UiT The Arctic University of Norway

Department of Psychology

Metacognition and decision-making in schizophrenia:

Exploring how aberrant processing and representation of uncertainty may explain cognitive-behavioral biases

Isabel Viola Kreis

A dissertation for the degree of Philosophiae Doctor – December 2020

Table of Contents

List of Tables

List of Figures

Acknowledgments

First and foremost, I would like to thank the Research Council of Norway for providing the funding for this project.

Further, I would like to express my deepest appreciation to my main supervisor, Prof. Gerit Pfuhl, and my co-supervisor, Assoc. Prof. Matthias Mittner, for their guidance, valuable advice and unwavering support. I am extremely grateful for everything that I have learned from you.

I also want to express my deepest gratitude to Prof. Steffen Moritz and his wonderful team - the Clinical Neuropsychology Working Group of the Department of Psychiatry and Psychotherapy of the University Medical Center Hamburg-Eppendorf (UKE), Germany. Thank you for opening your doors to me and 'my' portable eye-tracker, for your indispensable guidance regarding the clinical assessments I conducted and for always having an open ear. I have learned so much from you.

I also wish to thank UiT – The Artic University of Norway, and in particular the Department of Psychology, for providing such a fruitful, stimulating and collegial working and learning environment. Thank you to all the fantastic colleagues for exciting and helpful discussions, for scientific and motivational support. Here, a special thanks is reserved for Nya, the best office mate one could wish for, and the EPIC research group of cognitive psychology, a bunch of wonderful, sharp and inspiring people.

Above all, I am greatly indebted to the participants of my studies. Thank you for the trust you have placed in me and for the time and effort you invested in order to move science forward. Without you, none of this would have been possible.

Last but not least, I would like to thank my family and friends for believing in me and my choices and for their unconditional support.

Abstract

The cognitive profile associated with schizophrenia has long been a research subject of interest. A plethora of studies have not only revealed cognitive deficits but also cognitive biases, meaning qualitative deviations in the way information is processed and evaluated. Such biases are for example reflected in findings of diminished metacognitive accuracy, or premature and disadvantageous decision-making. Interestingly, though often studied separately from one another, many of these biases may arise from similar underlying mechanistic aberrancies related to the processing and representation of uncertainty.

This thesis aimed to explore this overarching role of uncertainty with a particular focus on metacognitive processes and decision-making in schizophrenia. Cognitive-behavioral assessments were conducted using computerized tasks, complemented with pupillometric measures and cognitive modelling techniques. In *Paper I*, the relationship between decision-making under uncertainty and metacognitive accuracy was investigated. In *paper II*, decision-making under different kinds of uncertainty was examined, while pupil size was recorded as an indicator of norepinephrinergic activity, a postulated neurochemical marker of uncertainty. In *paper III*, self-reported and objectively assessed effort were inspected as knowledge- and regulation-based components of metacognition. Across all studies, the uncertainty-related measures of cognitive-behavioral performance were surprisingly similar between individuals with schizophrenia and the respective control group. However, subtle differences emerged within subgroups and on cognitive model-based estimates of uncertainty representation. Furthermore, pupillometric measures revealed significant differences in the way individuals with schizophrenia process relevant information, pointing towards diminished effort allocation and decreased tracking of uncertainty-dependent informational salience. The findings are discussed within a wider framework regarding the potentially central role of uncertainty for various clinical and cognitive-behavioral symptoms of schizophrenia.

List of papers

Paper I

Kreis, I., Biegler, R., Tjelmeland, H., Mittner, M., Reitan, S. K., & Pfuhl, G. (2019, July 30). Overestimation of volatility in schizophrenia and autism? A comparative study using a probabilistic reasoning task. *Accepted for publication in PLOS ONE. Preprint:* https://doi.org/10.31219/osf.io/ea3kz

Paper II

Kreis, I., Zhang, L., Moritz, S., & Pfuhl, G. (2020, October 19). Spared performance but increased uncertainty in schizophrenia: evidence from a probabilistic decision-making task. *Preprint. Submitted for publication.* https://doi.org/10.31219/osf.io/qaupb

Paper III

Kreis, I., Moritz, S., & Pfuhl, G. (2020). Objective versus subjective effort in schizophrenia. *Frontiers in Psychology*, 11, 1469. https://doi.org/10.3389/fpsyg.2020.01469

List of abbreviations

ASD	Autism Spectrum Disorders
DSM	Diagnostic and Statistical Manual of Mental Disorders
DTD	Draws-to-Decision
HC	Healthy (i.e. non-psychiatric and neurotypically developing) Controls
HMM	Hidden Markov Model
IGT	Iowa Gambling Task
JTC	Jumping-to-Conclusions
LC	Locus Coeruleus
LC-NE	Locus Coeruleus-Norepinephrine
NE	Norepinephrine
PANSS	Positive and Negative Symptoms Scale
SZ	Schizophrenia
UKE	University Medical Center Hamburg-Eppendorf
WAIS-IV	Wechsler Adult Intelligence Scale, 4th edition

1 General introduction

Schizophrenia is a severe mental health disorder with a comparatively low global prevalence of approximately 0.28% as of 2016 (Charlson et al., 2018). However, the burden associated with this disorder is immense, both for the affected individual and the society at large, given that schizophrenia is one of the 15 leading causes of disability worldwide (Vos et al., 2017). The psychopathology of schizophrenia is diverse and heterogeneous (Lindenmayer, Bernstein-Hyman, Grochowski, & Bark, 1995), and symptoms are typically characterized along a positive and a negative dimension. Positive symptoms describe the presence of phenomena that are absent in the healthy individual, such as hallucinations, delusions and disorganized speech. In contrast, negative symptoms describe the absence of functions that are present in the healthy individual, such as avolition and diminished emotional expression (American Psychiatric Association, 2013). On top of these rather specific and characteristic symptoms, patients with schizophrenia typically demonstrate cognitive deficits regarding attention, executive control, working memory, episodic memory and language (Barch, 2005; Fioravanti, Bianchi, & Cinti, 2012; Kuperberg & Heckers, 2000; Schaefer, Giangrande, Weinberger, & Dickinson, 2013). While such cognitive impairments are based on quantitative differences between patients and healthy control groups, with the former achieving less in cognitive tests than the latter, patients also demonstrate qualitative differences in how information is obtained, processed and evaluated. This is often referred to as *cognitive biases*, rather than poor cognitive performance (Moritz et al., 2010). Typical examples include aberrant (probabilistic) decision-making such as the so-called 'Jumping-to-Conclusions' (JTC) bias, where (too) little information is gathered before a decision is reached; a bias against disconfirmatory evidence, where information that conflicts with prior beliefs is ignored; and an overconfidence in errors (Moritz et al., 2010). The latter is commonly described as a sign of deficient metacognition, i.e. 'thinking about thinking' (see e.g., Moritz, Woodward, & Chen, 2006; Moritz & Woodward, 2006). Indeed, evidence from various studies points towards impaired metacognition in schizophrenia, including, for example, the lack of correspondence between self-evaluation of performance and clinician-rated performance (Moritz, Ferahli, & Naber, 2004). Deficits and aberrant information processing are further reflected in more general learning and decision-making paradigms where patients with schizophrenia tend to make more disadvantageous decisions, for example in gambling tasks (Lee et al., 2007; Martino, Bucay, Butman, & Allegri, 2007; Ritter, Meador-Woodruff, & Dalack, 2004; Shurman, Horan, & Nuechterlein, 2005; Struglia et al., 2011), or reinforcement learning tasks (Deserno, Heinz, & Schlagenhauf, 2017; Frank, 2008; Maia & Frank, 2017).

Interestingly, and as will be outlined in detail in the following sections, many of these cognitive biases and performance differences can be interpreted within frameworks concerning the processing and representation of uncertainty (Broyd, Balzan, Woodward, & Allen, 2017). Impaired metacognitive abilities are for example reflected in overconfidence, i.e. decreased uncertainty regarding one's own cognitive performance. Furthermore, metacognitive monitoring processes conceptually overlap with cognitive control, which in turn has been linked to uncertainty (Mushtaq, Bland, & Schaefer, 2011). Metacognition might also moderate effort invested in the task at hand (Efklides, 2009), causing decreased task performance (see

section 1.2). Moreover, decision-making paradigms employed in schizophrenia research often require patients to make decisions under uncertainty. Here, suboptimal decision-making might be driven by aberrant processing of, and altered sensitivity to, uncertainty, possibly linked to metacognitive abilities (see section 1.3).

This overarching role of processing, representing and dealing with uncertainty, and its potential to explain various cognitive-behavioral biases is also reflected in recent hypotheses that conceptualize the phenomenology of schizophrenia in the context of a Bayesian brain account (see section 1.1). Here, aberrant uncertainty representation is proposed to explain the emergence of clinical symptoms as well as behavior in probabilistic decision-making tasks, including the JTC bias (see section 1.3.1), and findings of an increased tendency to switch between responses in so-called reversal learning tasks (see section 1.3.2). Furthermore, within this framework, higher-level belief instability can be considered a 'failure of metacognition' (Adams, Stephan, Brown, Frith, & Friston, 2013; see section 1.2).

This thesis aims to investigate the uncertainty-related concepts of these different fields of cognitive-behavioral research in conjunction in order to explore whether such overarching conceptualizations may indeed explain a variety of symptoms and behaviors observed in individuals with schizophrenia.

1.1 The Bayesian brain account of schizophrenia

Aberrant processing and representation of uncertainty lie at the core of recent hypotheses that suggest schizophrenia, and psychosis in particular, may result from abnormal Bayesian inference processes (Sterzer, Adams, et al., 2018). These hypotheses have been developed within a general Bayesian brain account which proposes that hierarchical Bayesian inference in the human brain underlies perception, action and cognition (Fletcher & Frith, 2009; Karvelis, Seitz, Lawrie, & Seriès, 2018). Specifically, it is assumed that perception and learning are the result of comparing prior expectations or beliefs to observed sensory data. In this process, the extent to which inference, and thus perception and learning, are affected by prior beliefs and/or sensory data depends on the inverse uncertainty (or: precision) associated with both (Sterzer, Adams, et al., 2018). Hence, updating of a prior belief will be less affected by new incoming information if this belief is held with high certainty and/or if the uncertainty about the new information is high. In perception, this would be similar to 'seeing what one expects to see', and in cognition this could, for example, translate into a bias against disconfimatory evidence, i.e. dismissing information that might indicate prior assumptions to be wrong. In contrast, a belief is likely to be changed and updated in response to incoming sensory data if this prior belief is associated with large uncertainty and/or if the certainty regarding the sensory data is high.

Abnormalities in this Bayesian inference process in terms of altered precision (i.e. inverse uncertainty) of prior beliefs and/or sensory data might explain symptoms and deficits observed in psychotic disorders such as schizophrenia (Adams et al., 2013; Corlett, Frith, & Fletcher,

2009; Fletcher & Frith, 2009; Sterzer, Adams, et al., 2018). Specifically, it has been suggested that psychosis might be associated with a tendency to 'overweight' sensory information, leading to an increased influence of such information on perception and belief updating (Sterzer, Adams, et al., 2018). Notably, this could be driven by either decreased precision (i.e. high uncertainty) of prior beliefs, increased precision (i.e. low uncertainty) of sensory information, or both (Adams et al., 2013; Corlett et al., 2009; Fletcher & Frith, 2009; Sterzer, Adams, et al., 2018). Such biased perception towards sensory data may cause the experience of 'strange percepts' in states of delusional mood, where objectively unimportant events or stimuli appear to be particularly salient to patients (Adams et al., 2013; Fletcher & Frith, 2009). Indeed, there is a range of trait-like characteristics of schizophrenia that could be explained by decreased precision of prior beliefs and subsequent overweighting of new data (Adams et al., 2013). Conversely, other phenomena are more likely the consequence of an increase in prior precision (Adams et al., 2013). Importantly, this increase might be of secondary and compensatory nature. For example, while low precision of prior beliefs may cause symptoms of delusional mood in an early stage of the disorder, delusional beliefs, which are commonly characterized by an increased subjective precision (i.e. high certainty or confidence) and a resistance to change, may arise and be maintained as an attempt to explain and resolve the 'strange percepts' (Fletcher & Frith, 2009). Such an increased precision of prior beliefs and a subsequent tendency for top-down guided perception may also explain the occurrence of hallucinations, i.e. seeing or hearing things that deviate from actual sensory evidence (Corlett et al., 2009; Powers, Mathys, & Corlett, 2017). The topic of hallucinations remains controversial, however, as others have proposed that they instead may arise from decreased prior precision and/or increased sensory precision, where reduced top-down suppression of self-generated sensory experiences caused by internal speech might for example explain the occurrence of auditory hallucinations (Fletcher & Frith, 2009). This also raises the question to what extent both decreased and increased precision of prior beliefs and/or sensory data could co-exist during the same stage in psychotic disorders. A solution seems to be offered by the fact that this Bayesian brain system is assumed to be hierarchical. Within this hierarchical belief system, the precisions of beliefs on different hierarchical levels might be independent and may differ across modalities (Corlett et al., 2019; Sterzer, Adams, et al., 2018).

Centering on uncertainty-related processes, the hypotheses developed within this Bayesian brain account of schizophrenia do not only offer an interesting approach to understanding the phenomenology of clinical symptoms but may also explain the dysfunctional decision-making observed in this disorder (Sterzer, Voss, Schlagenhauf, & Heinz, 2018; see section 1.3). Furthermore, they fit well with findings of impaired metacognition in patients as will be described in the following.

1.2 Metacognition in schizophrenia

Metacognition can broadly be described as 'thinking about thinking', i.e. reflective processes regarding one's own cognition and cognitive performance in a given state. This includes for example self-assessments of one's own memory performance (e.g. "I am very confident that I

recalled this story correctly."), representations of one's traits as a cognitive agent (e.g. "I generally struggle with memorizing numbers."), or might refer to strategies of regulating and improving future learning experiences (e.g. "I should repeat each set of numbers four times in my head in order to improve my recall accuracy."). As such, metacognition comprises 'beliefs about beliefs', suggesting a hierarchical structure of the human mind, similar to the Bayesian brain account. Within the Bayesian brain, subjective confidence and metacognitive processes are represented by (hierarchically) higher-level beliefs about lower-level beliefs and their associated precisions (i.e. inverse uncertainty; Adams et al., 2013; Friston, Stephan, Montague, & Dolan, 2014). Consequently, aberrant encoding of uncertainty in the Bayesian brain sense can be considered a 'failure of metacognition' (Adams et al., 2013).

However, the way metacognition is conceptualized and measured differs across the various disciplines of psychology and behavioral science (Norman et al., 2019). Early definitions emphasized the distinction between conscious reflective thought processes and automatic monitoring of one's own thoughts and cognitions (Flavell, 1979). In the same vein, Fernandez-Duque and colleagues (Fernandez-Duque, Baird, & Posner, 2000) propose that metacognition can be broken down into two main components: *metacognitive knowledge*, i.e. knowledge about one's own abilities and demands of the task at hand, and *metacognitive regulation*, such as cognitive monitoring (e.g. error detection), and cognitive control (e.g. error correction, inhibition, etc.). Based on this notion, they emphasize the close link between metacognition and executive function in general, which by definition relies on the ability to monitor and control information processing. Accordingly, they suggest that brain regions that are typically related to executive functions, i.e. regions within the (medial) frontal cortex, constitute the neurobiological basis for metacognitive processes. Their perspective was substantiated by Shimamura (Shimamura, 2000) who draws a comparison between metacognition and different aspects of executive control mechanisms, such as selection, maintenance, updating and rerouting of information, while emphasizing the role of the frontal cortex for all of these processes.

Both impaired executive functioning (Barch, 2005; Fioravanti et al., 2012; Kuperberg & Heckers, 2000; Schaefer et al., 2013) and deficits in metacognition are common findings in schizophrenia (Lysaker et al., 2011; Lysaker, Vohs, Hillis, et al., 2013). Patients seem to be overconfident in errors (Moritz & Woodward; Moritz, Woodward, & Rodriguez, 2006; Moritz, Woodward, Jelinek, & Klinge, 2008), fail to judge their own memory performance correctly (Kircher, Koch, Stottmeister, & Durst, 2007), and divert from clinicians' ratings when judging their general neuropsychological test performance (Moritz, Ferahli, & Naber, 2004). The association with impaired executive functions has been substantiated by correlations with the commonly found lack of awareness or insight regarding the disorder, which itself could be considered a form of metacognitive knowledge (Lysaker, Bryson, Lancaster, Evans, & Bell, 2003). Some studies have focused on more complex metacognitive processes, including the reflection about one's own and others' mental states and to what extent this ability to reflect is diminished in schizophrenia (see e.g., Lysaker & Dimaggio, 2014; Lysaker et al., 2014). Within patients, this measure of metacognitive reflection has been associated with the degree of disorder-related insight (Lysaker et al., 2019). Notably, this understanding of metacognition

moves beyond purely task-related error monitoring and knowledge-based confidence judgments and describes metacognitive ability as a more abstract and overarching construct. In an attempt to reconcile the different conceptualizations of metacognition, Lysaker and colleagues (Lysaker, Vohs, Ballard, et al., 2013) propose that different definitions could be arranged on a spectrum ranging from more discrete to more synthetic metacognitive actions. Here, more discrete activities refer to awareness and accuracy when judging one's own experiences, e.g. error detection and subjective judgment of task performance, whereas more synthetic activities regard the construction of complex and coherent representations of one's self and others.

Metacognitive deficits might contribute to the formation and maintenance of delusions (Garety & Freeman, 1999; Moritz, Woodward, Whitman, & Cuttler, 2005) and are therefore targeted in a cognitive-behavioral program called 'metacognitive training for psychosis', which trains patients to identify and reduce cognitive biases, develop new coping strategies, and aims to increase general metacognitive competence (Moritz & Woodward, 2007).

