the offers made by one team ("Controllable condition") but not the other ("Uncontrollable condition"). Participants were not instructed on the existence of different conditions. In the Controllable condition, participants could either increase or decrease the value of the next offer by rejecting or accepting the current offer, respectively. The amount of the offer change

was determined in a probabilistic manner: $\frac{1}{3}$ chance of raising the offer by \$2 (after rejecting) or decreasing the offer \$2 (after accepting), $\frac{1}{3}$ chance of raising/decreasing by \$1 (after rejecting/accepting), and $\frac{1}{3}$ chance of no change (Figure 1b). In contrast, under the Uncontrollable condition, offer amounts were sampled from a predetermined distribution (mean = \$5.0, SD = \$2.3) and their order of appearance was randomized for each participant. In both conditions, the initial offer was \$5 and the offers were constrained to be an integer between \$1 and \$9 (inclusive). At the end of the task, participants were asked to rate (using a sliding bar from 0–100%) how much influence they perceived themselves as having over the proposers on each team (Figure 1a).

2.3. Computational modeling

To probe the mechanisms underlying exploitation of social controllability in association with delusional traits, we used a set of forward thinking (FT) models (Na et al., 2021), which formalize how people compute the downstream effects of their current choices over future outcomes and use these projected effects to guide choices. In these FT models, the value of an action (a_{j} , acceptance or rejection) is the sum of (i) the utility of the expected immediate reward (r_{j}) given the internal norm (f_{j} ; one's internal understanding of the value of the 'fair' offer, which is updated based on observed offers instead of a 50/50 split as shown in (Gu et al., 2015)), and (ii) the utility of the simulated future outcomes given the expected influence (δ) that the agent assumes.

The utility of the immediate reward was computed as the reward amount subtracted by the aversion to norm violation as follows, where a (constrained to $0 \ a \ 1$) represents the sensitivity to norm violation (Gu et al., 2015):

$$U(r_i, f_i) = \begin{cases} r_i - \alpha \max[f_i - r_i, 0] & \text{if } r_i > 0(\text{accepted}) \\ 0 & \text{if } r_i = 0(\text{rejected}) \end{cases}$$

We also assumed that the norm was not static, but instead learned by observing partners' offers (*S*) trial-by-trial using the Rescorla-Wagner learning model (Sutton and Barto, 2018) with learning rate (0 ϵ 1). The initial norm was assumed to vary among individuals (\$0 f_0 \$20) (Gu et al., 2015).

In the FT models, the utility of future outcomes was defined as the sum of the discounted utilities of the future rewards, which could be mentally simulated from the hypothetical future choices (\bar{a} ; as visualized in Figure 3a). We assumed these hypothetical future choices to be greedy (dependent on the immediate utility of the choices at the trial) and deterministic (only one choice was selected rather than both choices being probabilistically considered). Importantly, in the decision tree shown in Figure 3a, the hypothetical trial-by-trial offer change was contingent on one's hypothetical choices and determined by the level of controllability that an agent assumes ('expected influence (δ)', hereafter; $2 \quad \delta \quad 2$). The expected influence δ was applied symmetrically, such that the offer changed by δ if the prior choice was rejected and $-\delta$ if accepted. The hypothetical offers were constrained to positive values, given the assumption that participants were aware that the offers would always be greater than \$0. The future discounting factor (γ), which described how much weight one

puts on future values, was fixed at 0.8 following a previous study (Na et al., 2021). We tested 1- to 4-steps of forward thinking horizons (*k*). In summary, the action value was computed as follows:

$$v|_{a_i} = U(r_i, f_i) + \sum_{j=1}^k \gamma^j \times U\left(\widehat{E}\left(r_{i+j} \middle| a_i, \overline{a}_{i+1}, \dots \overline{a}_{i+j}\right), f_i\right)$$

In addition, we also considered a valuation model without forward thinking ("0-step model"). This model had the same settings as the FT models (i.e., Rescorla-Wagner norm updates and aversion to norm violation) except that it did not have the future simulation components. Lastly, we also considered the model-free valuation in which the reward prediction error (θ_i) is updated with a learning rate of $\zeta(0 \ \zeta \ 1)$, assuming the deterministic greedy choice in the following trial:

$$Q_{i+1}(s_i, a_i) = Q_i(s_i, a_i) + \zeta \theta_i$$

$$\theta_i = U(r_i, f_i) + \gamma max(Q(s_{i+1}, 1), Q(s_{i+1}, 0)) - Q_i(s_i, a_i)$$

For each model, the difference in utility between accepting and rejecting was entered into a softmax function to compute the trial-by-trial probability of each choice. We fit the models at the individual level for the middle 24 trials only, to focus on the exploitation of controllability. The first 3 trials were excluded to account for the possibility that one might still be learning the game and understanding controllability, and the last 3 trials were excluded based on the assumption that strategic behavior would differ towards the end of the trial series.