As stated in the beginning of this section, metacognitive processes are inherently linked to representations of uncertainty. Within the Bayesian brain account, metacognition may for instance refer to higher-level beliefs about the certainty associated with lower-level cognitive representations, for example visual memories. Metacognitive deficits are then inaccurate representations of this certainty, for example when the visual memories have been encoded with much less certainty and more noise than assumed. This could translate into overconfidence judgments when self-assessing one's own recall performance. In schizophrenia, impaired metacognitive abilities (for example expressed as deviations between real and self-rated performance) may thus ultimately reflect misrepresentations of uncertainties within the belief hierarchy as proposed by the Bayesian brain account (Adams et al., 2013). This conceptualization of metacognitive deficits in schizophrenia as the consequence of uncertainty-related processes is substantiated by the aforementioned studies on aberrant confidence ratings regarding cognitive performance. It is further corroborated by the commonly found link between impaired metacognition and probabilistic reasoning. Probabilistic reasoning occurs when decisions have to be made or conclusions have to be drawn under conditions of uncertainty (see also section 1.3). It requires accurate processing and representation of the uncertainties (or: probabilities) at play for an appropriate integration of evidence with prior knowledge. Essentially, this is equivalent to the inference process described by the Bayesian brain account, where sensory experiences are constantly contrasted with prior beliefs, and perceptual and cognitive conclusions are drawn dependent on the uncertainties associated with both (Rausch et al., 2014). It has been suggested that impaired metacognition, such as overconfidence in errors, may conceptually be similar to probabilistic reasoning biases such as the aforementioned JTC bias, where too little evidence is considered before a decision is reached (Moritz et al., 2005; see section 1.3.1). Indeed, various studies found that in schizophrenia, the JTC bias, corresponding decision confidence and metacognitive deficits seem to be related (Andreou et al., 2015; Buck, Warman, Huddy, & Lysaker, 2012; Eisenacher & Zink, 2017; Huq, Garety, & Hemsley, 1988; Moritz et al., 2008; Takeda et al., 2018). Building on the understanding that probabilistic reasoning requires the formation of beliefs and

integration of experiences with prior knowledge, some authors even consider probabilistic reasoning and decision-making per se a function of metacognition (Rausch et al., 2014).

Interestingly, and in line with the postulated overlap of cognitive control and metacognitive processes, cognitive control has also been linked to uncertainty (Mushtaq et al., 2011). Here, the idea is that uncertainty may signal the extent to which cognitive control is needed to solve a given situation. In addition to the conceptual overlap, Mushtaq and colleagues (2011) highlight the findings of various neuroimaging studies that point towards the involvement of similar brain regions in cognitive control- and uncertainty-related processes, including the lateral prefrontal cortex, the parietal cortex, and the anterior cingulate cortex. Furthermore, they summarize results of electrophysiological studies which show how internal outcome monitoring (a key component to cognitive control, measured by event-related potentials) is affected by uncertainty. They conclude that uncertainty may play a major role in cognitive control and monitoring processes and suggest that findings of an association between intolerance of uncertainty and cognitive control functioning in individuals at a clinical high risk for psychosis as reported by Broome and colleagues (2007) reflect that deficits in cognitive control may ultimately be linked to an inability to adapt to uncertainty in these patient groups.

These assumptions fit well with the idea that metacognitive knowledge may, alongside other motivational processes such as personal goals and incentives, affect the amount of effort allocated to learning (Fisher & Ford, 1998) and any cognitive task at hand (Efklides, 2009). In cognitive tasks, mental effort is considered to be the mediating process between one's cognitive abilities and the demands imposed by the task on the one hand (determining the maximum level performance that could be achieved), and final performance on the other (Shenhav et al., 2017). Clearly, metacognitive representations of both one's own cognitive abilities and the demands of the task at hand will shape estimations of how much effort would be required to solve a task correctly (Efklides, 2009). These representations and estimations may in turn affect and interact with motivation and ultimately determine cognitive-behavioral performance. Awareness of high task difficulty, for example, will indicate that increased effort investment is necessary (Efklides, 2009). Moreover, metacognitive skills per se involve by definition strategies to regulate cognitive processing, including increase of effort (Efklides, 2009). Indeed, metacognitive self-regulation was found to be related to effort investment and resulted in higher exam scores in a sample of healthy students (Vrugt & Oort, 2008). Notably, self-reports of effort expenditure can itself be considered a form of metacognitive judgment as they require reflection and insight (Efklides, 2009).

In line with this intimate relationship between metacognition and effort and the metacognitive deficits reported for patients with schizophrenia, schizophrenia has also been associated with diminished effort investment (Green, Horan, Barch, & Gold, 2015). Here, patients, especially those scoring high on negative symptom severity ratings, have been found to avoid physically (Barch, Treadway, & Schoen, 2014; Bergé et al., 2018; Gold et al., 2013) or cognitively effortful tasks (Chang et al., 2019; Culbreth, Westbrook, & Barch, 2016; Gold et al., 2015; Reddy et al., 2018; Wolf et al., 2014), and to employ less effort during unavoidable tasks (Gorissen, Sanz, & Schmand, 2005; Granholm, Ruiz, Gallegos-Rodriguez, Holden, & Link,

2016; Granholm, Verney, Perivoliotis, & Miura, 2006). Evidently, this means that the cognitive performance deficits in neuropsychological tests and tasks observed in individuals with schizophrenia may arise not solely because of a lack of cognitive resources per se, but also due to reduced effort investment (Gorissen et al., 2005). However, results of decreased effort investment in schizophrenia are not always replicated and inconsistencies across studies may be related to differences in task paradigms and samples (Green et al., 2015).

1.3 Decision-making under uncertainty in schizophrenia

A major field of research that centers on uncertainty representation and processing in schizophrenia concerns decision-making under uncertainty. In general terms, decision-making under uncertainty takes place when the outcome following a particular action (i.e. a choice or decision) is not 100% predictable or when conclusions have to be drawn based on limited prior information (Krug et al., 2014).

Performance in decision-making tasks relies on several processes, including action selection and implementation, subjective experience of a given outcome and the formation of beliefs about action-outcome association and weighting of the available actions against each other (i.e. development of preferences; Ernst & Paulus, 2005). When uncertainty is introduced to the task-paradigm, both degree and type of uncertainty may affect all of these processes. Importantly, both in cognitive-behavioral task paradigms and in real life, different types of uncertainty can be distinguished (Payzan-LeNestour & Bossaerts, 2011; Yu & Dayan, 2005): *Irreducible uncertainty,* also referred to as *expected uncertainty* or *risk,* reflects a fixed cue- or action-outcome association. In a given task this might for example mean that a cue X is followed by a fixed outcome Z with a probability of 0.5 (high risk/uncertainty) or that a selection of choice A out of the available choices A and B leads to a reward with a probability of 0.9 (low risk/uncertainty). When this fixed risk is not known but has to be learned through experience (i.e. cue- or action-outcome observations), there is (additional) *estimation uncertainty,* sometimes also referred to as *informational uncertainty.* If this uncertainty cannot be reduced through learning, it is called *ambiguity*. Furthermore, known risks regarding cue- or action-outcome associations might change over time. This is captured by *unexpected uncertainty,* which increases with the instability or *volatility* of the learning environment. For example, while choice A was previously rewarded in 90% of all cases, this probability might drop or increase over time, requiring adaption and relearning (see also Mushtaq et al., 2011).

It has been shown that humans are generally capable of dealing with these uncertain decision contexts and arrive at appropriate decisions (e.g., Behrens, Woolrich, Walton, & Rushworth, 2007; Nassar, Wilson, Heasly, & Gold, 2010). To what extent this may be different in individuals with schizophrenia has been investigated in many different studies and within different lines of research. The most commonly employed paradigms to study decision-making under uncertainty in schizophrenia seem to be the so-called beads task as well as reinforcement and reversal learning tasks. Results of studies employing these tasks are summarized in more detail in the following sections.

1.3.1 Beads task

A plethora of studies has investigated decision-making under uncertainty in patients with schizophrenia using a probabilistic reasoning task called the beads or urn task. Here, a common finding is that patients seem to 'jump to conclusions', meaning they make hasty decisions based on very limited evidence (for review see e.g., Dudley, Taylor, Wickham, & Hutton, 2016; Evans, Averbeck, & Furl, 2015). In the beads task, beads are sampled from one out of two possible containers, e.g. bags (or, depending on task formulation: jars, urns, bags or bottles), containing unlike amounts of beads of different colors. Based on the observed beads, participants have to indicate the bag of origin (Huq et al., 1988). Uncertainty is introduced by the fact that participants do not know which bag the beads are drawn from and by the proportions of differently colored beads in the bags. In 'draws-to-decision' (DTD) versions of the task, participants are free to decide how many beads they want to sample until they feel they can reach a decision about the bag of origin. Here, the aforementioned JTC bias is characterized by premature decisions, i.e. decisions on the bag of origin after very few beads, sometimes only a single one, have been sampled (Dudley et al., 2016). In graded-estimates versions of the task, participants indicate after each draw the probability or certainty for the bead to come from either of the bags, with responses typically recorded on Likert scales (Balzan, Delfabbro, Galletly, & Woodward, 2012; Moritz & Woodward, 2005) or visual analogue scales (Speechley, Whitman, & Woodward, 2010). Here, the JTC bias has been described as over-adjustment in form of radical belief alterations following the confrontation with objectively modest disconfirmatory evidence (Balzan et al., 2012; Garety, Hemsley, & Wessely, 1991; Langdon, Ward, & Coltheart, 2010; Moritz & Woodward, 2005; Speechley et al., 2010). A related finding is that participants with schizophrenia (particularly if currently psychotic) make more extreme probability ratings at the start of a sequence of beads, i.e. they display a higher initial certainty than healthy controls (Adams, Napier, Roiser, Mathys, & Gilleen, 2018; Peters & Garety, 2006; Speechley et al., 2010).

Even though some studies have observed the JTC bias in both delusional and non-delusional schizophrenic patients (Menon, Pomarol-Clotet, McKenna, & McCarthy, 2006; Moritz & Woodward, 2005), there is a large amount of evidence linking the JTC bias to delusions (Broome et al., 2007; Dudley et al., 2016; Falcone et al., 2015; Fine, Gardner, Craigie, & Gold, 2007; Garety & Freeman, 2013; Garety et al., 1991; Huq et al., 1988). It has even been suggested, that the JTC bias essentially contributes to the development and maintenance of delusions (Garety & Freeman, 1999; Garety, Kuipers, Fowler, Freeman, & Bebbington, 2001; Lincoln, Salzmann, Ziegler, & Westermann, 2011). Consequently, the JTC bias has been integrated into cognitive models of psychosis (Garety, Bebbington, Fowler, Freeman, & Kuipers, 2007) and is targeted in the aforementioned metacognitive training intervention (Moritz & Woodward, 2007; see section 1.2).

Different mechanisms have been proposed to underlie this commonly observed reasoning bias in schizophrenia, including motivational deficits due to increased perceived costs of sampling information (Moutoussis, Bentall, El-Deredy, & Dayan, 2011), impulsivity (Dudley, John, Young, & Over, 1997), increased cognitive noise (Moutoussis et al., 2011), a general 'liberal acceptance' of hypotheses, i.e. a lowered threshold to reach a decision (Moritz et al., 2009;

Moritz, Woodward, & Lambert, 2007), and 'hypersalience' of new evidence (see below; Speechley et al., 2010). JTC has also been associated with deficits in working memory which are typically observed in schizophrenia, suggesting that the JTC bias is a secondary result of struggling with information maintenance (Broome et al., 2007; Freeman et al., 2014; Garety et al., 2013). Others have found a relationship between the JTC bias and impaired executive functioning (Woodward, Mizrahi, Menon, & Christensen, 2009) or general intelligence (Falcone et al., 2015; Merrin, Kinderman, & Bentall, 2007). However, neither the results for working memory (Falcone et al., 2015; Krug et al., 2014; Merrin et al., 2007) nor for executive functions or intelligence are always replicated (Krug et al., 2014).

Other explanatory accounts directly address the role that processing of, and dealing with, uncertainty might play in relation to the JTC bias. Here, it has been suggested that an increased 'intolerance of uncertainty' or a higher 'need for closure' might cause the JTC bias in patients. In other words, early conclusions might help to reduce distressing uncertainty (Freeman et al., 2014). While some findings corroborate this account (Broome et al., 2007), many studies did not find an association between JTC and either intolerance of uncertainty (Dudley et al., 2011; Freeman et al., 2014) or need for closure (Freeman et al., 2006; McKay, Langdon, & Coltheart, 2006; McKay, Langdon, & Coltheart, 2007), despite showing that both intolerance of uncertainty and need for closure were consistently associated with delusions.

Regarding the role of uncertainty for the JTC bias, the aforementioned hypersalience account is particularly interesting. The assumption that JTC results from a hypersalience of new evidence rests on the observation of drastic belief alterations in graded-estimates versions of the beads task, where patients tend to display a disproportionate over-adjustment behavior to new evidence (i.e. new colors of beads; Speechley et al., 2010). This interpretation fits well with the aberrant salience account of psychosis which postulates that dysfunctional dopaminergic firing drives attention towards objectively little relevant events and stimuli, skewing perception and belief updating in the cognitive system towards external events (Broyd et al., 2017; Heinz & Schlagenhauf, 2010; Kapur, 2003). This idea is in line with the notion of a Bayesian belief system that is biased towards incoming sensory evidence (Fletcher & Frith, 2009; Sterzer, Voss, et al., 2018). From the perspective of the Bayesian brain account, hypersalience of new evidence in schizophrenia may result from a comparatively high uncertainty of higher-level cognitive representations (or: beliefs; Adams et al., 2013). This, in-turn, nudges the perceptual-cognitive system to rely more on new incoming sensory data as opposed to prior beliefs (Adams et al., 2013; Corlett et al., 2009; Friston et al., 2014; Mishara & Sterzer, 2015), thereby increasing the extent to which processing is driven in a 'bottom-up' manner (Adams et al., 2013; Corlett, Honey, & Fletcher, 2007; Fletcher & Frith, 2009; Horga, Schatz, Abi-Dargham, & Peterson, 2014). Ultimately, this could explain why patients with schizophrenia show over-adjustment in response to disconfirmatory evidence in graded-estimates versions of the beads task: they dismiss their prior hypothesis (belief) about the origin of the sequence of beads quickly, because they hold this belief with high uncertainty, and base each decision more on whatever evidence (i.e. bead color) they are currently presented with. Accordingly, overweighting of new evidence as suggested by the Bayesian brain account of schizophrenia has been proposed as a possible cause underlying the JTC bias (Sterzer, Voss, et

al., 2018). Similar to the hypersalience account, the aforementioned 'cognitive noise' hypothesis (Moutoussis et al., 2011) has been linked to the Bayesian brain account of schizophrenia as well, in that increased uncertainty (i.e. 'noise') may reflect decreased precision of prior beliefs in schizophrenia (Corlett & Fletcher, 2014). As outlined in the previous section, probabilistic reasoning in the beads task can be considered a form of metacognitive functioning (Rausch et al., 2014) and seems to be related to metacognitive deficits in schizophrenia (Andreou et al., 2015; Buck et al., 2012; Eisenacher & Zink, 2017; Huq et al., 1988; Moritz et al., 2005; Moritz et al., 2008; Takeda et al., 2018).

1.3.2 Reinforcement and reversal learning

Another area of research on decision-making under uncertainty in schizophrenia is that of reinforcement learning. In reinforcement learning tasks, participants learn to choose the better of two (or more) available actions, i.e. the one that results in a positive feedback (reward) as opposed to a negative feedback (e.g. punishment, or simply the absence of a reward). Even though subjective uncertainty processing and representation are not always directly addressed in reinforcement learning studies, they may in fact heavily impact learning of action-outcome associations and affect the way individuals react to observed outcomes. Here, associations between action and outcome are commonly probabilistic: i.e. a particular choice will only result in a positive outcome with a given probability. This fixed probability (risk), as well as the fact that it is usually unknown and has to be learned over time (estimation uncertainty), introduces uncertainty on different levels.

To explore patients' sensitivity to these different types of uncertainty, Fujino and colleagues (2016) employed a version of the so-called Iowa Gambling Task (IGT) where participants can freely choose between different actions that with unknown probabilities lead to unknown magnitudes of rewards. This task variant included ambiguous (i.e. probabilities were unknown) and non-ambiguous (i.e. probabilities were known) trials, and studying patients' behavior revealed that ambiguity aversion seemed to be attenuated in schizophrenia. In contrast, Cheng and colleagues (Cheng, Tang, Li, Lau, & Lee, 2012) found that patients avoided risky choices and behaved more conservatively in two different kinds of tasks, even when this led to suboptimal decision outcomes. Studies using the classical IGT have also produced inconsistent results in terms of whether patients differ from controls or not (Sevy et al., 2007). Furthermore, while some authors found that disadvantageous decision-making in the IGT was related to positive symptoms (e.g., Struglia et al., 2011), others have found more evidence for a relationship with negative symptoms (Shurman et al., 2005). Nevertheless, impaired reinforcement learning in schizophrenia is considered a core finding (Deserno, Boehme, Heinz, & Schlagenhauf, 2013), substantiated by the results of a plethora of studies using other paradigms than the IGT. Here, the focus is often less on how uncertainty is being processed but more on how patients represent and react to rewards. In this vein, suboptimal decision-making in reinforcement learning tasks as displayed by patients has been linked to impaired reward representation and aberrant learning from reinforcements (Cicero, Martin, Becker, & Kerns, 2014; Frank, 2008; Gold, Waltz, Prentice, Morris, & Heerey, 2008; Stopper & Floresco, 2014;

Strauss, Waltz, & Gold, 2014). Given that reward processing is tracked by dopamine (e.g., Schultz, 2002), these abnormalities might arise from dysfunctional dopaminergic signaling which is assumed to be a neurochemical marker of schizophrenia (Deserno et al., 2013; Frank, 2008). Multiple studies have found that reinforcement learning deficits in schizophrenia are associated with a particular deficit in learning from positive feedback, while negative feedback processing seems to be relatively intact (Dowd, Frank, Collins, Gold, & Barch, 2016; Strauss et al., 2011; Waltz, Frank, Robinson, & Gold, 2007; Waltz, Frank, Wiecki, & Gold, 2011). Such reinforcement learning deficits seem to be exacerbated when negative symptom load is high, suggesting a role of motivational factors (e.g., Gold et al., 2012; Strauss et al., 2011; Waltz et al., 2011), though this finding is not always replicated (Cicero et al., 2014; Dowd et al., 2016; Hartmann-Riemer et al., 2017). Moreover, some studies have shown that in fact, learning from both positive and negative feedback is impaired in schizophrenia (Cicero et al., 2014; Frank, 2008; Hartmann-Riemer et al., 2017; Waltz & Gold, 2007). Addressing the possible impact of both negative and positive symptoms on reinforcement learning performance, Deserno and colleagues (2013) suggested that, similar to the beads task account, positive symptoms may be related to aberrant salience attribution during outcome processing, while negative symptoms may be related to impaired learning about the values of the different choice options, i.e. integration of action-outcome experiences over time.