2.4. Psychometric assessment

To measure delusional ideation, we used the 21-item Peters et al. Delusions Inventory (PDI) (Peters et al., 2004). The questionnaire consists of 21 yes/no questions (e.g., "Do you ever feel as if people are reading your mind?"). Each affirmative answer generates follow-up questions that ask the participant to rate feelings of distress, preoccupation, and conviction on a five-point Likert scale. Thus, five separate scores are obtained from each participant: y/n score (0–21), distress score (0–105), preoccupation score (0–105), conviction score (0–105), and total score (0–336; the aggregate of the other four scores). We also measured cognitive ability using the 16-item International Cognitive Ability Resource (ICAR scores) as a covariate in our analyses (Condon and Revelle, 2014).

2.5. Statistical analysis

Statistical analyses were carried out in MATLAB. We examined group differences in demographics, PDI scores, and model-agnostic task behaviors using a two-sample t-test (or a chi-square test for the sex variable) using the 'ttest2' function (or the 'crosstab' function for the chi-square test) in MATLAB. To examine the effect of delusional ideation on expected influence, we ran a repeated measures ANCOVA with the independent variables

of sex, age, ICAR scores, and group (high/low PDI) predicting the expected influence of each condition, using the 'fitrm' and 'ranova' functions in MATLAB. To compare the correlation coefficients, a Fisher r-to-z transformation and a z-test were conducted in http:// vassarstats.net/rdiff.html. To run a robust regression, we used the 'robustfit' function in MATLAB. The statistical threshold was P < .05.

3. Results

3.1. Demographic and clinical profiles

Confirming our selection of the two groups as shown in Figure 2a, the high PDI groups scored higher not only in total PDI scores, but also in all sub-factors (Table 1; total PDI score: t(249) = 40.92, $p \ll .00001$; PDI Y/N: t(249) = 42.91, $p \ll .00001$; distress: t(249) = 33.86, $p \ll .00001$; preoccupation: t(249) = 34.79, $p \ll .00001$; conviction: t(249) = 38.86, $p \ll .00001$). The two groups did not differ in sex ($\chi 2(2, N = 251) = 5.16$, p = .08); however, the high PDI group was significantly younger (t(249) = -2.78, p < .01) and had lower ICAR scores (t(249) = -2.45, p < .05) than the low PDI group. These variables were entered into subsequent ANCOVAs (Table 2, 3) to account for their potential effects on group differences.

3.2. Model-agnostic task behavior

Using the social controllability task, we first examined whether delusions were related to one's ability to distinguish between controllable and uncontrollable environments and consequently make strategic choices. Two model-agonistic behaviors were of interest here: 1) offer size and 2) rejection rates in both conditions. First, the high PDI group and the low PDI group both successfully raised the offers over time under the Controllable condition (Figure 2b), compared to the Uncontrollable condition in which offers were random (Figure 2c). The mean offer over all trials was also comparable between the two groups for the Controllable condition (mean_{high} = 5.71, mean_{low} = 6.02, t(249) = -1.26, p = .21).

Second, the two groups did not show significant differences in their rejection patterns in either the Controllable (low offers (\$1–3): mean_{high} = 63.1%, mean_{low} = 67.4%, $t(88) = -0.62 \ p = 0.54$; middle offers (\$4–6): mean_{high} = 64.1%, mean_{low} = 67.7%, $t(249) = -0.91 \ p = 0.37$; high offers (\$7–9): mean_{high} = 42.4%, mean_{low} = 44.3%, $t(172) = -0.51 \ p = 0.61$; Figure 2d) or the Uncontrollable condition (low offers (\$1–3): mean_{high} = 85.5%, mean_{low} = 82.4%, $t(249) = 0.87 \ p = 0.39$; middle offers (\$4–6): mean_{high} = 59.4%, mean_{low} = 55.7%, $t(249) = 0.80 \ p = 0.42$; high offer (\$7–9): mean_{high} = 13.8%, mean_{low} = 14.8%, $t(249) = -0.35 \ p = 0.72$; Figure 2e). These results suggest that both high and low PDI groups were able to distinguish between the controllable and uncontrollable environments.

3.3. Computational modeling of social controllability

3.3.1. Model comparison results—To further investigate the mechanisms underlying social controllability in relation to delusional traits, we used a set of FT models (Na et al., 2021) that formalize the exploitation of social controllability. These models allow us to measure the level of controllability mentally simulated in the forward-thinking process using the "expected influence" parameter (see 2.3. Computational modeling). Our previous

study using these models revealed that individuals simulate their influence over future offers regardless of the actual controllability of a social environment, yet the level of simulated controllability depends on actual controllability (Na et al., 2021). Here, we examined whether or not these results apply to our current sample and, if so, how the level of expected influence would be related to delusional traits.