Evidently, even if not always addressed directly, salience attribution and learning from reward-related feedback in these uncertain environments can be substantially affected by how (un)certainty is processed and represented to begin with. Additionally, reinforcement learning does not only build on the processing of rewards but also relies on uncertainty-driven exploration of alternative choice options (Strauss et al., 2011). Evidence from an electroencephalographic study of event-related potentials indeed points towards the potential impact of aberrant uncertainty representation in schizophrenia on reward and feedback processing, as patients were fundamentally impaired in distinguishing certain from uncertain task contexts during feedback perception (Clayson et al., 2019). Others have shown that while reinforcement learning impairments sometimes persist across different levels of action-outcome probabilities (Waltz et al., 2007), group differences between patients and controls can be moderated by the probabilities chosen for the task at hand (Koch et al., 2010; Yılmaz, Simsek, & Gonul, 2012), suggesting particular sensitivities to different degrees of uncertainty.

Clearly, the cognitive processes involved in reinforcement learning such as reward-integration, updating and maintenance of the beliefs about the values of the different available actions, and the tendency to explore alternative choice options (Strauss et al., 2014) may depend on other cognitive core functions, such as working memory. Deficits in working memory and other cognitive resources may thus at least in part contribute to impaired performance in reinforcement learning tasks as observed in schizophrenia (Collins, Brown, Gold, Waltz, & Frank, 2014; Collins & Frank, 2012; Deserno, Schlagenhauf, & Heinz, 2016).

Within the field of reinforcement learning, much research has been conducted on the topic of reversal learning, assessing how well participants with or without schizophrenia adapt to changes of the cue- or action-outcome probabilities. Adaptation to such changes might be

related to executive functioning, where cognitive flexibility is a central concept. Many studies on cognitive flexibility in schizophrenia have found impaired performance on tasks measuring task- or set-shifting abilities, i.e. attentional switches between different tasks, tasks sets or task dimensions, indicating a reduced ability to switch and increased perseveration (Ceaser et al., 2008; Pantelis et al., 1999; Pantelis et al., 2009). Interestingly, while set-shifting deficits in these tasks seem to be a reasonably robust findings, results of reversal learning studies conducted within a reinforcement learning context seem to point into an opposite direction. In those tasks, participants have to learn which of two optional choices is most likely linked to a post-choice reward and relearn and adapt their choices once the reward-contingencies for both options are reversed. Using these paradigms, a range of different studies have found *increased* behavioral switching in patients with schizophrenia (e.g., Culbreth, Gold, Cools, & Barch, 2016; Deserno et al., 2020; Kaplan et al., 2016; Schlagenhauf et al., 2014; Waltz et al., 2013), leading to sub-optimal overall performance. Similar to results of traditional reinforcement learning paradigms, impaired reversal learning performance is sometimes (Murray et al., 2008; Waltz et al., 2013), but not consistently (Culbreth, Gold, et al., 2016; Schlagenhauf et al., 2014) related to negative symptoms. In fact, some studies show a relationship between aberrant reversal learning and positive symptoms (Li, Lai, Liu, & Hsu, 2014; Schlagenhauf et al., 2014). Like reinforcement learning deficits, impaired reversal learning might be linked to dopaminergic dysfunction in schizophrenia (Schlagenhauf et al., 2014).

Notably, performance in reversal learning tasks relies on a range of different processes. For one, and similar to simpler reinforcement learning tasks, the contingencies (i.e. cue- or action-outcome probabilities) have to be learned. Studies have shown that patients already struggle to learn these initial contingencies (Culbreth, Westbrook, Xu, Barch, & Waltz, 2016; Murray et al., 2008; Waltz et al., 2013) and that value representations of the different choice options seem to be unstable (Culbreth, Gold, et al., 2016). Then, changes in contingencies have to be detected. This detection happens based on feedback but may further be affected by initial beliefs about change probabilities. Accordingly, and similar to the results of simple reinforcement learning paradigms, impaired performance in reversal learning tasks in schizophrenia has been linked to a particular insensitivity to positive reinforcement (i.e. rewards; Li et al., 2014; Schlagenhauf et al., 2014). However, some studies have shown that the increased switching behavior observed in schizophrenia follows both positive and negative feedback (Culbreth, Gold, et al., 2016; Deserno et al., 2020; Waltz et al., 2013). Others have demonstrated that patients as well as individuals at high risk for psychosis seem to have a generally increased expectation for the different contingencies to change, i.e. a higher-level belief about environmental volatility (Cole et al., 2020; Deserno et al., 2020; Kaplan et al., 2016; Schlagenhauf et al., 2014). Such augmented beliefs about volatility are thought to increase the salience of new evidence, driving individuals with schizophrenia to update their beliefs and concurrent choices heavily based on current evidence and thus explaining the increased tendency to switch (Cole et al., 2020; Deserno et al., 2020). These interpretations were formulated within the context of the Bayesian brain account of schizophrenia, where increased subjective volatility is suggested to represent increased uncertainty surrounding prior beliefs on higher cognitive levels (Cole et al., 2020; Deserno et al., 2020). This explanation possibly extends to the maladaptive and increased switching behavior observed in probabilistic reversal learning tasks where volatility was not

assessed directly in that reduced precisions of beliefs (i.e. increased belief uncertainty) in the Bayesian sense might increase weighting of incoming sensory evidence, ultimately causing increased switching behavior. This demonstrates a strong overlap with the Bayesian brain and hypersalience account of the JTC bias (see section 1.3.1), which might result from similar abnormalities in the uncertainty-dependent balance between high- and low-level cognitive representations (Cole et al., 2020).

Similar to simple reinforcement learning, cognitive core processes such as working memory are likely involved in reversal learning as well. In reversal learning tasks, performance highly depends on internal representations of the task environment overall, including cue- or action-outcome contingencies and probabilities of change, and good performance requires maintenance and updating of the different values associated with the current choice options. Hence, reversal learning deficits in schizophrenia may be linked to impaired cognitive control processes, including working memory (Deserno et al., 2013).

1.4 Neurochemical correlates of uncertainty processing in schizophrenia

Uncertainty is thought to be encoded by neuromodulatory systems (Friston, Kilner, & Harrison, 2006; Yu & Dayan, 2005) and abnormalities in different neurotransmitter systems have been proposed to explain aberrant uncertainty processing in schizophrenia.

The dopamine hypothesis is arguably the most prominent etiological neurochemical hypothesis of clinical symptoms schizophrenia and while it underwent multiple revisions, the crucial role of dopamine still seems to be undisputed (Howes & Kapur, 2009). It has been postulated that positive (i.e. psychotic) symptoms are mediated by a hyperactive mesolimbic dopaminergic pathway, whereas negative symptoms are mediated by hypoactive prefrontal dopaminergic pathways (Davis, Kahn, Ko, & Davidson, 1991). However, this specification has been criticized for lack of evidence and simplification of complex cortical abnormalities (Howes & Kapur, 2009). Beyond the strong focus on dopaminergic dysfunction, the potential role of other neurotransmitter has been acknowledged, including glutamate, serotonin, and gamma-aminobutyric acid (Gill & Grace, 2016). Additionally, abnormal norepinephrinergic signaling might contribute to at least some manifestations of schizophrenic disorders. Findings of a number of studies, including endogenous level studies of norepinephrine (NE) and studies of norepinephrinergic drug-induced modulation of the disorder, seem to support the hypothesis that elevated norepinephrinergic signaling plays a particularly prominent role in the paranoid subtype of schizophrenia (Fitzgerald, 2014; van Kammen & Kelley, 1991). Importantly, Fitzgerald (2014) points out that this does not contradict theories regarding the role of dopamine or glutamate, but instead proposes NE as an additional factor that should be considered.

Out of these different neurotransmitters, dopamine has thus far received most attention in the context of decision-making under uncertainty. The ‘hypersalience’ account (Broyd et al., 2017; Heinz & Schlagenhauf, 2010; Kapur, 2003) inherently links the JTC bias to dopaminergic

signaling by suggesting that it arises from disrupted interactions of dopaminergic pathways between the striatum and cortical regions (Evans et al., 2015; Speechley et al., 2010). Still, despite this presumed dopaminergic involvement, neither a dopaminergic agonists nor a dopaminergic antagonist seemed to affect decision-making in the beads task in a healthy sample (Andreou, Moritz, Veith, Veckenstedt, & Naber, 2014). Moreover, comparison of first episode schizophrenia patients' performance on a classical beads task prior to and two weeks after administration of (typically dopamine-antagonistic) antipsychotic medication did not indicate significant behavioral changes (Menon, Mizrahi, & Kapur, 2008). Together, this suggests a more complex picture regarding the role of dopamine for the JTC bias, which may be case-, state-, and task-dependent, and potentially moderated by other neurotransmitters. Impaired reinforcement and reversal learning have also been proposed to arise from dysfunctional dopaminergic signaling, in particular because of the implications of ventral striatal areas for reward and feedback processing (Deserno et al., 2013; Frank, 2008; Schlagenhauf et al., 2014).

Within the Bayesian brain account of schizophrenia, the possible roles of dopamine, acetylcholine and glutamate for encoding of precision (i.e. inverse uncertainty) and modulation of top-down vs. bottom-up signaling in the brain have been discussed (Adams et al., 2013; Corlett et al., 2009; Fletcher & Frith, 2009; Sterzer, Adams, et al., 2018).

In contrast, NE has received little attention, despite its potential etiological involvement in the disorder and its supposed involvement in uncertainty-related processes. It has been suggested that NE might encode unexpected uncertainty, and as such reflects higher-level uncertainty about the statistical regularities within a given, more or less volatile environment (Yu & Dayan, 2005). NE originates from the subcortical locus coeruleus (LC) from where it is transmitted to widespread sites of the brain (Aston-Jones & Cohen, 2005; Sara, 2009). This 'LC-NE' system is implicated in a broad range of different cognitive functions, including attention and memory (Sara, 2009), as well as executive functions (Logue & Gould, 2014); likely because of its crucial effects on the modulation of arousal (Samuels & Szabadi, 2008). The 'adaptive-gain' theory postulates that the level of activity within the LC determines certain modes of attention and behavior by modulating the 'neural gain', i.e. the responsivity of neurons, through NE. Whereas intermediate LC activity promotes the *exploitation* mode, where attention is focused and task engagement is high, high activity promotes *exploration*, a mode of increased distractibility and proneness to attentional switching (Aston-Jones & Cohen, 2005). This theory is often contrasted with the aforementioned 'unexpected uncertainty' theory. Nevertheless, since contextual change promotes a revision of prior beliefs through exploration and learning, there certainly seems to be conceptual overlap between the unexpected-uncertainty account (Yu & Dayan, 2005) and the adaptive gain theory (Aston-Jones & Cohen, 2005).

Interestingly, norepinephrinergic transmission also seems to be involved in metacognitive processes. A recent study demonstrated that metacognitive ability was enhanced through inhibition of NE in a healthy sample but unaffected by dopamine blockade (Hauser et al., 2017). This was interpreted in line with the neural gain hypothesis according to which NE modulates information processing by regulating general neural gain within the brain, which in turn affects learning (Aston-Jones & Cohen, 2005; Eldar, Cohen, & Niv, 2013). Here, it is assumed that NE

usually increases the contrast between weak and strong signals, potentially leading to a negligence of subtle signals which may be relevant for appropriate metacognitive judgments. Inhibition of norepinephrinergic transmission may then have hindered such NE-modulated information loss, promoting metacognitive awareness (Hauser et al., 2017). The authors point out that their findings could just as well be interpreted when considering that NE in fact signals unexpected uncertainty as hypothesized by Yu and Dayan (Yu & Dayan, 2005). In this context, task-related errors are thought to trigger phasic NE bursts which will erase the information that is currently maintained. The unavailability of this information might then cause poorer metacognitive judgments (Hauser et al., 2017).

Taken together, the hypotheses and findings reviewed here illustrate the promising role of NE as a neurochemical candidate for studies regarding uncertainty-related processes in schizophrenia.

1.5 Measuring uncertainty-related processes

To assess how patients with schizophrenia process, represent and deal with uncertainty, a wide range of different methods can be applied. Some uncertainty-related concepts can be assessed by self-reports and questionnaires, such as intolerance of uncertainty (Freeman et al., 2014) or need-for-closure (McKay et al., 2006). Similarly, while metacognition can be assessed in many different kinds of ways (Norman et al., 2019), research focusing on more synthetic metacognitive abilities in schizophrenia (see section 1.2) typically employ questionnaires, such as the Metacognition Assessment Scale (Semerari et al., 2003) or abbreviated versions thereof. With this scale, metacognition is gauged by a range of different questions that address the extent to which an individual reflects on and understands their own and others' mental processes (Lysaker & Dimaggio, 2014; Lysaker et al., 2014). Yet, with this questionnaire-based approach to metacognition, uncertainty-related processes are not measured directly and cannot easily be inferred from the responses. In contrast, studies on metamemory in schizophrenia, where correspondence of subjective confidence ratings with objective recall performance is of main interest, directly assess degrees of (un)certainty by measuring confidence, using Likert scales (Moritz et al., 2008; Moritz, Woodward, & Rodriguez, 2006; Kircher et al., 2007) or visual analogue scales (Danion, Gokalsing, Robert, Massin-Krauss, & Bacon, 2001). Similar confidence-based analyses have been conducted for other cognitive tasks, regarding for example executive functioning (Koren et al., 2004).

In order to investigate the role of uncertainty-dependent processes in decision making tasks, the degree or level of uncertainty can for example be manipulated in different task conditions so that contrasting performance between those allows for inferences about particular uncertainty-sensitivities. In this vein, patients' performance on an IGT under different degrees of ambiguity has been compared (Fujino et al., 2016), beads task performance was investigated using different ratios of colored beads in the containers (Balzan, Ephraums, Delfabbro, & Andreou, 2017; Ross, McKay, Coltheart, & Langdon, 2015), and both reinforcement and reversal

learning were studied using different degrees of action-outcome contingencies (Cicero et al., 2014; Waltz et al., 2007).

However, several short-comings to these approaches exist. Measures derived from questionnaires or Likert scales, for example, are verbal and explicit by nature. As such, they depend on a participant's particular understanding of the formulations at hand and might be sensitive to response biases. Further, there are caveats when inferring uncertainty-related processes from the observed behavior in different task conditions in the context of decision-making, as there is usually no direct measure of what exactly might drive observed performance differences. Here, more latent or 'implicit' measures provide an additional or alternative approach to explicit assessments. These include for example cognitive modelling of the latent processes that lie behind the observed behavior in decision-making tasks. Another level of explanation can be added through neuro- or psychophysiological correlates. Given the postulated relationship between NE and uncertainty encoding, measures that can capture norepinephrinergic signaling are of particular interest.

1.5.1 Cognitive modelling of latent decision variables

Cognitive models can be used to quantify the latent cognitive processes employed during completion of cognitive tasks. Such models can help to explain observed behavioral data, identify and capture particular cognitive stages, and define the exact mechanisms that underlie those cognitive processes (Lewandowsky & Farrell, 2010; Anticevic, Krystal, & Murray, 2018). They are built on theoretical and mathematical models of how a task is assumed to be solved and which information processing related parameters might affect task behavior. Fitted to task events and the participant's choices, the models then allow to estimate the subjective parameters that seem to drive each participant's behavior and between-subject comparisons of these parameters can shed light on the extent to which participants solved a task differently. Different classes of cognitive models exist, including reinforcement learning models and Bayesian cognitive models (Kriegeskorte & Douglas, 2018). Applied to choice behavior in reinforcement learning tasks, reinforcement learning models (Rescorla & Wagner, 1972; Sutton & Barto, 1998) assume that the task-solving agent strives to maximize future rewards (i.e. positive outcomes of their actions) and learns the values of the different available actions in an accumulative manner on a trial-and-error basis. In these models, uncertainty may be reflected in the learning rate, a parameter which expresses the extent to which a prediction error (i.e. a mismatch between expected and received reward) is weighted during belief updating. Here, a belief concerns the values of the different available actions. In volatile environments, for example, unexpected rewards or punishments may indicate a change in the underlying action-outcome contingencies and should thus strongly influence the action value updating process, i.e. a higher learning rate should be employed. In stable environments, the learning rate should be lower as unexpected events may simply reflect the inherent but stable action-outcome contingences and thus should not promote strong belief changes (see e.g., Behrens et al., 2007; Browning, Behrens, Jocham, O'Reilly, & Bishop, 2015). Bayesian models include for example Bayesian inference models of perception, where prior beliefs and sensory data are combined to

produce the final perceptual experience (Kriegeskorte & Douglas, 2018). However, Bayesian models can also be applied to higher cognitive processes such as decision-making. Bayesian inference models like the Hidden Markov Model can for example describe behavior in reversal learning tasks (Hämmerer et al., 2019; Schlagenhauf et al., 2014). Here, it is assumed that participants use an internal model of the task environment to guide their actions, representing the different hidden task states that may change or reverse over time. Unlike simple reinforcement learning models this assumes a more strategic and less accumulative decision-making and belief updating process. Here, volatility is captured by a parameter called transition probability, which reflects the probability for switches between the different states of the task.

Importantly, different models might be capable of explaining the same kind of data or observations. Model selection criteria, e.g. regarding the goodness-of-fit of models to data, can help to select the best fitting one, though there will always be an unknown amount of alternative (and maybe not yet specified) models that may describe or explain the data equally well (Lewandowsky & Farrell, 2010). Furthermore, any model is still a simplification of the complex cognitive processes it tries to describe or explain and will essentially always be 'wrong' to an extent. Nevertheless, the simplification of processes that otherwise would be too complex to understand can in fact be perceived as beneficial and even a wrong model can be useful in aiding scientific reasoning (Lewandowsky & Farrell, 2010).