To determine the best fitting model for the two PDI groups combined, we used Draper's Information Criteria (DIC) scores. Consistent with our previous study (Na et al., 2021), the 2-step FT model was the elbow point in the DIC graph for both the Controllable (Figure 3b) and the Uncontrollable conditions (Figure 3c), suggesting that the FT models best explain participants' choice patterns and that 3- or 4-step FT models do not provide more explanatory power than the 2-step FT model. These results were consistent with our previous finding that participants simulate exploitation of controllability, regardless of whether the environment is controllable. Accordingly, we extracted the parameters estimated using the 2-step FT model for further analyses.

3.3.2. Effect of trait delusion on expected influence—The expected influence parameter from the 2-step FT model allows us to measure (in dollars) how much influence people thought they had over the offers. Particularly for the Uncontrollable condition, in which the true influence is \$0, all expected influence estimates larger than \$0 indicate illusion of control because the actual offers are independent of participants' actions. We hypothesized that those with greater delusional traits would demonstrate higher illusion of control in both controllability conditions as captured by this expected influence. To test this hypothesis, we conducted a repeated measures ANCOVA to examine whether the expected influence differed by group, while controlling for sex, age, and cognitive ability. This analysis showed a significant interaction effect between condition and group (F(1, 245) =4.55, p < .05; Table 2). Indeed, a t-test confirmed that the expected influence estimates were higher for the high PDI group than the low PDI group under the Uncontrollable condition $(\text{mean}_{\text{high}} = 1.08, \text{mean}_{\text{low}} = 0.76 \ t(249) = 2.32 \ p < .05;$ Figure 4a), whereas they were comparable between the two groups under the Controllable condition (mean_{high} = 1.14, $\text{mean}_{\text{low}} = 1.20 \ t(249) = -0.49 \ p = .62$; Figure 4a). That is, the high PDI group mentally simulated a higher level of influence over their partners compared to the low PDI group when they did not have control over future monetary offers.

3.3.3. Effect of trait delusion on perceived social controllability—Next, we examined whether trait delusion also had an impact on self-reported beliefs about controllability using a similar ANCOVA analysis. We expected that the amplified illusion of control in the high PDI group found in the expected influence estimates would be confirmed by the self-reported measure. Echoing our previous results on the expected influence parameter, we found an interaction effect of group and condition on self-reported perceived controllability (F(1, 245) = 5.55, p < .05; Table 3). Specifically, the high PDI group reported higher perceived controllability than the low PDI group under the Uncontrollable condition (mean_{high} = 29.37, mean_{low} = 19.53 t(249) = 2.91 p < .01; Figure 4b), but not under the Controllable condition (mean_{high} = 54.41, mean_{low} = 57.89 t(249) = -0.83 p = .41; Figure

4b). Together, these findings demonstrate that delusional ideation is associated with both expected influence and self-reported perceived controllability.

3.3.4. Disconnect between choice behavior and subjective belief about

social controllability—Thus far, we showed that high PDI individuals have a greater illusion of control than low PDI individuals in both the expected influence and the perceived social controllability, supporting our first hypothesis. Our second hypothesis was that high delusion individuals would also show a disconnect between their beliefs and behavior. To test this, we ran correlations between expected influence and self-reported controllability for each condition and each group. When future monetary offers were controllable, the high PDI group's self-reported controllability was not correlated with expected influence (r = .15, p =.09; Figure 4c). In contrast, the low PDI group's self-report was highly correlated with the expected influence (r = .47, p < .0001; Figure 4d). The correlation coefficients were also statistically different between the two groups (z = -2.73, p < .01). When monetary offers were uncontrollable, neither the high PDI group (r = .16, p = .07; Figure 4e) nor the low PDI group (r = .06, p = .51; Figure 4f) showed significant correlation between self-reported controllability and expected influence. The coefficients were also not statistically different between the two groups (z = 0.80, p = .42). Our comparison of correlation coefficients can be read as a comparison of the coefficients of determination; namely, the variance explained or R-squared. On this view, the metacognitive deficit means that objective estimates of control explained less variance of subjective estimates, relative to individuals with low delusional scores.

We further ran a robust regression (Table 4, 5) for each condition to deal with the outliers, in which we regressed the self-reported controllability on group (the high PDI group was coded as 1 and the low PDI group as 0), expected influence, and the interaction between the expected influence and the group along with the controlling factors of sex, age, and ICAR scores. Consistent with our findings from the correlation analysis, we found that for the Controllable condition, the regression coefficients for expected influence (b = 15.47, p < .0001) and for the interaction term between group and expected influence were significant (b = -9.87, p < .05). Such effects were not significant for the Uncontrollable condition (expected influence: b = 1.83, p = .38; interaction: b = 3.12, p = 0.34). These findings suggest that in individuals with high delusional traits, self-reported beliefs about controllability were not grounded in their own estimation of controllability. Additionally, even among individuals with low delusional traits, subjective beliefs reflected their own estimation only in the controllable environment.

4. Discussion

In this study, we investigated the computational mechanisms of illusion of control in a community sample of participants with high or low trait delusion. We found that, despite a similar ability to estimate controllability in controllable environments, individuals with high delusional traits estimated an unrealistically high level of control and reported a stronger sense of control in uncontrollable environments compared to those with low trait delusion. Furthermore, delusional individuals demonstrated a disconnect between their exerted and perceived controllability when their social interactions were actually controllable. These

findings suggest that aberrant computation and perception of environmental controllability both contribute to illusion of control in delusional individuals.

Our main finding was that individuals with high trait delusion exhibited illusion of control at both objective (i.e., behavior) and subjective (i.e., self-reports) levels. Specifically, we found that in our social exchange game, the high delusion group estimated greater levels of influence over monetary offers made by others and also subjectively reported a stronger sense of controllability compared to the low delusion group when partners' offers were uncontrollable. These results provide evidence for illusion of control in individuals with high delusional ideation. Although our participants were recruited from a community sample, the current results are consistent with previous findings on patients with schizophrenia that demonstrate high positive symptoms were associated with the excessive illusion of control (Moritz et al., 2014). Thus, our findings here extend the link between trait delusion and illusion of control beyond those with a confirmed psychotic disorder diagnosis to the general population.

A second major finding of the current study is the disconnect between the subjective belief about control and exerted control expressed through behavior in highly delusional individuals when the environment was indeed controllable. Given the seemingly intact choice behavior in the high delusion group under the Controllable condition, we speculate that such a disconnect indicates deficits in belief formation at the metacognitive level associated with delusional ideation. This suggests that metacognitive processes may be separate from lower-level decision processes and may not always be grounded in actual evidence. Indeed, deficits in metacognition have been described as a stable feature of schizophrenia (Moritz et al., 2016; Morrison and Wells, 2003; Vohs et al., 2014). Although metacognitive impairments have been reported to relate primarily to negative symptoms (Hamm et al., 2012; McLeod et al., 2014; Trauelsen et al., 2016), they are also shown to contribute to delusions in patients (Bruno et al., 2012). Furthermore, metacognitive intervention has been found to reduce delusion, at least in the short-term (Andreou et al., 2017; Hepworth et al., 2011). In the non-clinical populations, delusion proneness has been related to deficits in metacognitive processes such as the ability to monitor and control one's thoughts (Larøi and Van der Linden, 2005; Stainsby and Lovell, 2014). Our finding of a disconnect between the self-reported belief and the estimated controllability further supports the association of metacognitive impairments and delusions.

Delusions have been formulated as false inference in the predictive coding framework, where a failure of belief updating has been linked to aberrant precision-weighting of priors and prediction errors (Sterzer et al., 2018). For example, patients with schizophrenia show a failure of sensory attenuation, as evinced by their superior performance on force-matching tasks compared to controls (Shegrill et al., 2005). In addition, the failure to predict one's actions has been linked to delusions of passivity, where one feels as if one's actions are controlled by an alien force (Sterzer et al., 2018; Frith et al., 1989; Synofzik et al., 2013). While delusions of passivity are consistent with the helplessness often experienced in schizophrenia, these observations are not necessarily incompatible with illusion of control in the same disorder. For instance, while a delusion of being persecuted by a government agency may be conceived of as an example of perceived helplessness,

the perceived omniscience and capacity to "see through" such a conspiracy can be seen as illusion of control (Moritz et al., 2014). Furthermore, delusions of passivity are typically linked to deficits in precision-weighting in the sensorimotor system, though our study uses a social cognitive paradigm. Different types of delusions might therefore arise within different modalities but co-exist in the same individuals, although further studies are needed to tease apart these differences.

The current study has a few limitations. First, our participants were recruited from a community sample and did not include clinical patients. Thus, caution is needed to apply our findings to the entire spectrum of delusion. Second, the current study lacks a non-social condition and thus cannot directly test if illusions of social and non-social control are correlated, a question that needs to be explored in future studies. Finally, the question of how task-specific cognitive functioning remains intact under the Controllable condition - despite aberrant controllability beliefs in delusional individuals - is not investigated as we considered it to be beyond the scope of the present study. Answering this question would necessitate the design of a new task to tease apart metacognitive belief updating and controllability learning processes.

Despite these limitations, the current study demonstrates that illusion of control and aberrant belief formation are linked to delusional traits in a community sample. These findings provide a computational account for how cognitive and metacognitive deficits might both contribute to illusion of control, which could inform future studies to further investigate the neural deficits in patients with schizophrenia.

Acknowledgements

We thank the two anonymous reviewers whose comments and suggestions helped improve and clarify this manuscript. This study is supported by internal funding from the Icahn School of Medicine at Mount Sinai. DC is supported by the National Research Foundation of Korea [Grant number: NRF-2018R1D1A1B07043582]. VGF is funded by the Mental Illness Research, Education, and Clinical Center (MIRECC VISN 2) at the James J. Peter Veterans Affairs Medical Center, Bronx, NY.X.G. is supported by the National Institute of Mental Health [grant number: R21MH120789, R01MH122611, R01MH123069]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The data in this paper were used in a dissertation as partial fulfillment of the requirements for a PhD degree at the Graduate School of Biomedical Sciences at Mount Sinai.