1.5.2 Pupillometry: assessing norepinephrinergic activity

It has been established that activity in the LC-NE system is reflected in pupil size changes (Joshi et al., 2016; Rajkowski et al., 1994; Samuels and Szabadi, 2008). Accordingly, different theories about the role of NE for cognition have been tested using pupillometric measures. Pupil size, measured as diameter or area, is usually recorded by an eye tracker. During a cognitive task, pupil dilation is typically caused by the presentation of a task-relevant stimulus, reaching a dilation maximum around 1 – 1.5s after stimulus onset (Eckstein, Guerra-Carrillo, Singley, & Bunge, 2017). Notably, whereas pupil diameter can vary in size between 1.5 to 9 mm, with brighter light conditions promoting constriction and darker light conditions promoting dilation (Sirois & Brisson, 2014), the size of cognitively induced pupil dilation, while highly variable (Mathot, 2018), is very small, with maximum values reported as a ca. 0.5 mm diameter increase (Beatty & Lucero-Wagoner, 2000).

With regards to uncertainty processing, particularly the results of studies testing the adaptive gain theory and the unexpected uncertainty theory are of interest here. Indeed, evidence of pupillometry studies speaks in favor of the adaptive gain theory, demonstrating that explorative behavior is associated with large baseline pupils (Jepma & Nieuwenhuis, 2011), and smaller phasic pupil responses (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010), indicative of high tonic LC-NE activity. Results of pupillometry studies using probabilistic (reversal) learning tasks provide even more direct evidence for the relationship between pupil dilation and uncertainty. Here, it has been found that pupil dilation is affected by the degree of outcome surprise (an inverse indicator of uncertainty) as well as unexpected uncertainty, i.e.

environmental volatility (Browning et al., 2015; Hämmerer et al., 2019; Lawson, Mathys, & Rees, 2017; Nassar et al., 2012; Preuschoff, t Hart, & Einhauser, 2011). This has rarely been investigated in schizophrenia, though early studies demonstrated that in patients, pupil dilation seems to be less affected by the probabilities, i.e. uncertainties, associated with the presented stimuli (Steinhauer, Hakerem, & Spring, 1979; Steinhauer & Zubin, 1982).

It shall be noted, that pupil size and pupil dilation have been related to many more cognitive processes as summarized in multiple reviews (Eckstein et al., 2017; Laeng, Sirois, & Gredebäck, 2012; Mathot, 2018; Sirois & Brisson, 2014; van der Wel & van Steenbergen, 2018), including mental effort and cognitive control, task difficulty and demand, and motivational salience (Eckstein et al., 2017; Laeng et al., 2012; van der Wel & van Steenbergen, 2018). These widespread associations are likely due to the central role of LC activity for general arousal (Samuels & Szabadi, 2008).

1.6 Aims of the thesis

As outlined in the preceding sections, many of the cognitive-behavioral deficits and biases observed in schizophrenia may be related to similar underlying abnormalities regarding processing and representation of uncertainty. As hypotheses derived from the Bayesian brain account propose, these may further explain the emergence and maintenance of clinical symptoms. Despite this overlap, metacognition, JTC, reversal and reinforcement learning are rarely studied in conjunction, and the potentially shared role of higher-level uncertainties, e.g. volatility-driven unexpected uncertainty, for all of these processes seems to be seldom discussed. Furthermore, the role of NE regarding uncertainty processing in schizophrenia is has received little attention. The aim of this thesis was to investigate these topics using different tasks, cognitive modelling and pupillometry. *Paper I* explored whether decision-making under uncertainty in the beads task might be affected by higher order uncertainty processes, testing whether patients would show a particular overestimation of volatility, and whether this would be related to metacognition, tapping the same sort of aberrant underlying mechanisms in terms of higher-level uncertainty representation. Following these investigations, *paper II* aimed to investigate in more detail whether patients with schizophrenia would show a particular sensitivity to different kinds of uncertainty and if this was accompanied by abnormal norepinephrinergic activity as indexed by pupil dilation. Investigating a subgroup of the same sample, *paper III* assessed working memory performance, which has been linked to decision-making under uncertainty, as well as self-reported and objectively measured effort. This taps into metacognitive processes, as self-evaluation of invested effort requires insight and reflection, and it further gives an indication to what extent measured performance can be considered a reliable measure of actual cognitive ability. This consideration is vital for the interpretation of virtually all cognitive-behavioral results obtained from patient samples.

2 Methods and materials

Table 1 presents an overview of the different samples, tasks, and measures used in each of the papers.

Table 1. Overview of study designs

	Paper I	Paper II	Paper III
Sample[a]	SZ (n = 21) ASD (n = 19) HC (n = 46)	SZ (n = 31) HC (n = 30)	SZ (n = 29) HC (n = 30)
Cognitive tasks	Beads task Visual working memory task	Probabilistic prediction task	Digit span task
Questionnaires and scales	-	PANSS	PANSS Motivation and task demand questionnaire
Pupil measure	-	Trial-wise pupil dilation to outcome	Condition-wise pupil dilation at last digit
Cognitive models	Ideal Bayesian observer	HMM_{RP} (winning model)	-

Notes: SZ = disorders from the schizophrenia spectrum, ASD = autism spectrum disorders, HC = healthy (i.e. non-psychiatric) control group; PANSS = Positive and Negative Symptoms Scale; HMM_{RP} = Hidden Markov model with separate learning sensitivities for positive and negative feedback

[a] equivalent core sample for paper II and paper III, with 2 participants excluded for paper III

2.1 Participants and ethics

Inclusion criteria for *paper I* were: (1) 18 to 60 years of age, (2) no current suicide intent, (3) no substance dependence, (4) no PTSD, acute anorexia, acute psychosis, (5) IQ above 80, and (6) a primary diagnosis from the schizophrenia spectrum (SZ group) or high-functioning autism/Asperger (ASD group) or no psychiatric diagnosis at all (HC group); according to the 5th version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V; American Psychiatric Association, 2013). In total, 93 participants were recruited. Patients with schizophrenia were contacted through a clinician at St. Olavs hospital in Trondheim, Norway,

persons with autism were recruited through *Autismeforeningen* and traditional recruiting methods including fliers and social media posts.

Inclusion criteria for *paper II* and *paper III* were: (1) 18 to 65 years of age, (2) capacity to give informed consent, (3) very good command of the German language, (4) IQ above 80, (5) normal or corrected-to-normal eyesight, (6) no history of neurological disorders, (7) no substance dependence, (8) no recreational drug consumption within one week prior to the assessment (excluding alcohol, nicotine, and caffeine), (9) a primary diagnosis of schizophrenia or schizoaffective disorder (SZ group; DSM-V, American Psychiatric Association, 2013) or no psychiatric diagnosis at all (HC group). Participants in the SZ group were in- and outpatients of the Department of Psychiatry and Psychotherapy of the University Medical Center Hamburg-Eppendorf (UKE), Germany. The sample for both papers was equivalent, with the exception of two participants who were not included in *paper III* due to a data recording error. Hence, *paper II* included 61 and *paper III* 59 participants. For *paper II* and *paper III*, SZ group and HC group were matched on the demographics variables age, gender and education.

Participants gave written informed consent prior to all studies which were conducted in accordance with the guidelines of the Declaration of Helsinki. The study conducted in Norway (*paper I*) was approved by the Central-Norwegian regional committee for medical and health research ethics [reference number: 2014/1648]. The study conducted in Germany (*paper II* and *paper III*) was approved by the local ethics committee of psychologists at the UKE.

2.2 Cognitive-behavioral tasks

2.2.1 Working memory and metamemory

In *paper I*, a visual working memory task (see Fig. 1 and *paper I* for details) was applied, where both working memory accuracy and implicit metamemory were measured. Here, participants had to memorize a target shape that was presented on screen and then point out its location in an array of similar shapes. In this array, continuously modified shapes were arranged according to their similarity, with shapes in closer proximity appearing more alike. Thus, the angular deviation between the target shape and the shape pointed to by the participants represented a continuous error of recall accuracy. When selecting the supposed target shape in the array of shapes, participants also set a visual capture area around as many of the surrounding shapes as they deemed necessary to certainly include the target shape. The proportion of trials in which their capture area indeed included the target shape thus represented an implicit index for metamemory, as overconfidence would lead to a smaller proportion.

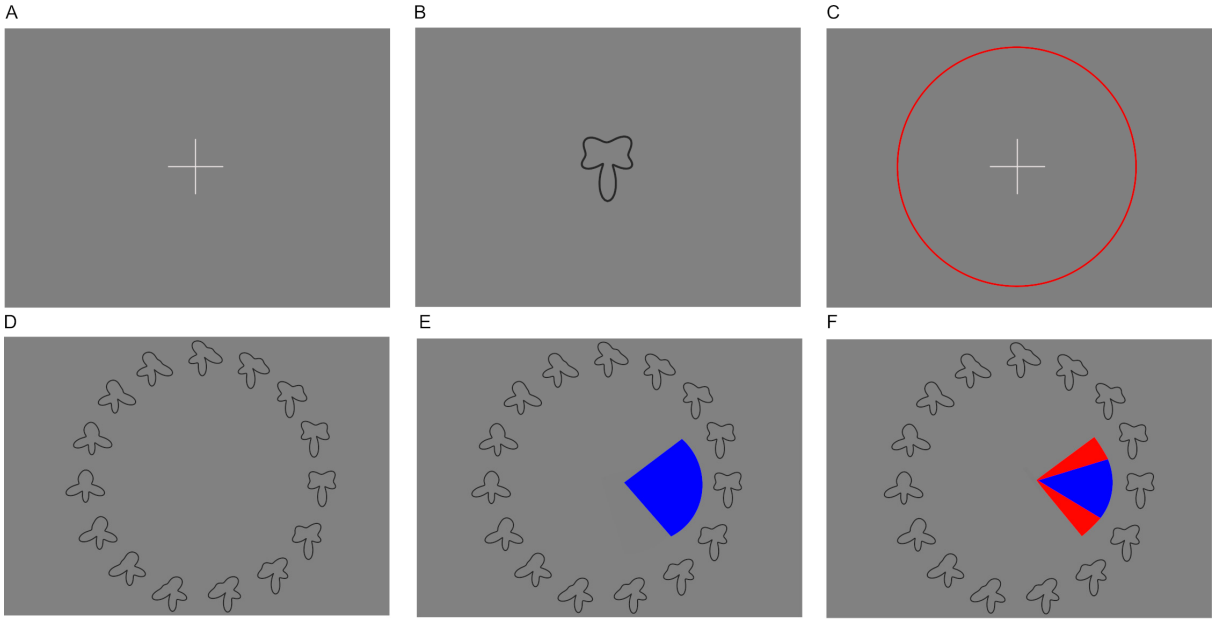


Figure 1. Exemplary display of the visual working memory task

Figure is taken from *paper I*. (A) The appearance of a fixation cross marks the start of a trial. Clicking on it initiates the presentation of the target shape that has to be memorized (B). It is drawn at random from a pool of 30 continuously modified shapes. The target shape remains on screen for 1s, and is followed by another fixation cross (C). By clicking on the fixation cross, participants initiate the recall phase (D). Here, all 30 shapes are presented in a circular array (for better visibility only 15 shapes are displayed). (E) Participants make their guess on the location of the target shape by clicking on it in the circle. They set a capture area symmetrically surrounding their first guess, while adjusting its size to make sure that the target shape is truly included. Task points are subtracted according to the size of overshoot, i.e. setting the capture area larger than necessary, in order to demotivate participants from covering the area of the whole circle each time. (F) Participants receive feedback about the true location of the target shape. If the capture area includes the target (as in the example Figure), the region of overshoot of the capture area is shown in red.

In *paper III*, a visual and computerized version of the classical digit span forward task of the Wechsler adult intelligence scale (WAIS-IV; Wechsler, 2008) was administered (see *paper III* for details). On each trial, digits were presented one after another with a 1s interval. The amount of digits presented in a row increased every two trials from a minimum amount of two to a maximum amount of nine digits, thus providing trials of different task load conditions. After each trial, participants had to recall all digits in the correct order and indicate their responses on a keyboard. Proportion of correctly recalled digits until the first error was made was recorded per trial (trial-wise accuracy) and the maximum amount of digits recalled correctly during the whole task (maximum digit span) was recorded as a measure of working memory capacity. The data of this assessment were also used in *paper II* to control for working memory capacity differences between the HC and the SZ group.

2.2.2 Decision-making under uncertainty

For *paper I*, a modified version of a graded-estimates variant of the beads task was implemented to test for the effect of environmental instability, i.e. volatility, on decision-making and JTC-like variables (see Fig. 2 and *paper I* for details). Five different sequences of 20 beads each were drawn with replacement from one of two bags containing 80 white and 20 black beads and the converse. The first bag to start drawing from was chosen at random at the beginning of each sequence, but the bag of origin could change during the sequence with a probability of 0.04 per draw. Participants were instructed that this amounted to a ca. 50% chance of observing a bag change during a sequence of beads. Naturally, the drawn sequence depended on the color distributions in the bag (i.e. the fixed probabilities for 'white' and 'black') as well as the probability of change, i.e. the degree of volatility. After each draw, participants indicated their certainty about the bag of origin on a visual analogue scale that ranged from 0 = 'absolutely sure it comes from the bag with more black beads', to 1 = 'absolutely sure it comes from the bag with more white beads'. The indicated locations on the scale were mapped to probabilities. These probabilities were used for the calculation of JTC-like outcome measures, with initial certainty calculated as the average size of probability ratings for the first bead of each sequence, and disconfirmatory belief updating calculated as the average belief (i.e. probability) change in favor of the alternative hypothesis upon occurrence of a bead colored differently than the two or more preceding beads. For both variables, higher values were equivalent to increased JTC-like behavior.

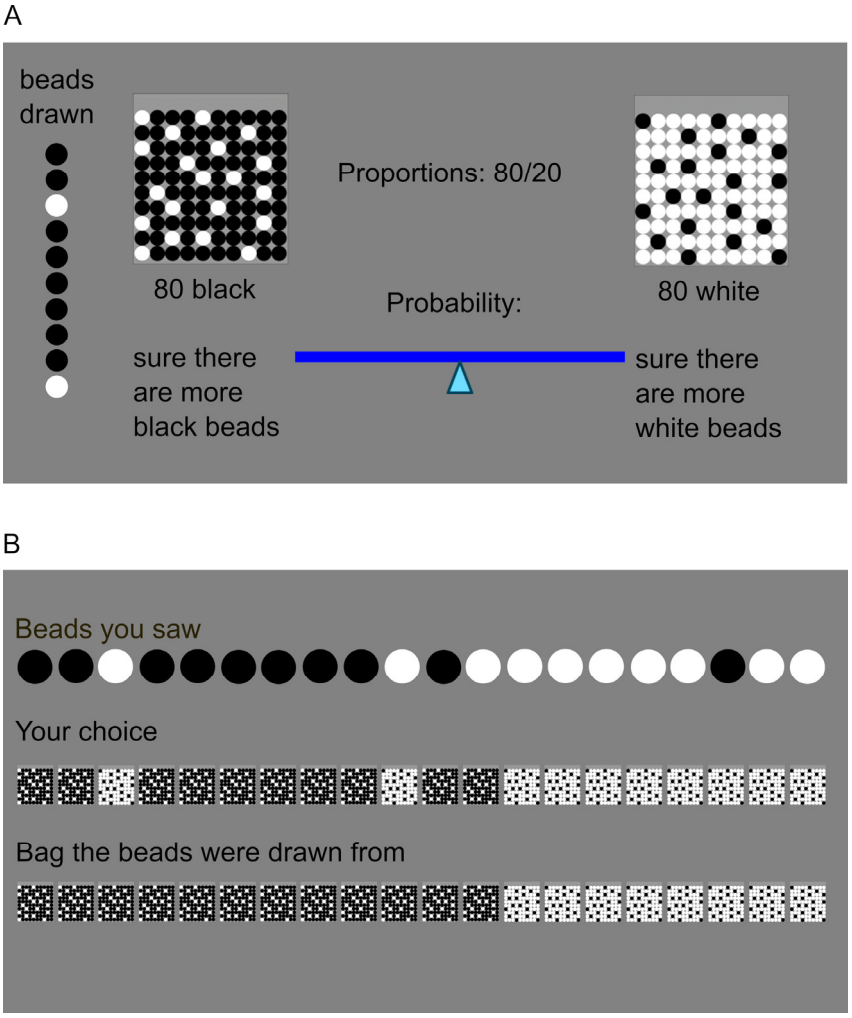


Figure 2. Exemplary display of the beads task

Figure is taken from *paper I*. (A) The 10th trial of a given sequence is displayed as an example. The two bags from which beads are drawn, containing either 80 black and 20 white beads or the converse, are presented on screen. The color of the drawn bead on a given trial is displayed together with the history of all beads of the given sequence. Participants indicate their certainty about the bag of origin by adjusting the position of the marker on a visual scale within a 10s time window after a bead was drawn. After every trial, this slider is reset to the center of the scale. (B) Once all 20 beads of a sequence have been drawn, feedback is provided regarding the history of observed beads, choices made (in terms of favoring one bag over the other), and the actual bag of origin for each draw.

For *paper II*, a probabilistic decision-making task similar to standard reversal and reinforcement learning tasks was developed: the probabilistic prediction task. The task included different conditions of volatility and risk to assess to what extent processing of these different kinds of uncertainty would be affected in schizophrenia (see *paper II* for details). On each trial, participants had to predict whether the upcoming stimulus was going to be a Gabor patch tilted to the left or the right of the center (orientation ± 45°; see Fig. 3A). They were informed that with a fixed probability over the course of an unknown number of trials, one would be more likely than the other. Participants were instructed to try and find out which one that was and to adapt their predictions to make as few prediction errors as possible. In a first, volatile block of

the task they were further informed that these probabilities would change at unannounced change points during the task and encouraged to keep track of such hidden changes, i.e. adapt their predictions accordingly. In a second, cued block of the task such changes were announced while the actual probabilities for either 'left' or 'right' remained hidden. The probabilities for either 'left' or 'right' alternated between 85:15 and 60:40 and the converse (15:85, 40:60) after every 20 ± 4 trials in both blocks (see Fig. 3B). This set-up defined task periods of low (85:15, 15:85) and high (60:40, 40:60) risk, whereas the different blocks, each spanning 160 trials, constituted task environments of low (second, cued block) and high (first, volatile) volatility. The main behavioral variables of interest were proportion of accurate predictions, defined as a prediction of the current majority stimulus (left or right Gabor patch), and proportion of choice switches (i.e. a change in a prediction from trial t to trial $t + 1$).

In both *paper I* and *paper II*, subjectively perceived volatility as a higher-level belief of uncertainty was a measure of major interest. For both papers, this subjective volatility was estimated with the help of cognitive models as described in section 2.5.

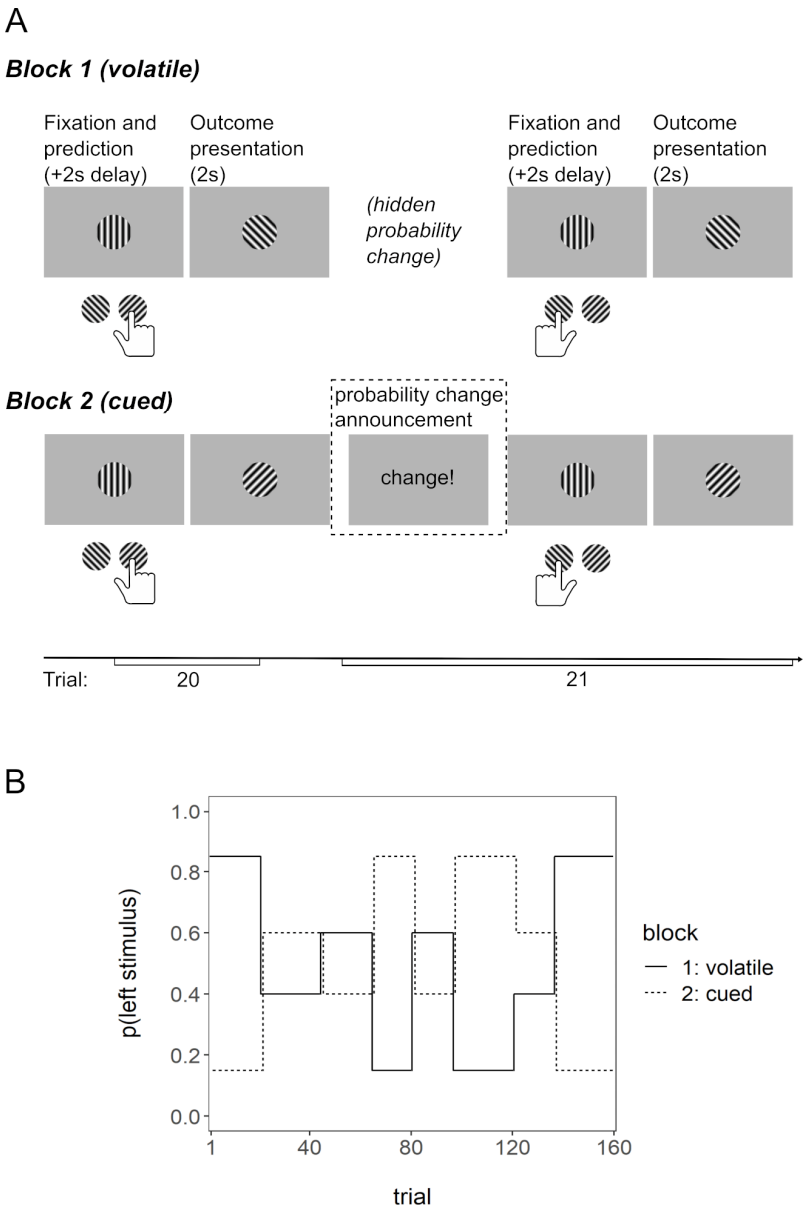


Figure 3. Probabilistic prediction task

Figure adapted from *paper I*. (A) Example trials 20 and 21 for the volatile and the cued task block, respectively. A trial begins with the presentation of a vertically striped Gabor patch. Upon occurrence, participants predict whether it is going to tilt to the left or the right from the center by pressing corresponding 'left' and 'right' buttons on a keyboard. Two seconds after a prediction is made, the outcome is displayed for another 2s. A new trial (i.e. a new prediction) is prompted by the re-occurrence of the vertically striped patch. Within the first, volatile task block, changes in the underlying probabilities for 'left' or 'right' are hidden. Within the second, cued task block, these changes are announced in the beginning of the respective change trial (see B) with a 'change' message on screen. To ensure participants' registration of that change message, they have to press 'enter' upon occurrence before continuing with the task. (B) In both task blocks, probabilities for the left- (p(left)) and the right-tilted (1-p(left)) Gabor patch change every 20 ± 4 trials. These changes are hidden in the volatile block (solid line), and announced (see A) in the cued block (dashed line). The order of probability conditions as well as the timing of change points are identical across blocks, while the identity of the respective majority stimulus (left or right) is inverted.

2.3 Questionnaires and interviews

For *paper I*, the demographic variables age, sex and education were collected in a questionnaire. For patients, currently used medication type was additionally recorded.

For *paper II* and *paper III*, demographic (such as age, gender and education) and clinical variables (such as time since onset of the disorder, type and dose of medication), were recorded in a short interview. To confirm presence (SZ group) or absence (HC group) of clinical diagnoses, the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998) was administered. Current severity of positive and negative symptoms was assessed using the Positive and Negative Symptoms Scale (PANSS; Kay, Fiszbein, & Opler, 1987). Negative symptom scores were calculated based on a suggestion by van der Gaag et al. (van der Gaag et al., 2006), as the validity of the original negative symptoms scale of the PANSS has been criticized (Khan et al., 2013; van der Gaag et al., 2006). Premorbid verbal intelligence was assessed with the German multiple choice vocabulary test (Lehrl, Triebig, & Fischer, 1995).

In *paper III*, a newly compiled motivation and task demand questionnaire was applied to assess self-reported motivation, invested effort, and subjective task demand after completion of the digit span task. Items were adapted from the Momentary Influences, Attitudes and Motivation Impact on Cognitive Performance Scale (Moritz, Klein, et al., 2017; Moritz, Stöckert, et al., 2017) and the NASA Task Load Index (Hart & Staveland, 1988), and responses were recorded on a 4-point Likert scale ranging from 1 (completely disagree) to 4 (completely agree; see supplementary material of *paper III* for details).

2.4 Pupillometry

During completion of the probabilistic prediction task in *paper II*, pupil size was recorded with an infrared video-based eye tracker (Eyelink 1000, SR Research) at a sampling rate of 500 Hz. In seven cases the right eye was recorded, and in all remaining participants the left. Measures of trial-wise pupil size changes in response to feedback (i.e. outcome presentation: left or right tilted Gabor patch) were used to assess whether uncertainty-related feedback processing differed between groups (see section 2.5.2).

In *paper III*, pupil size was recorded with the same eye tracker system during digit presentation (i.e. the encoding period) of the digit span task, again with a sampling rate of 500 Hz. Here, pupil dilation at the last digit of each trial indicated invested mental effort as established previously (e.g., Granholm et al., 2016).

For both *paper II* and *paper III*, the recorded pupil signal was cleaned from eye blinks and other artefacts based on a customized filter, using the signal's velocity and employing cubic-spline interpolation (Mathôt, Fabius, Van Heusden, & Van der Stigchel, 2018). The cleaned signal was subsequently smoothed with a 3 Hz low pass Butterworth filter. If blinks or artifacts caused a missing signal for more than 1000 consecutive milliseconds, the pupil signal in the respective period was not interpolated but treated as missing data in subsequent analyses.

For *paper II*, the final signal was z-scored per participant for each block of the prediction task. For each trial, baseline-corrected pupil dilation was calculated by subtracting the average z-scored signal over the 500ms preceding the onset of the outcome from each sample of the signal during outcome presentation. If 50% of the signal during these periods of interest were missing or interpolated, data of the whole trial was treated as missing during data analysis.

For *paper III*, trial-wise baseline measures were calculated as average pupil size of the 200ms preceding the presentation of the very first digit. For each trial, pupil dilation at the last presented digit of that trial was then calculated as percentage change from the baseline, averaged across the 1s presentation duration. For subsequent analysis, all trials where more than 25% of the data were missing and/or more than 50% of the signal during last digit presentation was interpolated, were treated as missing in subsequent analyses.

2.5 Data analysis

For all papers, normality of variables and of statistical model residuals were assessed through visual inspection of QQ-plots and frequency distributions, as well as Shapiro-Wilk tests. When variables violated this assumption, non-parametric methods were used for simple group comparisons (i.e. Mann-Whitney U tests, Kruskall-Wallis tests), and for simple tests of variable associations (i.e. Spearman correlations). When model residuals violated this assumption, dependent variables were transformed in correspondence with the nature of the skew in the data (e.g. cube root, square, or log transformation).

Statistical testing was conducted with a significance level of 0.05 using R (R version 3.5.1; R Core Team, 2018). Paper specific analysis details are summarized in the following.

2.5.1 Paper I

2.5.1.1 Cognitive modelling

To estimate participants' perceived volatility of the beads task environment, participants' trial-wise probability estimates (indicated on the visual analogue scale) were compared to an ideal Bayesian observer model (see supplement of *paper I* for details). In the ideal Bayesian model, trial-wise probabilities are based on the colors of all draws before and including the current trial n, and on the presumed volatility parameter v. Similarly, a participant's probability rating for a given trial n in a given sequence k ($\tilde{p}_{k,n}$) should be their estimate of the theoretical probability $P(x_{k,n}|z_{k,1}, \dots, z_{k,n}, v)$, with x indicating the bag of origin ($x_{k,n}$ = 0 if bag with more black beads, $x_{k,n}$ = 1 if bag with more white beads) and z indicating the color of the given bead ($z_{k,n}$ = 0 if white, $z_{k,n}$ = 1 if black). The volatility parameter v was therefore estimated as the parameter value that would minimize the distance between the theoretical probabilities and a participant's estimated probabilities ($\tilde{p}_{k,n}$).

2.5.1.2 Statistical analyses

To compare performance between the three groups (SZ, ASD, and HC group), one-way ANOVAs were performed and followed up with Tukey's Honest Significant Difference tests in case of a significant group difference. Here, disconfirmatory belief updating was the only variable that was log-transformed. Using the untransformed version of all variables, all ANOVA results were verified with Kruskal-Wallis tests and followed up with Bonferroni corrected Dunn's Tests in case of a significant group difference. Accordingly, both parametric (η^2) and non-parametric (ε^2) effect sizes were reported. To investigate the relationship between variables across the whole sample, Spearman correlations were calculated. Due to the bimodal distribution of volatility, exploratory analyses were added using Gaussian Mixture models. For details, see *paper I*.

2.5.2 Paper II

2.5.2.1 Cognitive modelling

Various alternative cognitive models were fitted to the data of the prediction task in order to allow for the fact that different strategies may be employed when solving the task. These models encompassed a win-stay-loose-shift model (Worthy & Todd Maddox, 2014), four Reinforcement Learning models (den Ouden et al., 2013; Gläscher, Hampton, & O'Doherty, 2008; Pearce & Hall, 1980; Rescorla & Wagner, 1972), and two versions of a Hidden Markov Model (HMM; Schlagenhauf et al., 2014; see supplementary material of *paper II* for details). Models were fitted for each block and separately for the SZ and the HC group, using participants' trial-wise choices (prediction of 'left' or 'right' stimulus) and outcome observations (prediction correct or incorrect). Importantly, additional variants of all models were formulated for the cued block, specifying belief resets at every point of an announced probability change. Model estimation was conducted within the hierarchical Bayesian framework, using a Markov chain Monte Carlo method (Ahn, Haines, & Zhang, 2017; Gelman et al., 2013). The winning model for all groups and task blocks, i.e. the model that explained the observed behavior best, was a variant of the HMM. This model assumes that participants have an internal model of the task that includes information about the instability (i.e. volatility) of the task environment. Specifically, participants are assumed to make their choices depending on what state of the task ('left is currently the majority stimulus' vs. 'right is currently the majority stimulus') they believe to be in. Trial-wise states beliefs are updated based on both the history of observed choice-outcome pairs and the assumed transition probability γ, which represents the expected change rate between the two states. In addition to this parameter γ, i.e. participants' subjective volatility of the task environment, the parameters c and d were estimated, with c reflecting the sensitivity to positive (correct prediction) and d reflecting the sensitivity to negative (incorrect prediction) feedback. For each task block, all those parameters were further estimated on the group level. Additionally, trial-wise measures of Bayesian surprise and belief entropy were calculated from the extracted state beliefs (see *paper II* for details). Bayesian surprise indicates the extent to which a current state belief should be updated

given a new outcome observation, whereas belief entropy reflects current uncertainty regarding the hidden states.

2.5.2.2 Statistical analyses

SZ and HC group were compared on demographic variables, premorbid verbal intelligence and working memory capacity using Chi-squared tests and Mann–Whitney U tests.

Linear mixed-effects models were employed to investigate the relationships between behavioral performance, task conditions (risk and volatility) and group membership on the one hand, and latent variables derived from the HMM (i.e. Bayesian surprise and belief entropy), task conditions, group membership and pupil dilation on the other.

To test for effects of task block, group and their interaction on group-level parameters of the HMM, i.e. transition probability as well as sensitivity to positive and negative feedback, posterior sampling distributions were contrasted (Zhang, Lengersdorff, Mikus, Gläscher, & Lamm, 2020). The relationship between the corresponding subject-level parameters and clinical symptoms as measured with the PANSS was assessed using Spearman correlations.

2.5.3 Paper III

Responses to the items of the post-assessment motivation and task demand questionnaire were submitted to an exploratory factor analysis with varimax rotation to identify the latent factors behind the single responses. This revealed two factors. First, a 'motivated effort' scale, which reflected self-reported motivation to do well and to invest effort into solving the task. Second, a 'perceived ease' scale, which reflected subjective difficulty of the task and the (inverse) extent to which participants felt challenged and strained by the task (see *paper III* for details). Average scale scores for each participant were used in subsequent analyses. For group comparisons of working memory capacity, questionnaire responses, and clinical assessments (such as medication), Mann–Whitney U-tests were used due to violated normality assumptions. Accordingly, non-parametric effect sizes were reported as Cliff's delta d_C. To test for relationships between variables, Spearman correlations were conducted. For trial-wise analyses of recall accuracy, load condition, group, and pupil dilation, linear mixed-effects models were built hierarchically and compared with the likelihood-ratio chi-squared test (see supplementary material of *paper III* for details of model comparison).

3 Summary of papers

3.1 Paper I

Aims and background:

Recent Bayesian brain hypotheses have suggested that aberrant representation of uncertainties regarding prior beliefs and/or incoming sensory information leads to excessive 'bottom-up noise' in schizophrenia, which might explain the emergence of delusions and cognitive biases such as the JTC bias (e.g., Adams et al., 2013; Corlett & Fletcher, 2014; Fletcher & Frith, 2009). This might be linked to a perception of the world as unstable, i.e. volatile (Cole et al., 2020; Deserno et al., 2020): in an unstable world, the most recent information may seem most reliable, hence, the salience of new events is increased.

Furthermore, a misrepresentation of (un)certainties regarding higher-level beliefs possibly extends to metacognition (Adams et al., 2013), i.e. beliefs about beliefs, explaining findings of reduced metacognitive ability in patients. This conception fits well with the observed correlations between the JTC bias, decision confidence and metacognitive deficits in schizophrenia (Moritz et al., 2008). Interestingly, within the Bayesian brain account, similar hypotheses have been proposed to explain clinical and cognitive symptoms observed in autism spectrum disorders (e.g., Van de Cruys et al., 2014; van Schalkwyk, Volkmar, & Corlett, 2017). Therefore, this study aimed to test to whether uncertainty-related biases such as overestimation of volatility and deficits in metacognition were unique to patients with schizophrenia (SZ group) when compared to both a healthy control group (HC group) and individuals with autism spectrum disorders (ASD group). Across the whole sample, the association between probabilistic-decision making biases, metamemory (as a proxy for implicit metacognitive ability) and aberrant volatility processing was investigated. Moreover, their relationship with working memory accuracy was assessed to control for the potential effect of general cognitive resources.

Methods:

A variant of the beads task was administered where volatility was introduced by including a probability for the bag of origin to change throughout a sequence of drawn beads. Participants indicated probabilities for the bag of origin on a trial-wise basis, based on which their assumed subjective volatility was estimated using a Bayesian observer model. Visual working memory and implicit metamemory were measured with a separate task where participants had to remember one shape at a time and were asked to identify it in an array of shapes after a 1s delay. They were told to set a capture area within the array of shapes to make sure the actual target shape was included. Thus, this indicated their certainty about their first guess on the target shape location. Participants where the capture area rarely included the target shape were assumed to be overconfident.

For method details, see section 2.2 and *paper I*.

Results:

No significant group differences were found in any of the variables assessing probabilistic decision-making in the beads task, including initial certainty about the bag of origin after the first drawn bead, belief updating when faced with disconfirmatory evidence, and model-based estimated volatility. Groups also did not differ in terms of the metamemory scores. However, the SZ group demonstrated significantly lower visual working memory accuracy. Since estimated volatility followed a bimodal distribution, with ca. half of the participants estimating it close to the optimal as designed by the task paradigm (first cluster) and the other half clearly overestimating volatility (second cluster), additional exploratory analyses were conducted. The bimodality was modelled with Gaussian mixture models, where the winning model revealed that within the second cluster of volatility, while participants of all groups were represented, volatility estimates were higher for both the SZ and the ASD group. This suggests that among those participants who overestimated volatility, individuals with SZ and ASD did so to an even larger extent.

Correlations between the different variables revealed a significant relationship between volatility and disconfirmatory belief updating, between volatility and metamemory, as well as between volatility and memory accuracy.

Conclusion:

Despite the absence of main group differences, results of the exploratory analyses suggest that there may be an overestimation of volatility in subgroups of participants with autism and schizophrenia. Both conditions are very heterogeneous in nature, possibly explaining why cognitive-behavioral performance was not consistent across participants within the same clinical sample.

The correlations revealed associations between different variables reflecting (mis)estimation of uncertainty, including volatility estimation and metacognitive processes. This suggests that overestimation of environmental uncertainty (such as volatility) and misestimation of one's own cognitive capacity may be affected by similar mechanisms, potentially driven by higher-level uncertainty calculations in the mind's belief hierarchy. This corresponds to the interpretation of aberrant uncertainty processing and representation as a 'failure of metacognition' (Adams et al., 2013). Notably, the correlation of uncertainty-related variables with memory accuracy might point towards an additional role of general cognitive ability in driving these processes. Deficits in general cognitive ability may for example limit the ability to properly understand task instructions and to translate the instructed volatility value into behavior.