Role of the funding source

The funder did not have any role in conducting the study.

References

- American Psychiatric Association, American Psychiatric Association, 2013. DSM 5 American Psychiatric Association 70.
- Andreou C, Wittekind CE, Fieker M, Heitz U, Veckenstedt R, Bohn F, Moritz S, 2017. Individualized metacognitive therapy for delusions: a randomized controlled rater-blind study. Journal of behavior therapy and experimental psychiatry 56, 144–151. [PubMed: 27919404]
- Appelbaum PS, Robbins PC and Roth LH, 1999. Dimensional approach to delusions: comparison across types and diagnoses. American Journal of Psychiatry, 156(12), 1938–1943. [PubMed: 10588408]
- Baker SC, Konova AB, Daw ND, Horga G, 2019. A distinct inferential mechanism for delusions in schizophrenia. Brain 142, 1797–1812. [PubMed: 30895299]

- Balzan R, Delfabbro P, Galletly C, Woodward T, 2013. Confirmation biases across the psychosis continuum: The contribution of hypersalient evidence-hypothesis matches. British Journal of Clinical Psychology 52, 53–69. [PubMed: 23398112]
- Balzan RP, Delfabbro PH, Galletly CA, Woodward TS, 2013. Illusory correlations and control across the psychosis continuum: the contribution of hypersalient evidence-hypothesis matches. The Journal of nervous and mental disease 201, 319–327. [PubMed: 23538977]
- Bruno N, Sachs N, Demily C, Franck N, Pacherie E, 2012. Delusions and metacognition in patients with schizophrenia. Cognitive neuropsychiatry 17, 1–18. [PubMed: 22216943]
- Condon DM, Revelle W, 2014. The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. Intelligence 43, 52–64.
- Eichner C and Berna F, 2016. Acceptance and efficacy of metacognitive training (MCT) on positive symptoms and delusions in patients with schizophrenia: a meta-analysis taking into account important moderators. Schizophrenia bulletin, 42(4), 952–962. [PubMed: 26748396]
- Eisenacher S, Zink M, 2017. Holding on to false beliefs: The bias against disconfirmatory evidence over the course of psychosis. Journal of behavior therapy and experimental psychiatry 56, 79–89. [PubMed: 27608522]
- Fehr E, 2004. Human behaviour: don't lose your reputation. Nature 432, 449. [PubMed: 15565133]
- Fehr E, Schmidt KM, 1999. A theory of fairness, competition, and cooperation. The quarterly journal of economics 114, 817–868.
- Fotopoulou A, Rudd A, Holmes P and Kopelman M, 2009. Self-observation reinstates motor awareness in anosognosia for hemiplegia. Neuropsychologia, 47(5), 1256–1260. [PubMed: 19428388]
- Frith CD and Done DJ, 1989. Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. Psychological medicine, 19(2), 359–363. [PubMed: 2762440]
- Garrison JR, Fernandez-Egea E, Zaman R, Agius M and Simons JS, 2017. Reality monitoring impairment in schizophrenia reflects specific prefrontal cortex dysfunction. NeuroImage: Clinical, 14, 260–268. [PubMed: 28203529]
- Griffin JD, Fletcher PC, 2017. Predictive processing, source monitoring, and psychosis. Annual review of clinical psychology 13, 265–289.
- Gu X, Wang X, Hula A, Wang S, Xu S, Lohrenz TM, Knight RT, Gao Z, Dayan P, Montague PR, 2015. Necessary, yet dissociable contributions of the insular and ventromedial prefrontal cortices to norm adaptation: computational and lesion evidence in humans. Journal of Neuroscience 35, 467–473. [PubMed: 25589742]
- Hamm JA, Renard SB, Fogley RL, Leonhardt BL, Dimaggio G, Buck KD, Lysaker PH, 2012. Metacognition and social cognition in schizophrenia: stability and relationship to concurrent and prospective symptom assessments. Journal of clinical psychology 68, 1303–1312. [PubMed: 22886716]
- Hepworth C, Startup H, Freeman D, 2011. Developing treatments of persistent persecutory delusions: the impact of an emotional processing and metacognitive awareness intervention. The Journal of nervous and mental disease 199, 653–658. [PubMed: 21878778]
- Huys QJ, & Dayan P (2009). A Bayesian formulation of behavioral control. Cognition, 113(3), 314– 328. [PubMed: 19285311]
- Johnson DD, Fowler JH, 2011. The evolution of overconfidence. Nature 477, 317–320. [PubMed: 21921915]
- Lachman ME (1998). Weaver SL (1998). The sense of control as a moderator of social class differences in health and well-being
- Larøi F, Van der Linden M, 2005. Metacognitions in proneness towards hallucinations and delusions. Behaviour research and Therapy 43, 1425–1441. [PubMed: 16159586]
- Maier SF and Seligman ME, 1976. Learned helplessness: theory and evidence. Journal of experimental psychology: general, 105(1), 3.
- Maier SF, & Watkins LR (2005). Stressor controllability and learned helplessness: the roles of the dorsal raphe nucleus, serotonin, and corticotropin-releasing factor. Neuroscience & Biobehavioral Reviews, 29(4–5), 829–841 [PubMed: 15893820]

- Makridakis S and Moleskis A, 2015. The costs and benefits of positive illusions. Frontiers in psychology, 6, p.859. [PubMed: 26175698]
- Marsh H, 2017. Admissions: a life in brain surgery Hachette UK.
- Mathys C, Daunizeau J, Friston KJ, Stephan KE 2011. A Bayesian foundation for individual learning under uncertainty. Frontiers in human neuroscience, 5, 39. [PubMed: 21629826]
- McLean BF, Mattiske JK, Balzan RP, 2020. Jumping to conclusions in the less-delusion-prone? Preliminary evidence from a more reliable beads task. Journal of Behavior Therapy and Experimental Psychiatry 68, 101562. [PubMed: 32105906]
- McLeod HJ, Gumley AI, MacBeth A, Schwannauer M, Lysaker PH, 2014. Metacognitive functioning predicts positive and negative symptoms over 12 months in first episode psychosis. Journal of psychiatric research 54, 109–115. [PubMed: 24725651]
- Montague PR, Lohrenz T, 2007. To detect and correct: norm violations and their enforcement. Neuron 56, 14–18. [PubMed: 17920011]
- Moritz S, Balzan RP, Bohn F, Veckenstedt R, Kolbeck K, Bierbrodt J, Dietrichkeit M, 2016. Subjective versus objective cognition: Evidence for poor metacognitive monitoring in schizophrenia. Schizophrenia research 178, 74–79. [PubMed: 27591821]
- Moritz S, Andreou C, Schneider BC, Wittekind CE, Menon M, Balzan RP and Woodward TS, 2014. Sowing the seeds of doubt: a narrative review on metacognitive training in schizophrenia. Clinical psychology review, 34(4), 358–366. [PubMed: 24866025]
- Moritz S, Thompson SC, Andreou C, 2014. Illusory control in schizophrenia. Journal of Experimental Psychopathology 5, 113–122.
- Moritz S and Woodward TS, 2007. Metacognitive training in schizophrenia: from basic research to knowledge translation and intervention. Current opinion in psychiatry, 20(6), 619–625. [PubMed: 17921766]
- Morrison AP, Wells A, 2003. A comparison of metacognitions in patients with hallucinations, delusions, panic disorder, and non-patient controls. Behaviour research and therapy 41, 251–256.
 [PubMed: 12547384]
- Na S, Chung D, Hula A, Perl O, Jung J, Heflin M, Blackmore S, Fiore VG, Dayan P and Gu X, 2021. Humans use forward thinking to exploit social controllability. Elife, 10, e64983. [PubMed: 34711304]
- Overmier JB (1968). Interference with avoidance behavior: Failure to avoid traumatic shock. Journal of Experimental Psychology, 78(2p1), 340. [PubMed: 5722450]
- Peters E, Joseph S, Day S, Garety P, 2004. Measuring delusional ideation: the 21-item Peters et al. Delusions Inventory (PDI). Schizophrenia bulletin 30, 1005–1022. [PubMed: 15954204]
- Sadock BJ, Sadock VA, 2003. Kaplan and Sadock's Synopsis of Psychiatry (9th ed.). Philadelphia, PA. Lippincott Williams & Wilkins.
- Sadock BJ, Sadock VA, 2007. Kaplan and Sadock's Synopsis of Psychiatry: Behavioral Sciences/ Clinical Psychiatry (10th edition). Philadelphia, PA. Lippincott Williams & Wilkins.
- Shergill SS, Samson G, Bays PM, Frith CD and Wolpert DM, 2005. Evidence for sensory prediction deficits in schizophrenia. American Journal of Psychiatry, 162(12), 2384–2386. [PubMed: 16330607]
- Simons JS, Garrison JR and Johnson MK, 2017. Brain mechanisms of reality monitoring. Trends in cognitive sciences, 21(6), 462–473. [PubMed: 28462815]
- So SH, Kwok NT, 2015. Jumping to conclusions style along the continuum of delusions: delusionprone individuals are not hastier in decision making than healthy individuals. PloS one 10, e0121347. [PubMed: 25793772]
- Stainsby LM, Lovell GP, 2014. Proneness to hallucinations and delusions in a non-clinical sample: Exploring associations with metacognition and negative affect. Australian Journal of Psychology 66, 1–7.
- Sterzer P, Adams RA, Fletcher P, Frith C, Lawrie SM, Muckli L, Petrovic P, Uhlhaas P, Voss M and Corlett PR, 2018. The predictive coding account of psychosis. Biological psychiatry, 84(9), 634–643. [PubMed: 30007575]
- Sutton RS, Barto AG, 2018. Reinforcement learning: An introduction MIT press.