3.2 Paper II

Aims and background:

Using a newly developed probabilistic prediction task, *paper II* investigated in more detail if individuals with schizophrenia are particularly sensitive to volatility or whether processing of lower-level uncertainties such as risk is also affected. Unlike in *paper I*, the probabilistic value of volatility was not instructed but had to be inferred from experience to minimize the effect that (mis)understanding of task instructions may have on performance. To further control for working memory capacity, participants with schizophrenia (SZ group) were compared to the healthy control group (HC group) on a working memory capacity measure derived from a simple digit span task. In order to test the hypothesis that aberrant uncertainty processing in schizophrenia is linked to abnormal norepinephrinergic signaling, trial-wise pupil dilation was recorded.

Methods:

Groups were matched on relevant demographic and educational variables, including premorbid verbal intelligence.

In the probabilistic prediction task, participants had to predict the upcoming stimulus on each trial to be either left- or right-tilted. They knew one would be more likely than the other but that this would change throughout the task, and that these changes would be hidden in the first, volatile task block, and announced in the second, cued block. Further, risk conditions were manipulated by including sub-blocks of high risk (probabilities for left/right stimulus were 60:40 or 40:60) and of low risk (probabilities for left/right stimulus were 85:15 or 15:85).

Task performance was measured as the proportion of times that the current majority stimulus was correctly predicted and the proportion of times that participants changed their predictions from one trial to the next. Fitting a cognitive model to task behavior allowed for the estimation of subjective volatility, as well as subjective parameters indicating to what extent participants based their belief updates on positive or negative feedback. Based on the model, latent trial-wise variables were extracted: Bayesian surprise as an indicator to what extent a new outcome should invoke belief updating, and belief entropy as an indicator of (subjective) uncertainty.

The relationship between trial-wise pupil dilation and both belief uncertainty as well as Bayesian surprise was investigated, with a focus on the extent to which this relationship would be moderated by group and task conditions such as high/low risk and high/low volatility.

For method details, see sections 2.2.2, 2.3, 2.4 and *paper II*.

Results:

No differences were found between SZ and HC group regarding working memory capacity, the behavioral prediction task variables accuracy and switching, or the group parameters from the cognitive model. However, within the SZ group, a negative correlation between positive

symptoms and sensitivity to positive feedback in the cued task block suggested that when volatility was low, with changes of the probability conditions announced, participants with more severe delusions and hallucinations learned less appropriately from positive feedbacks (i.e. correct predictions) than those with a lower symptom load. This decreased sensitivity may reflect an increased subjective perception of positive feedback to be unreliable.

Despite the lack of behavioral group differences, patients with SZ showed on average more belief entropy, i.e. uncertainty, particularly within the volatile task block. Moreover, while pupil dilation to outcome presentation was positively associated with subjective trial-wise uncertainty, this relationship was significantly attenuated in the SZ group.

Conclusion:

As groups were matched on demographic and educational variables and did not differ regarding working memory capacity, the lack of group differences on some of the main prediction task variables may be due to the general neurocognitive fitness of the selected patient sample. Nevertheless, patients still showed an increased uncertainty in the volatile block, hinting in part at a potentially increased sensitivity to the environment's volatility, even though this did not translate into a significantly increased model-based volatility estimate. Aberrant processing of uncertainty in the SZ group is also implied by the decreased adaption of pupil size to trial-wise uncertainty. On trials where uncertainty is high, a given outcome should be perceived as highly salient as its integration into prior beliefs would drive learning and help reduce uncertainty. Thus, pupil dilation would be expected to increase as well, reflecting increased neural gain (Eldar et al., 2013). Individuals with SZ, however, seem to fail to differentiate between high and low salient, or 'informative', outcomes. This fits well with the aberrant salience account which postulates that a misbalance in attributing salience to more or less informative events is a core feature of schizophrenia (Heinz & Schlagenhauf, 2010; Kapur, 2003).

3.3 Paper III

Aims and background:

Paper III aimed to investigate the relationship between working memory capacity, recall accuracy, and objective as well as subjective measures of effort and motivation in the same core sample (excluding two participants) as described in *paper II*. Both working memory and motivation have been implied to play a role in decision-making under uncertainty (see section 1.3). Hence, general group differences on these measures are relevant to consider when interpreting the cognitive-behavioral results of other tasks. Furthermore, investigating to what extent both objective and subjective measures of invested effort converge can give insights regarding metacognitive ability in this sample.

Methods:

A visual, computerized version of the digit span task forward was administered to measure both general working memory capacity and trial-wise recall accuracy across different load conditions (min. 2 – max. 9 digits to recall). While digits were presented on screen (i.e. during the encoding process), pupil size was recorded as an index of objectively invested effort. Self-reported motivated effort, and perceived 'ease' regarding the task were assessed after completion of the digit span task with a newly compiled questionnaire.

For method details, see sections 2.2.1, 2.3, 2.4, and *paper III*.

Results:

Groups did not differ regarding general working memory capacity assessed as maximum digit span, i.e. maximum number of digits recalled in the correct order. Conversely, the SZ group demonstrated decreased recall accuracy on a trial-by-trial basis. While there was no group difference in self-reported motivated effort, pupil dilation was reduced in the SZ group across all load conditions, suggesting objectively decreased effort investment. Substantiating the interpretation of pupil size as a measure of effort investment, trial-wise pupil dilation was positively associated with trial-wise recall accuracy. Yet, a significant interaction with group indicated that this positive relationship was smaller in the SZ group. Across the whole sample, objectively measured effort in terms of pupil dilation was not related to subjective measures of effort as assessed by self-reports.

Conclusion:

The lack of group differences in working memory capacity was surprising, given well-replicated findings of working memory deficits in schizophrenia (Lee & Park, 2005) and the results of *paper I*. However, particularly for the digit span forward, findings have been quite heterogenous and inconsistent (Forbes, Carrick, McIntosh, & Lawrie, 2009), and performance sometimes seems to be spared (Barch, 2005). Given this similar general capacity, the trial-wise recall accuracy deficits found in the SZ group are likely to have been caused by reduced effort and allocation of attention rather than by a general lack of cognitive resources. In line with this interpretation, the pupillometric results also indicated reduced effort investment in patients with SZ. Still, in light of the significant interaction between pupil size and group on trial-wise accuracy it remains unclear to what extent this alone can explain the decreased recall performance. If interpreted as reflecting metacognitive ability, the lack of a direct link between objective and subjective measures of effort for the whole sample suggests impaired insight for all participants. This may in part be due to response biases on the questionnaire measure as well as the fact that, while pupil dilation as an objective measure was recorded on each trial, self-reports were only collected after the whole task was completed. Nevertheless, the findings highlight the importance to consider reduced effort as a factor when interpreting neuropsychological test results in individuals with schizophrenia at the same time as they suggest that self-reported motivation or effort might have to be interpreted with caution.

4 General discussion

4.1 Discussion of findings

One of the most striking overall findings across the papers summarized in this thesis is the spared performance of individuals with schizophrenia regarding many of the cognitive-behavioral variables. The patient sample that was assessed in both *paper II* and *paper III* did not differ significantly from the control group regarding their general working memory capacity. Similarly, no group differences were found for metamemory in *paper I* and self-reported (subjective) effort, as well as the relationship between subjective and objective effort, in *paper III*. Furthermore, in both *paper I* and *paper II*, decision-making under uncertainty was similar in patients and healthy controls. However, in the studies of all three papers, additional and more latent measures revealed underlying group differences in decision parameters and variables beyond the directly observable behavior. These results are discussed in detail in the following sections.

4.1.1 Working memory and meta-cognitive processes

4.1.1.1 Working memory performance

Working memory performance was assessed in all papers, using two different tasks. While individuals with schizophrenia showed a significantly smaller recall accuracy in the visual working memory task used in *paper I*, their working memory capacity assessed with the digit span forward in *paper II* and *paper III* (same sample and same assessment) was comparable to that of the control group. The finding of *paper I* fits well with previously established (visual) working memory deficits in schizophrenia (e.g., Freeman et al., 2014; Gold, Wilk, McMahon, Buchanan, & Luck, 2003; Horan, Braff, et al., 2008; Tek et al., 2002).

However, as outlined in section 3.3, findings for the digit span forward task as used in the sample that was investigated in *paper II* and *paper III* are inconsistent, with many showing preserved performance (Barch, 2005). Given the fact that patients assessed in *paper II* and *paper III* were to a large extent relatively well adapted outpatients, well matched with controls on central demographic variables, patients' overall neurocognitive fitness might have prevented impairments on a relatively simple working memory task such as the digit span forward. Interestingly, and despite the intact general working memory capacity, participants of the schizophrenia group demonstrated a reduced trial-wise digit recall accuracy in *paper III*. This effect was not moderated by task load (i.e. the number of digits to be recalled). Together, these findings suggest that decreased trial-wise performance might more likely be the result of momentary attentional fluctuations and/or fluctuations in invested effort, as opposed to diminished abilities per se. This was possibly, but not necessarily, driven by motivation (Engelmann et al., 2009; see also section 3.3). Similarly, it has been suggested that the working memory impairments commonly observed in patients with schizophrenia may result from attentional deficits and not a lack of cognitive storage resources (Gold et al., 2003). Furthermore, it has been acknowledged that impaired performance on neurocognitive tests may

be the consequence of a range of confounders alternative to or in addition to actual cognitive impairments, including poor motivation and momentary impairments, for example related to distraction by current symptoms (Moritz, Klein, et al., 2017).

4.1.1.2 Metacognitive processes

In *paper I*, metamemory was assessed in the visual domain in a more implicit manner than in many previous studies, using a capture area instead of verbal self-reports of confidence. Here, individuals with schizophrenia did not show the expected overconfidence in errors that was expected based on prior findings (e.g., Moritz & Woodward; Moritz, Woodward, & Rodriguez, 2006; Moritz et al., 2008). On the one hand, this absence of a significant group difference may have been caused by the fact that metamemory was measured implicitly as opposed to explicitly. In fact, implicit and explicit metacognitive processes may rely on separate cognitive systems (Shea et al., 2014) and might therefore be affected differently in schizophrenia. Indeed, it has been found that implicit self-monitoring, a form of metacognition (see section 1.2) seems to be intact in patients (Knoblich, Stottmeister, & Kircher, 2004). Accordingly, administration of a variant of the visual working memory task to a different patient sample revealed again no significant impairments regarding implicit metamemory (Hegelstad, Kreis, Tjelmeland, & Pfuhl, 2020). On the other hand, many of the previous studies on metamemory in schizophrenia have shown that the problem is not overconfidence per se, but that individuals with schizophrenia simply fail to adapt their confidence ratings in accordance with their performance, leading to overconfidence in errors and underconfidence in correct responses when compared to a healthy control group (e.g., Moritz, Woodward, & Chen, 2006; Moritz, Woodward, & Rodriguez, 2006; Moritz & Woodward, 2006). This so-called 'confidence gap' (Moritz, Woodward, & Rodriguez, 2006) could not be assessed with the metamemory measure in *paper I,* where recall performance was not categorical.

In *paper III*, metacognitive processes are partly captured in the task-related effort measures. As outlined in section 1.2, the amount of effort allocated to a task may be driven by metacognitive knowledge about task demands and one's own skills and abilities (Efklides, 2009). In line with this, some suggest that (effort-related) cognitive control is driven by uncertainty, which in turn may reflect metacognitive judgments (Mushtaq et al., 2011). Lastly, the definition of metacognitive regulation itself comprises processes of monitoring, adaptation and regulation, i.e. cognitive control processes (Fernandez-Duque et al., 2000). Hence, effort allocation might not only be driven by metacognitive knowledge but also be an inherent part of metacognitive regulation. Given previous reports of diminished effort investment in schizophrenia, patients were expected to show reduced effort allocation in the digit span task of *paper III*. Interestingly, their degrees of self-reported 'motivated effort' did not differ from that of the healthy control group. Conversely, pupil size, an implicit psychophysiological effort measure, showed the expected result, with smaller pupil dilation to the last digit of the encoding period in patients. This effect was not moderated by load (i.e. number of digits), indicating that patients did not reduce effort as a result of being overwhelmed with the task demands but instead invested less effort throughout. While a positive relationship between trial-wise pupil dilation and recall

accuracy seemed to substantiate the idea that pupil dilation indeed reflected invested effort, the fact that this relationship was decreased in the schizophrenia group challenged the interpretation of their diminished trial-wise performance to be a result of reduced effort and proposes the involvement of additional factors (see also section 4.2.3 on the interpretation of pupillometric measures). Notably, for the whole sample assessed in *paper III*, pupil dilation did not correspond to self-reported effort or ease (experienced task demands), raising important questions about the validity and comparability of explicit (subjective) and implicit (objective) measures of effort. However, this may in part reflect a weakness of the task paradigm, where self-reported effort was only measured once after task completion whereas objective effort (pupil dilation) was measured on a trial-by-trial basis. This decision was based on the concern that explicitly measuring invested effort on each trial may trigger self-reflection and –regulation, which might ultimately lead to changes in the measured construct itself. As such, results would be little comparable to the prior findings of decreased effort investment in schizophrenia.

4.1.2 Decision-making under uncertainty

4.1.2.1 Belief updating and choice switching: behavioral results

Assessing probabilistic decision-making via directly observable variables of task behavior such as belief updating and choice switching, the findings of neither *paper I* nor *paper II* pointed towards deficits in individuals with schizophrenia.

As outlined in section 3.1 and contrasting previous findings (Speechley et al., 2010), participants with schizophrenia did not show over-adjustment to disconfirmatory evidence or increased initial uncertainty in the beads task. This might in part be due to the implemented task paradigm, which varied from traditional graded-estimates versions of the beads task in that volatility, i.e. a probability for the bag of origin to change, was explicitly introduced. Interestingly, it has been suggested that even in standard versions of the task, participants may misinterpret the instructions and base their certainty estimates about the origin of the drawn beads based on whatever bead is currently represented (Balzan et al., 2012), constituting a similar effect as one that is introduced by the probability for the bags to change. Announcing volatility explicitly through task instructions may then have diminished potential group differences between those individuals that naturally tend to overestimate volatility (and/or assume they should only judge the currently presented beads, i.e. individuals with schizophrenia) and those that do not (healthy controls). Notably, even with standard, DTD versions of the beads task, findings of a JTC bias in schizophrenia cannot always be replicated. In fact, Rausch and colleagues (2014) found that patients scoring low on positive symptoms actually sampled more beads than controls.

To circumvent the issue of explicit volatility instruction and assess more directly to what extent manipulation of the degree of volatility in the task environment may affect patients and controls differently, *paper II* implemented a probabilistic reversal learning task. Here, increased choice switching might indicate an increased tendency to update prior beliefs, similar to

disconfirmatory belief updating in the beads task. Surprisingly, neither accuracy regarding the identification of the current majority stimulus nor choice switching differed significantly between patient and control group in either high or low volatile task conditions. This was again at odds with previous findings (Culbreth, Gold, et al., 2016; Deserno et al., 2020; Murray et al., 2008; Waltz et al., 2013), and in part possibly related to differences in task paradigms, regarding for example the choice of the risk conditions (i.e. probabilities for the 'left' and 'right' stimuli to appear) and the lack of an external monetary reward. In studies where a monetary reward is awarded as a function of participants' performance, group differences may emerge due to differences in sensitivity to and valuation of such reward (Chang et al., 2019; Culbreth, Westbrook, & Barch, 2016). However, studies of probabilistic and reversal learning in schizophrenia have produced inconsistent results, with some finding preserved performance in large subgroups of outpatients, for example (Reddy, Waltz, Green, Wynn, & Horan, 2016). Hence, given that the sample for *paper II* also consisted of a majority of outpatients, differences in the particular clinical and neurocognitive characteristics of the samples assessed may further explain the discrepancy with other findings (see also section 4.2.1). Nevertheless, if the comparative psychopathological and neurocognitive 'fitness' of the sample in *paper II* was indeed the reason for the relatively spared performance, this could still not explain why then performance in *paper I* was preserved as well, where the sample consisted of inpatients with likely more severe and acute symptoms. In order to investigate this more closely, measures of symptoms and other clinical details should have been added to the assessments of *paper I*. This constitutes an important limitation, particularly because the JTC-variables measured with the beads task are commonly related to delusions (Broome et al., 2007; Dudley et al., 2016; Falcone et al., 2015; Fine et al., 2007; Garety & Freeman, 2013; Garety et al., 1991; Huq et al., 1988); an association that could not be tested for here. The absence of group difference in the currently more acute and thus more severely affected patient sample of *paper I* may in part be related to the low power of the study. The patient sample size was very small ($n = 21$). This was improved in *paper II*, although one might argue that even a sample size of 30 is barely big enough, particularly when effect sizes might be small. In light of the relative fitness of this sample, effect sizes of increased switching or a heightened volatility parameter might indeed be smaller than those derived from studies that tested more acute patients. Another limitation that should particularly be mentioned for *paper I* is the lack of suiting control conditions for the beads task. As mentioned above, the explicit instruction about volatility may have diminished some potentially present and inherent group differences. Ideally, the task would have been compared to a version of the task where volatility was present but not instructed and to a version where volatility was completely absent. This would provide more insight into which task manipulations patients might be most sensitive to, and whether they show behavior that generalizes across different versions of the task.

4.1.2.2 Volatility estimation and choice uncertainty: modelling results

In addition to assessing the effect of volatility on directly observable behavior, in both *paper I* and *paper II* cognitive models were employed to derive participant- and group-specific estimates of volatility. In *paper I*, statistical tests conducted on this parameter did not reveal the

hypothesized overestimation of volatility in the patient group. Similar to the behavioral results, the explicit instruction about the actual size of volatility in the beads task may have suppressed the emergence of group differences by providing all participants with the same prior belief about volatility. Notably, in each group there were some participants with an estimated volatility parameter close to the instructed value (first cluster), and others where volatility was much larger (second cluster). Within the second cluster, volatility values were larger for individuals with schizophrenia, indicating at least in part a tendency to overestimate volatility in a subgroup of this sample. The bimodality observed across the whole sample may on the one hand reflect differences in processing modes employed by the participants (Freeman & Dale, 2013). For example, participants in the low volatility cluster may have employed a more model-based strategy, where decisions are based on a complex cognitive model of the task structure and associated state transition probabilities. In contrast, participants in the high volatility cluster may have engaged in a more model-free strategy, where decision are made in a more habitual kind of way, based on trial-and-error feedback (Daw, Niv, & Dayan, 2005). Such a model-free mode might evoke hypersensitivity to color changes in terms of belief changes which in turn could be captured by an increased volatility parameter. Interestingly, patients with schizophrenia have been found to engage less in model-based as opposed to model-free decision-making (Culbreth, Westbrook, Daw, Botvinick, & Barch, 2016). On the other hand, the bimodality may be explained by differences in understanding task instructions. Participants in the high volatility cluster may have mistakenly assumed that bag changes could occur with a 50% chance during each bead draw instead of a 50% chance per sequence. This may have resulted in more and stronger subjective probability changes in favor of the different bags, increasing the volatility estimate. Inclusion of control conditions where volatility is present but not instructed, and where volatility is completely absent, would have been useful to elucidate these questions. The study design of *paper II* addressed this issue by removing the potential effect of (mis)understanding volatility-related task instructions and extending the assessment with contrasts between high and low volatile task conditions. Surprisingly, just like for the behavioral results, groups did not differ on the model-based latent parameter reflecting subjective volatility, again contradicting previous results (Schlagenhauf et al., 2014). Yet, within the volatile task condition, individuals with schizophrenia seemed to be on average more uncertain about the current task state on a given trial, which may reflect in part an increased sensitivity to volatility. The finding of decreased pupil size adaptation to uncertainty in patients was further in accordance with an aberrant uncertainty processing account. This is because pupil dilation to a given stimulus is thought to reflect processing resources allocated to this stimulus, indexing neural gain and learning (Eldar et al., 2013). On trials where uncertainty about the current task states is high, a presented outcome should be perceived as a highly informative teaching signal and thus receive greater attention. Since patients do not seem to differentiate between more and less informative outcomes, this might indicate abnormal neurochemical processing of cognitive-psychological uncertainty, which further might hinder long-term learning and uncertainty reduction.

Unlike *paper I*, *paper II* also assessed the relationship between decision parameters and psychopathological symptoms. Here, patients with more pronounced positive symptoms demonstrated a decreased sensitivity to positive feedback in the low volatility condition. This

suggests that they relied less on positive feedback when updating their state beliefs, even though changes of task states (i.e. probability conditions) were announced. These findings are in line with previous studies that revealed decreased sensitivity to positive feedback, as well as its association with positive symptom severity in schizophrenia (Reddy et al., 2016; Schlagenhauf et al., 2014). The absence of this correlation for the volatile task block may be explained by the fact that here, the perceived unreliability of positive feedback extended to all participants, given that the unreliability of feedback was generally higher due to the hidden changes of stimulus probabilities.

Again, the discrepancy of the remaining findings in comparison to other studies, particularly regarding the volatility parameters in *paper I* and *paper II*, may be related to the same limitations that have been outlined for the behavioral findings in the preceding section, including the particular task paradigms chosen, and the size (*paper I*) and relative fitness (*paper II*) of the patient samples.

4.1.3 The relationship between uncertainty-driven processes: metacognition, probabilistic reasoning, and the role of working memory

The question to what extent the different uncertainty-related processes investigated in the papers of this thesis overlap, was most directly addressed in *paper I,* where correlations across several measures were calculated. In addition, some general profiles can be inferred from the findings of *paper II* and *paper III*, which were based on the same core sample of patients and control participants. Across papers, careful conclusions can further be drawn about schizophrenia samples as a whole. Still, a crucial caveat here is certainly the heterogeneity of schizophrenia in general and within the samples of this thesis in particular, where *paper I* included a majority of inpatients, *paper II* and *III* a majority of outpatients, and where *paper I* did not include any symptoms assessments, whilst *paper II* and *III* provided positive and negative symptom scores.

Across the whole sample of *paper I*, working memory accuracy was lower in participants that showed larger disconfirmatory belief updating scores. Given that disconfirmatory belief updating is considered a graded-estimates-version of the JTC bias, this seems to be in line with previous reports about an association between working memory and the tendency to jump to conclusions (Freeman et al., 2014; Takeda et al., 2018), illustrating the potential role of general cognitive ability. Participants with lower working memory capacity might for example struggle with remembering the history of previously drawn beads and/or certain details of the task instructions. The memory measure also correlated with the metamemory measure, though this relationship might be due to the fact that the measures were heavily correlated by design. Across tasks, metamemory was negatively associated with subjective volatility, suggesting that participants with lower metacognitive ability also tended to overestimate the objectively low volatility of the beads task environment. This seems to fit well with the idea that general mechanisms related to processing and representation of uncertainty may underlie different

symptoms and behaviors in schizophrenia, including aberrant probabilistic decision-making and impaired metacognitive abilities. Surprisingly, patients did not differ from controls on any of these measures in *paper I*. Nevertheless, the correlation substantiates the conceptual overlap and it is likely that in a study with a larger and more appropriately assessed patient group, group comparisons on both measures would have rendered significant and more consistent differences.

When characterizing the sample assessed in *paper II* and *paper III* along the reported findings, the first important point to note is the relative neurocognitive fitness of the schizophrenia group. Working memory capacity in this sample was comparable to healthy controls and probabilistic learning as assessed by summarized behavioral outcome measures seemed to be intact as well. As discussed in the preceding sections, this may in part be due to the fact that the sample contained a large proportion of relatively stable outpatients. However, in both studies the more fine-grained analyses of trial-wise and latent decision variables revealed the presence of subtle differences. In *paper II*, this concerned trial-wise belief uncertainty and in *paper III* recall accuracy per trial. Furthermore, in both studies pupil size seemed to track relevant information less reliably. Since working memory capacity was not significantly impaired, such subtle deficits seem to be independent of storage resources. Instead, they may arise from fluctuations of attention devoted to the information presented on a given trial. Notably, while pupil size served as an indicator of uncertainty processing in *paper II* and of invested effort in *paper III*, the failure to adapt pupil size according to the information at hand in both tasks may reflect similar disturbances in the norepinephrinergic system. Generally put, NE moderates the 'signal-to-noise' ratio, determining the attention devoted to new incoming information and the regulation of neural gain for learning (Eldar et al., 2013). Within the adaptive gain account, a high baseline activity level within the LC-NE system is for example thought to promote exploratory behavior, which is associated with increased distractibility and attentional switching regarding the task at hand (Aston-Jones & Cohen, 2005). This processing mode has been found to evoke larger baseline pupil size (Jepma & Nieuwenhuis, 2011), and consequently, smaller (baseline corrected) phasic pupil responses (Gilzenrat et al., 2010). As such, the reduced adaptation of phasic pupil responses to the informational value of the outcome at hand (belief uncertainty in *paper II* and digits to memorize in *paper III*) may be the consequence of an increased proneness to exploratory behavior. The attentional switching associated with this mode may then explain the decreased trial-wise recall accuracy in *paper III* and the failure to reduce uncertainty through appropriate integration of trial-wise outcomes in *paper II*. A similar interpretation can be derived in the context of the unexpected uncertainty framework, where unexpected uncertainty is thought to be encoded by tonic levels of NE (Yu & Dayan, 2005). Chronically increased unexpected uncertainty, caused by the perception of the world as inherently volatile and unstable, inevitable stimulates exploratory behavior. Decreased adaptation of pupil size to the uncertainty-dependent informational salience of a given outcome may then reflect the fact that with elevated levels of unexpected uncertainty and associated tonic NE, all kinds of outcomes are perceived as equally relevant and 'worthy of exploring'. These explanations fit well with the idea that tonic NE levels might be increased in some types of schizophrenia (Fitzgerald, 2014).

As outlined in the preceding section and section 1.2, effort investment can conceptually be linked to metacognitive processes in that metacognitive knowledge indicates the need to invest effort and metacognitive regulation reflects the implementation of effort and cognitive control processes. Accordingly, effort investment might be corrupted when metacognitive ability is impaired. Comparing uncertainty-related processes of the metacognitive and the probabilistic reasoning domain across papers then demonstrates that the same sample of schizophrenia patients that showed increased belief uncertainty in *paper II* also demonstrated decreased, possibly metacognitively moderated, cognitive control and effort investment in *paper III*. This seems to substantiate a general 'uncertainty account' of schizophrenia, where aberrant processing and representation of uncertainty might occur on several levels in the cognitive system, translating into abnormal metacognitive processes and affecting belief formation.

Interestingly, the overlap between impaired metacognitive processes and reasoning in the probabilistic prediction task of *paper II* is also reflected in the negative correlation between positive symptoms and sensitivity to positive feedback found for the low volatile task condition. In line with previous findings (Reddy et al., 2016; Schlagenhauf et al., 2014) this indicates an increased uncertainty about the reliability of positive feedback (i.e. correct predictions) in patients with higher severity of delusions and hallucinations. Such misbeliefs about one's own correct performance may be similar to the findings of increased underconfidence in objectively accurate responses as demonstrated in other cognitive tasks (Moritz, Woodward, Cuttler, Whitman, & Watson, 2004; Moritz, Woodward, & Rodriguez, 2006; Moritz et al., 2005). Moreover, this association between uncertainty-related feedback processing and positive symptoms fits with the idea that these symptoms in and of themselves may be the result of aberrant uncertainty processing as proposed by the Bayesian brain account of schizophrenia (see section 1.1). Nevertheless, this interpretation is weakened by the fact that correlations with symptoms were absent for the other uncertainty-related parameters, including sensitivity to negative feedback and subjective volatility. It is conceivable that some of these associations are more subtle and might therefore only be revealed in a sample containing a wide range of different symptom scores. Given that the sample was reasonably stable and well medicated, the lack of high symptom severity cases may explain the absence of more significant correlations between symptoms and cognitive model parameters.

4.2 Methodological considerations and limitations

Many of the limitations of this thesis have already been mentioned in the previous sections. In the following sections, those limitations and according methodological considerations are summarized in a more systematic manner.

4.2.1 Selected samples

First and foremost it should be noted that recruiting a sufficient amount of participants for patient studies is always challenging, particularly when inclusion criteria are strict, when task

paradigms are complex and require in-person assessments, and when the studies have to be completed within a given time frame. This often leads to rather small sample sizes in clinical studies. A more lenient attitude towards inclusion criteria might help to increase statistical power, but might hinder interpretation and generalization of the findings. For all studies reported in *paper I – III*, the patient group consisted of individuals with a diagnosis from the schizophrenia spectrum and was not further divided into different disorders of the spectrum. Importantly, in the newest version of the DSM (DSM-V, American Psychiatric Association, 2013) subtypes of schizophrenia have in fact been eliminated. However, the division of the schizophrenia spectrum into categories such as schizophrenia, schizoaffective disorder and delusional disorder, persists. In this thesis, patients with schizophrenia were included in the same group as patients with a schizoaffective disorder. This helped to increase the amount of participants in favor of a higher statistical power but prevents the generalization of the results to specific subgroups and may further have overshadowed effects that might have been present only for very particular diagnoses. It is noteworthy that many of the hypotheses regarding aberrant uncertainty processing in schizophrenia have been linked to psychosis in particular and are thus not expected to differ substantially between schizoaffective and schizophrenic disorders. Adding symptom measures such as the PANSS as done in *paper II* and *paper III* can then assist in finding trends even across the different subgroups of patients. The lack of such symptom measures for *paper I* constitutes one of the major limitations of that study. It shall be noted that the distinction between schizoaffective and schizophrenic subtypes is a controversial topic. Many have acknowledged the psychopathological heterogeneity of schizophrenia (Buchanan & Carpenter, 1994; Horan, Blanchard, Clark, & Green, 2008; Joyce & Roiser, 2007; Lindenmayer et al., 1995; Picardi et al., 2012), and multiple studies point towards a lack of reliability in distinguishing schizoaffective disorder from schizophrenia and other psychotic or affective disorders (Jäger, Haack, Becker, & Frasch, 2011; Kempf, Hussain, & Potash, 2005; Lake & Hurwitz, 2007; Maier, 2006). The DSM-V has been acknowledged for improving this reliability (Malaspina et al., 2013). Yet, many patients that have been diagnosed prior to its publication (including many of those involved in the studies of this thesis) may still carry a potentially inappropriate label of 'schizoaffective' disorder. Also with regards to neurocognitive impairments there is no consistent evidence as to whether individuals with schizoaffective disorder differ from those with schizophrenia. While some have reported significant differences in the cognitive profile of both disorders (Heinrichs, Ammari, McDermid Vaz, & Miles, 2008; Hill et al., 2013; Torniainen et al., 2012), others found no evidence for this (Fiszdon, Richardson, Greig, & Bell, 2007; Townsend, Malla, & Norman, 2001). Consequently, many have questioned whether schizoaffective disorder should be considered a separate entity to begin with, or rather represents either a subtype of schizophrenia, of affective and bipolar disorders, or a stage on a continuum between the two (Abrams, Rojas, & Arciniegas, 2008; Madre et al., 2016).

Despite the generous inclusion criteria, sample sizes of the papers presented here remained rather small, particularly for *paper I*. As acknowledged in the previous sections, this may have hindered the detection of effects that are actually present in the real population. Another limitation, which is quite common for clinical studies, is the inclusion of medicated patients. Many of the disorder specific impairments are thought to be associated with neurochemical

imbalances inherent to the disorder. Dopaminergic dysfunction is for example thought to underlie both reinforcement learning deficits (Deserno et al., 2013; Frank, 2008) and aberrant belief updating in probabilistic reasoning tasks like the beads task (Evans et al., 2015; Speechley et al., 2010). It is then possible that such deficits are less prominent in patients that are medicated with antipsychotics, which usually have a strong dopamine antagonistic effect. This may explain the patients' intact performance on many of the variables of interest across all papers of this thesis. Some studies have tried to circumvent this issue by investigating samples of unmedicated and/or first episode patients instead, in order to obtain assessments untainted of medication effects (e.g., Falcone et al., 2015; Schlagenhauf et al., 2014). However, this might lead to a selection bias where the resulting samples consist of less severe cases and/or cases where the diagnostic process has not been thoroughly completed.

When conducting studies with medicated participants one may want to control for the confounding effect of medication on measured outcome variables through inclusion of covariates (see *paper II* and *paper III*), but this approach is less straight forward than it seems. Again, heterogeneity is a potential issue, with patients receiving different types and dosages of medication with varying neurochemical effects. Even though there are ways of calculating general medication load to render different medication types comparable (e.g. Chlorpromazine equivalents; Atkins, Burgess, Bottomley, & Riccio, 1997; or percentage of the clinically recommended maximum dosage; Kane, Leucht, Carpenter, & Docherty, 2003), these approaches do not account for the variety of differently affected neurotransmitter systems and their interactions. Besides, it remains questionable whether medication load is directly comparable across different individuals, even when medication type does not differ. The amount of medication needed may be inherently related to the degree with which neurochemical systems are imbalanced to begin with. A high correlation between medication and an observed behavioral variable may then for example reflect a more indirect relationship that is in fact driven by the underlying neurochemical disturbances and not the medication per se. Last but not least, even in well medicated subjects, clinical symptoms remain (see for example *paper II*). Hence, the question of how to best control for medication effects in clinical research remains a difficult one to resolve.

An ideal but complex approach might be the implementation of longitudinal studies, following patients from an untreated first episode stage to later, more chronic and then often medically treated stages. Such an approach would also be tremendously valuable in order to characterize cognitive and behavioral profiles at different developmental stages of the disorder. As stated in section 1.1, many of the uncertainty-based hypotheses that were formulated within a Bayesian brain account of schizophrenia postulate a shift in processing style from early to later stages. Assessing uncertainty-based processing styles in the mixed samples of *paper I and paper II*, both of which contained large proportions of chronic patients, might have limited the detectability of certain effects if they were to be more pronounced in early stages. Furthermore, longitudinal studies are vital when trying to determine the extent with which any of the cognitive-behavioral measures are relevant for the prediction of clinical outcomes (Deserno et al., 2016).

4.2.2 Chosen task paradigms

Another potential limitation of any cognitive-behavioral study concerns the task paradigms chosen. On the one hand, the development of new tasks and/or the adaptation of established tasks can facilitate a more detailed investigation of certain cognitive or decision-making components whilst removing factors that may have biased previous findings. On the other hand, continuous adaptation of task designs complicates the comparison of findings across studies.

For example, the beads task of *paper I* was adapted to introduce volatility as a potential mediating effect on belief overadjustment in schizophrenia. This created a new version of the task that differed on more than one factor from previous versions, rendering comparisons to prior findings problematic. As mentioned in section 4.1.2, it would have been beneficial to include additional control tasks, such as a version where volatility was present but not instructed to the participants, and a classic graded-estimates version of the beads task. However, every additional task may increase the strain perceived by the participants, which might be particularly detrimental in clinical studies, as patients might be more susceptible to strain (see e.g. *paper III*) and more easily exhausted. Dividing such studies into separate measurement sessions may alleviate the strain effect but might make recruitment more difficult and introduce additional confounders, as symptom severity, medication, and many other intra-individual factors may vary between the different sessions.

The prediction task of *paper II* was also newly constructed and designed to investigate probabilistic and reversal learning in more and less volatile and risky environments, without the potential confounder of monetary reward evaluation. This approach was chosen because monetary reward evaluation might be reduced in schizophrenic populations (Chang et al., 2019; Culbreth, Westbrook, & Barch, 2016) and thus could be one explanation for why patients perform worse on many of the 'classical' reinforcement and reversal learning tasks where monetary rewards are received based on performance (Culbreth, Gold, et al., 2016; Deserno et al., 2020; Waltz et al., 2013). One might wonder to what extent the lack of a monetary reward affected participants' motivation to 'do well' on the task at hand. Notably, there is evidence that responses to both external and 'internal' rewards such as good performance are generally comparable. Brain areas associated with the processing of reward, such as the ventral striatum, respond significantly stronger to feedback of correct performance than to feedback of incorrect performance (Tricomi & Fiez, 2008; Ullsperger & von Cramon, 2003). Furthermore, this relationship exists even when performance feedback is not provided, indicating a responsiveness to internal performance-based 'reward' processing (Han, Huettel, Raposo, Adcock, & Dobbins, 2010; Satterthwaite et al., 2012; Wolf et al., 2011). In line with this, average performance was above chance level in all groups across all conditions of the task, indicating that motivation to perform well was present even in the absence of a monetary reward (see *paper II*). Nevertheless, comparison of the findings to other reversal learning studies remains difficult, as the tasks used in those studies commonly only include one risk condition, often based on action-outcome associations of 80:20 and 20:80 (Culbreth, Gold, et al., 2016; Deserno et al., 2020; Waltz et al., 2013). Evidently, in those paradigms reversals are much easier to detect. In contrast, reversals in the prediction task of *paper II* might have been harder to detect when changes went from an 85:15 to a 60:40 condition, and even harder when a 60:40

condition was followed by a 40:60 condition, for example. This may have diminished potential subtle differences between individuals with schizophrenia and healthy controls. Likewise, the risk conditions themselves may have been too easy (85:15/15:85) or too hard (60:40/40:60) to cause significant group differences in terms of behavioral performance regarding accuracy or choice switching.