- Synofzik M, Vosgerau G and Voss M, 2013. The experience of agency: an interplay between prediction and postdiction. Frontiers in psychology, 4, 127. [PubMed: 23508565]
- Taylor SE, Brown JD, 1994. Positive illusions and well-being revisited: separating fact from fiction
- Taylor SE, Brown JD, 1988. Illusion and well-being: a social psychological perspective on mental health. Psychological bulletin 103, 193. [PubMed: 3283814]
- Trauelsen AM, Gumley A, Jansen JE, Pedersen MB, Nielsen H-GL, Trier CH, Haahr UH, Simonsen E, 2016. Metacognition in first-episode psychosis and its association with positive and negative symptom profiles. Psychiatry research 238, 14–23. [PubMed: 27086205]
- Van Os J, Linscott RJ, Myin-Germeys I, Delespaul P, Krabbendam L, 2009. A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistenceimpairment model of psychotic disorder. Psychological medicine 39, 179. [PubMed: 18606047]
- Vohs JL, Lysaker P, Francis M, Hamm J, Buck K, Olesek K, Outcalt J, Dimaggio G, Leonhardt B, Liffick E, 2014. Metacognition, social cognition, and symptoms in patients with first episode and prolonged psychoses. Schizophrenia Research 153, 54–59. [PubMed: 24503175]
- Ward T, Garety PA, 2019. Fast and slow thinking in distressing delusions: A review of the literature and implications for targeted therapy. Schizophrenia research 203, 80–87. [PubMed: 28927863]
- Woodward TS, Buchy L, Moritz S, Liotti M, 2007. A bias against disconfirmatory evidence is associated with delusion proneness in a nonclinical sample. Schizophrenia bulletin 33, 1023–1028. [PubMed: 17347526]
- Zawadzki JA, Woodward TS, Sokolowski HM, Boon HS, Wong AHC, Menon M, 2012. Cognitive factors associated with subclinical delusional ideation in the general population. Psychiatry Research 197, 345–349. [PubMed: 22421072]

Na et al.



Figure 1. Task Paradigm.

(a) The figures show the screens that were presented to participants. Participants played the game with two different teams in a row. The order of the teams was counterbalanced. For each team assigned, participants viewed a set of avatars and the team name and then played 30 rounds of the game with different team members. In each round, participants viewed an avatar with a name. In the following screen, an offer amount was presented and then participants could choose to accept or reject the proposal. Then the outcome was presented. At the end of the task, participants were asked to rate how much influence they thought they had on the offers from each team. (b) In the Controllable condition, the next offer increased (decreased) by \$0, 1, or 2 with a uniform probability (1/3 each) if participants rejected (accepted).

Na et al.



Figure 2. Model-agnostic behavioral results.

(a) We subsampled the low PDI group and the high PDI group below and above 1 SD on the PDI score (PDI mean = 47.99, SD = 37.58; low PDI group n = 126 (cutoff score = 10.41), high PDI group n = 125 (cutoff score = 85.57)). (b) The offer trajectories under the Controllable condition were comparable between the low PDI group and the high PDI group (mean offer_{high} = 5.71, mean offer_{low} = 6.02, t(249) = -1.26, p = 0.21). (c) The offer trajectories under the Uncontrollable condition (mean offer = 5 for both groups). (d) For the Controllable condition, we did not find significant differences in the rejection patterns by offer size (rejection rates for the low offers: mean_{high} = 63.1%, mean_{low} = 67.4%, t(88) = -0.62, p = 0.54; medium offers: mean_{high} = 64.1%, mean_{low} = 67.7%, t(249) = 0.91, p = 0.37; high offers: mean_{high} = 42.4%, mean_{low} = 44.3%, t(172) = -0.51, p = 0.61). (e) The rejection rates were comparable for the Uncontrollable condition as well. (rejection rates for the low offers: mean_{high} = 59.4%, mean_{low} = 55.7%, t(249) = 0.80, p = 0.42; high offers: mean_{high} = 13.8%, mean_{low} = 14.8%, t(249) = -0.35, p = 0.72).



Figure 3. Computational modeling of choices.

(a) The forward thinking model assumes that an agent simulate their future actions and consequential future offers based on the partner's current offer (s_i) , "expected influence (δ) ", and value functions. The thick and solid line shows an example of a simulated path. (**b-c**) The mean DIC scores showed that the 2-step model was the elbow point for both the Controllable and the Uncontrollable conditions.

Na et al.



Figure 4. Model estimated and self-reported social controllability.

(a) The expected influence did not differ between the low and high PDI groups for the Controllable condition (mean_{highPDI} = 1.14, mean_{lowPDI} = 1.20, t(249) = -0.49, p = .62) whereas the expected influence was higher for the high PDI group than the low PDI group for the Uncontrollable condition (mean_{highPDI} = 1.08, mean_{lowPDI} = 0.76, t(249) = 2.32, p < .05 (b) Self-reported controllability was comparable between the two group for the Controllable condition (mean_{highPDI} = 54.41, mean_{lowPDI} = 57.89, t(249) = -0.83, p = .41) whereas, for the Uncontrollable condition, the high PDI group reported perceiving higher controllability than the low PDI group (mean_{highPDI} = 29.37, mean_{lowPDI} = 19.53, *t*(249) = 2.91, p < .01). (c-d) For the Controllable condition, the high PDI group did not have a significant correlation between the expected influence and the self-reported controllability (r = 0.15, p = 0.09) whereas the low PDI group showed a significant association between the expected influence and the self-reported controllability (r = 0.47, p << 0.0001). The correlation coefficients were significantly different between the two groups (z = -2.73, p < -2.73) .01). (e-f) For the Uncontrollable condition, the correlation between the expected influence and self-reported controllability was not significant for either the high PDI group (r = 0.16, p = 0.07) or the low PDI group (r = 0.06, p = 0.51). The correlation coefficients were not statistically different between the two groups (z = 0.80, p = 0.42).

Table 1.

Participant characteristics.

	High PDI group (n=125) Mean (SD)	Low PDI group (n=126) Mean (SD)	χ ^{2/t}	р
Sex (male/female/other)	50/74/1	67/69	5.16	0.08
Age	34.3 (12.8)	39 (13.9)	-2.78	<.01
ICAR (cognitive ability)	5.9 (3.0)	6.9 (3.7)	-2.45	<.05
Total PDI	118.3 (31.1)	3.9 (4.2)	40.92	<<.0001
PDI Y/N	10.9 (2.7)	0.5 (0.5)	42.91	<<.0001
PDI distress	33.4 (10.7)	1.0 (1.2)	33.86	<<.0001
PDI preoccupation	34.0 (10.7)	0.8 (1.0)	34.79	<<.0001
PDI conviction	40.0 (11.0)	1.6 (1.8)	38.86	<<.0001

Table 2.

Repeated measures ANCOVA: modelled controllability

	Sum of Squares	DF	Mean Square	F	Р
(Intercept)	1.91	1	1.91	1.71	0.19
Group (PDI low/high)	1.67	1	1.67	1.50	0.22
Sex	11.17	2	5.58	5.02	0.01
Age	0.58	1	0.58	0.53	0.47
ICAR (cognitive ability)	0.76	1	0.76	0.68	0.41
Error	272.61	245	1.11	1.00	0.50
(Intercept):condition	1.00	1	1.00	0.90	0.34
Group:condition	4.55	1	4.55	4.10	0.04
Sex:condition	11.49	2	5.75	5.17	0.01
Age:condition	0.02	1	0.02	0.02	0.89
IQ:condition	0.24	1	0.24	0.22	0.64
Error(condition)	272.30	245	1.11	1.00	0.50

Table 3.

Repeated measures ANCOVA: Self-reported controllability

	Sum of Squares	DF	Mean Square	F	Р
(Intercept)	25,759.12	1	25,759.12	29.09	0.00
Group (PDI low/high)	922.44	1	922.44	1.04	0.31
Sex	1,246.00	2	623.00	0.70	0.50
Age	4,033.15	1	4,033.15	4.55	0.03
ICAR (cognitive ability)	252.74	1	252.74	0.29	0.59
Error	216,935.38	245	885.45	1.00	0.50
(Intercept):condition	10,964.64	1	10,964.64	16.03	0.00
Group:condition	3,798.69	1	3,798.69	5.55	0.02
Sex:condition	1,109.47	2	554.74	0.81	0.45
Age:condition	487.86	1	487.86	0.71	0.40
IQ:condition	9,012.74	1	9,012.74	13.17	0.00
Error(condition)	167,623.29	245	684.18	1.00	0.50

Author Manuscript

Table 4.

Robust regression predicting the self-reported controllability for the Controllable condition

	b	SE	t	р
Intercept	49.92	8.80	5.67	0.0000
Sex	-7.93	3.90	-2.03	0.0433
Age	-0.53	0.15	-3.58	0.0004
ICAR	2.13	0.58	3.64	0.0003
Group	9.56	6.03	1.59	0.1141
$\delta_{\text{Controllable}}$	15.47	2.82	5.48	0.0000
$\delta_{Controllable} \!\!\times \!\! Group$	-9.87	3.88	-2.55	0.0115

Table 5.

Robust regression predicting the self-reported controllability for the Uncontrollable condition

	b	SE	t	р
Intercept	22.04	7.21	3.06	0.0025
Sex	1.16	3.44	0.34	0.7359
Age	0.13	0.13	0.99	0.3242
ICAR	-1.58	0.50	-3.15	0.0018
Group	3.98	4.70	0.85	0.3983
$\delta_{Uncontrollable}$	1.83	2.06	0.89	0.3757
$\delta_{\text{Uncontrollable}}\!\!\times\!\!\text{Group}$	3.12	3.26	0.95	0.3406

Author Manuscript