Unlike the newly developed tasks used *paper I* and *paper II*, the digit span task employed in *paper III* was designed to simulate the original digit span task from the WAIS-IV (Wechsler, 2008) as closely as possible, with only two trials per load condition. While this might assist the comparison of the findings with previously reported ones, it was not ideal regarding the analysis of the corresponding pupil data. Here, it would have been beneficial to obtain more than two measures per load condition to prevent data loss due to blinks in crucial time windows and to extract a more robust pupil response average less susceptible to noise.

4.2.3 Pupillometry

It might seem striking that similar pupillometric measures were employed to track conceptually different processes, such as uncertainty processing in *paper II* and effort-investment in *paper III*. As summarized in sections 1.4 and 1.5.2, both pupil size and related LC-NE activity have been linked to various cognitive processes. This is likely due to the LC-NE system's crucial role for the moderation of arousal (Samuels & Szabadi, 2008). In fact, a recent review concluded that "anything that somehow activates the mind […] also causes the pupil to dilate" (Mathot, 2018, p. 12). In light of this, attempts of linking phasic pupil responses to particular cognitive processes might seem "doomed to fail" (Mathot, 2018, p. 12). This is a legitimate concern that should be kept in mind when interpreting the results of *paper II* and *paper III*. However, even if arousal is the ultimate driving force behind pupil size changes in a cognitive effort task such as the one in *paper III*, an intimate link between this arousal and the effort invested is still very likely. Arousal might for example indicate the need to invest more effort. Accordingly, tonic arousal changes during attentional tasks have been linked to perceived mental effort (Howells, Stein, & Russell, 2010). Moreover, multiple studies have provided evidence for a link between pupil dilation and effort investment (van der Wel & van Steenbergen, 2018) and given that both arousal and effort reflect activation of the mind and affect pupil size, their effects seem to be comparable (Mathot, 2018). Similarly, pupil responses to uncertainty in a probabilistic reasoning or decision making task such as the one in *paper II*, may ultimately be linked to levels of arousal (Alamia, VanRullen, Pasqualotto, Mouraux, & Zenon, 2019). On the one hand, uncertainty has been found to affect arousal (Ramsøy, Friis-Olivarius, Jacobsen, Jensen, & Skov, 2012; Urai, Braun, & Donner, 2017). On the other, arousal has been found to affect uncertainty processing and decision-making in uncertain environments (Allen et al., 2016; FeldmanHall, Glimcher, Baker, & Phelps, 2016).

With this in mind, patients' reduced pupil responses to the different task variables in both studies might in fact reflect general differences in arousal. As described in section 4.1.3, increased baseline arousal, accompanied by high levels of tonic NE, might for example render

phasic pupil dilations less responsive and limit their potential to track online information, such as uncertainty in *paper II* and digits presented in *paper III*. In both tasks, this may contribute to a reduced differentiation between more and less relevant signals, i.e. an abnormal 'signal-to-noise' ratio (Aston-Jones & Cohen, 2005), affecting attention devoted to the task at hand, which could explain deficient recall performance on a trial-by-trial basis in *paper III*, and aberrant processing of informational salience associated with a given outcome in *paper II*. Indeed, schizophrenia has in part been associated with increased tonic levels of NE (Fitzgerald, 2014) and states of hyperarousal (Depue & Fowles, 1973; Kornetsky & Mirsky, 1965; Nuechterlein & Dawson, 1984; Yamamoto & Hornykiewicz, 2004).

Nevertheless, given the multiple neurotransmitter systems that are assumed to be implicated in schizophrenia (see section 1.4), as well as the additional effects of psychoactive medication on these systems, one may wonder to what extent pupil responses in patients are comparable to those of healthy controls. Here, particularly the anticholinergic effects inherently associated with some of the common antipsychotics (Stahl & Stahl, 2013) or caused by additionally administered anticholinergic agents to treat extrapyramidal side effects (Desmarais, Beauclair, & Margolese, 2012; Ogino, Miyamoto, Miyake, & Yamaguchi, 2014) have to be considered, as they possibly affect pupil size (Naicker, Anoopkumar-Dukie, Grant, Neumann, & Kavanagh, 2016). To control for this, the relationship between baseline pupil size measures and anticholinergic load was therefore investigated in *paper II* and *paper III*. However, calculations were based on small sub-samples and the converted anticholinergic load measure can only be considered a rough approximation. Furthermore, complex interactions between the different neurochemical systems involved in schizophrenia are likely (see e.g., Carlsson et al., 2001), which limits the extent to which a final conclusion can be drawn regarding the hypothesis of norepinephrinergic dysfunction in schizophrenia.

4.2.4 The validity of cognitive models

Last but not least it shall be mentioned that cognitive models, such as the ones employed in *paper I* and *paper II*, are only approximations and are constrained by the researcher's assumptions of how a task is solved. Notably, various studies have shown that the same cognitive model does not always provide a comparatively appropriate fit to all of the patients' behavior (see e.g., Schlagenhauf et al., 2014, where behavior of a subgroup of patients was appropriately fitted with the HMM, whereas for others a simple Rescorla-Wagner reinforcement learning model provided the best fit). In cognitive modelling studies it remains therefore essential to consider more than just one model and compare their performance against each other. This can help to identify what strategies were most likely used by which subgroups of the investigated samples. Nevertheless, this does not guarantee that the winning model was indeed the best available approximation to the observed behavior. It is important to weight the results within a wider context, informed not only by cognitive-psychological theories, but also assumptions regarding the neurophysiological systems that constitute the mechanistic basis for the cognitive processes of interest. Ongoing processes of such internal and external model evaluation will help to improve cognitive models in the long run.

5 Conclusions and future directions

The central idea of this thesis proposed that various cognitive-behavioral biases and deficits observed in schizophrenia may be rooted in overarching abnormalities regarding the representation of uncertainty and processing of uncertain information. As suggested within the Bayesian brain account of the disorder, such uncertainty-related aberrancies may be the very same mechanisms that underlie psychopathological symptoms such as delusions and hallucinations. Extending the behavioral assessments of metacognitive processes and probabilistic decision-making with pupillometric measures and cognitive models allowed for an investigation of the latent uncertainty-related processes and decision parameters behind directly observable task performance. Across the different papers, many of the main outcome variables reflecting decision-making and belief updating seemed to be unimpaired in patients, whereas differences emerged on pupillometric measures and model-based assessments of uncertainty representation. Those differences may have been too subtle to translate into performance deficits; possibly related to the relative psychopathological and neurocognitive 'fitness' of the patient samples, and the low power across studies due to sample size limitations. However, the relevance of findings of spared cognitive performance in schizophrenia should not be underestimated. In fact, cognitive deficits may be less severe than often assumed, and impaired performance may in part result from secondary factors such as motivation or worry regarding the outcome of the assessment (Moritz, Klein, et al., 2017). Furthermore, this might reflect the heterogeneity among individuals with disorders from the schizophrenia spectrum, suggesting that certain subtypes are neurocognitively less affected than others, and/or that the degree of impairment fluctuates with developmental stages of the disorder.

Nevertheless, the results illustrate the potential of psychophysiological and cognitive modelling methods for the investigation of uncertainty-related processes in schizophrenia and invite for future research on larger samples with a wider range of current symptomatology. Here, particularly longitudinal studies are of interest, where the same sample of patients could be tested at different developmental stages of the disorder, characterized by fluctuating symptom severity as well as varying medication types and dosages. Relating cognitive-behavioral and psychophysiological variables as measured in the lab to functional outcomes, symptoms and symptom development over longer time scales would be particularly relevant for testing the clinical significance of such lab-based assessments. In addition, the effect of metacognitive training (Moritz & Woodward, 2007) on the uncertainty-related variables and parameters should be explored, as this intervention aims at altering the way patients process and represent uncertain information. To improve the reliability and validity of the cognitive-behavioral assessments, tasks could further be broken down into variants assessing more distinct cognitive sub-components, limiting the amount of potential confounders. This could guide conclusions about the cognitive-mechanistic origin of observed performance deficits. While assessing the particular effects of neurochemical transmission in schizophrenia remains challenging, it would be intriguing to study the effects of medically modulated NE increase and decrease in terms of its relevance for uncertainty-related processes in this disorder. In patients and healthy controls, one could test to what extent such modulation may change sensitivity to and processing of different kinds of uncertainty, as well as cognitive control and metacognitive performance.

References

Abrams, D. J., Rojas, D. C., & Arciniegas, D. B. (2008). Is schizoaffective disorder a distinct categorical diagnosis? A critical review of the literature. *Neuropsychiatr Dis Treat, 4*(6), 1089-1109. doi:10.2147/ndt.s4120

Adams, R. A., Napier, G., Roiser, J. P., Mathys, C., & Gilleen, J. (2018). Attractor-like dynamics in belief updating in schizophrenia. *Journal of Neuroscience, 38*(44), 9471-9485.

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Front Psychiatry, 4*, 47. doi:10.3389/fpsyt.2013.00047

Ahn, W. Y., Haines, N., & Zhang, L. (2017). Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. *Comput Psychiatr, 1*, 24-57. doi:10.1162/CPSY_a_00002

Alamia, A., VanRullen, R., Pasqualotto, E., Mouraux, A., & Zenon, A. (2019). Pupil-linked arousal responds to unconscious surprisal. *The Journal of Neuroscience*, 3010-3018. doi:10.1523/jneurosci.3010-18.2019

Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *Elife, 5*. doi:10.7554/eLife.18103

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5* (5 ed.). Washington DC: American Psychiatric Association.

Andreou, C., Moritz, S., Veith, K., Veckenstedt, R., & Naber, D. (2014). Dopaminergic modulation of probabilistic reasoning and overconfidence in errors: A double-blind study. *Schizophrenia Bulletin, 40*, 558-565.

Andreou, C., Schneider, B. C., Balzan, R., Luedecke, D., Roesch-Ely, D., & Moritz, S. (2015). Neurocognitive deficits are relevant for the jumping-to-conclusions bias, but not for delusions: A longitudinal study. *Schizophrenia Research: Cognition, 2*(1), 8-11. doi:https://doi.org/10.1016/j.scog.2015.02.001

Anticevic, A., Krystal, J. H., & Murray, J. D. (2018). Meeting Emerging Challenges and Opportunities in Psychiatry Through Computational Neuroscience. In A. Anticevic & J. D. Murray (Eds.), *Computational Psychiatry* (pp. xiii-xxxi): Academic Press.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu Rev Neurosci, 28*, 403-450. doi:10.1146/annurev.neuro.28.061604.135709

Atkins, M., Burgess, A., Bottomley, C., & Riccio, M. (1997). Chlorpromazine equivalents: a consensus of opinion for both clinical and research applications. *Psychiatric Bulletin, 21*(4), 224-226.

Balzan, Delfabbro, Galletly, & Woodward. (2012). Over-adjustment or miscomprehension? A re-examination of the jumping to conclusions bias. *Australian and New Zealand Journal of Psychiatry, 46*, 532-540.

Balzan, R. P., Ephraums, R., Delfabbro, P., & Andreou, C. (2017). Beads task vs. box task: The specificity of the jumping to conclusions bias. *Journal of Behavior Therapy and Experimental Psychiatry, 56*, 42-50. doi:https://doi.org/10.1016/j.jbtep.2016.07.017

Barch, D. M. (2005). The Cognitive Neuroscience of Schizophrenia. *Annual Review of Clinical Psychology, 1*(1), 321-353. doi:10.1146/annurev.clinpsy.1.102803.143959

Barch, D. M., Treadway, M. T., & Schoen, N. (2014). Effort, anhedonia, and function in schizophrenia: Reduced effort allocation predicts amotivation and functional impairment. *Journal of Abnormal Psychology, 123*(2), 387-397. doi:10.1037/a0036299

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In *Handbook of psychophysiology, 2nd ed.* (pp. 142-162). New York, NY, US: Cambridge University Press.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10*, 1214. doi:10.1038/nn1954

Bergé, D., Pretus, C., Guell, X., Pous, A., Arcos, A., Pérez, V., & Vilarroya, O. (2018). Reduced willingness to invest effort in schizophrenia with high negative symptoms regardless of

reward stimulus presentation and reward value. *Comprehensive Psychiatry, 87*, 153-160. doi:https://doi.org/10.1016/j.comppsych.2018.10.010

Broome, M. R., Johns, L. C., Valli, I., Woolley, J. B., Tabraham, P., Brett, C., . . . McGuire, P. K. (2007). Delusion formation and reasoning biases in those at clinical high risk for psychosis. *British Journal of Psychiatry, 191*(S51), s38-s42. doi:10.1192/bjp.191.51.s38

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci, 18*(4), 590-596. doi:10.1038/nn.3961

Broyd, A., Balzan, R. P., Woodward, T. S., & Allen, P. (2017). Dopamine, cognitive biases and assessment of certainty: A neurocognitive model of delusions. *Clinical Psychology Review, 54*, 96-106. doi:https://doi.org/10.1016/j.cpr.2017.04.006

Buchanan, R. W., & Carpenter, W. T. (1994). Domains of psychopathology: An approach to the reduction of heterogeneity in schizophrenia. *Journal of Nervous and Mental Disease, 182*(4), 193-204. doi:10.1097/00005053-199404000-00001

Buck, K. D., Warman, D. M., Huddy, V., & Lysaker, P. H. (2012). The Relationship of Metacognition with Jumping to Conclusions among Persons with Schizophrenia Spectrum Disorders. *Psychopathology, 45*(5), 271-275. doi:10.1159/000330892

Carlsson, A., Waters, N., Holm-Waters, S., Tedroff, J., Nilsson, M., & Carlsson, M. L. (2001). Interactions between monoamines, glutamate, and GABA in schizophrenia: new evidence. *Annual review of pharmacology and toxicology, 41*(1), 237-260.

Ceaser, A. E., Goldberg, T. E., Egan, M. F., McMahon, R. P., Weinberger, D. R., & Gold, J. M. (2008). Set-Shifting Ability and Schizophrenia: A Marker of Clinical Illness or an Intermediate Phenotype? *Biological Psychiatry, 64*(9), 782-788. doi:https://doi.org/10.1016/j.biopsych.2008.05.009

Chang, W. C., Westbrook, A., Strauss, G. P., Chu, A. O. K., Chong, C. S. Y., Siu, C. M. W., . . . Chen, E. Y. H. (2019). Abnormal cognitive effort allocation and its association with amotivation in first-episode psychosis. *Psychological Medicine*, 1-11. doi:10.1017/S0033291719002769

Charlson, F. J., Ferrari, A. J., Santomauro, D. F., Diminic, S., Stockings, E., Scott, J. G., . . . Whiteford, H. A. (2018). Global Epidemiology and Burden of Schizophrenia: Findings From the Global Burden of Disease Study 2016. *Schizophrenia Bulletin, 44*(6), 1195-1203. doi:10.1093/schbul/sby058

Cheng, G. L. F., Tang, J. C. Y., Li, F. W. S., Lau, E. Y. Y., & Lee, T. M. C. (2012). Schizophrenia and risk-taking: Impaired reward but preserved punishment processing. *Schizophrenia Research, 136*(1), 122-127. doi:https://doi.org/10.1016/j.schres.2012.01.002

Cicero, D. C., Martin, E. A., Becker, T. M., & Kerns, J. G. (2014). Reinforcement learning deficits in people with schizophrenia persist after extended trials. *Psychiatry Res, 220*(3), 760-764. doi:10.1016/j.psychres.2014.08.013

Clayson, P. E., Wynn, J. K., Infantolino, Z. P., Hajcak, G., Green, M. F., & Horan, W. P. (2019). Reward processing in certain versus uncertain contexts in schizophrenia: An event-related potential (ERP) study. *Journal of Abnormal Psychology, 128*(8), 867-880. doi:10.1037/abn0000469

Cole, D. M., Diaconescu, A. O., Pfeiffer, U. J., Brodersen, K. H., Mathys, C. D., Julkowski, D., . . . Stephan, K. E. (2020). Atypical processing of uncertainty in individuals at risk for psychosis. *NeuroImage: Clinical, 26*, 102239. doi:https://doi.org/10.1016/j.nicl.2020.102239

Collins, A. G. E., Brown, J. K., Gold, J. M., Waltz, J. A., & Frank, M. J. (2014). Working memory contributions to reinforcement learning impairments in schizophrenia. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 34*(41), 13747-13756. doi:10.1523/JNEUROSCI.0989-14.2014

Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience, 35*(7), 1024-1035. doi:10.1111/j.1460-9568.2011.07980.x

Corlett, P. R., & Fletcher, P. C. (2014). Computational psychiatry: a Rosetta Stone linking the brain to mental illness. *Lancet Psychiatry, 1*(5), 399-402. doi:10.1016/S2215-0366(14)70298-6

Corlett, P. R., Frith, C. D., & Fletcher, P. C. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology, 206*(4), 515-530. doi:10.1007/s00213-009-1561-0

Corlett, P. R., Honey, G. D., & Fletcher, P. C. (2007). From prediction error to psychosis: ketamine as a pharmacological model of delusions. Journal of Psychopharmacology, 21(3), 238–252. https://doi.org/10.1177/0269881107077716

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., & Powers III, A. R. (2019). Hallucinations and strong priors. *Trends in Cognitive Sciences, 23*(2), 114-127.

Culbreth, A. J., Gold, J. M., Cools, R., & Barch, D. M. (2016). Impaired Activation in Cognitive Control Regions Predicts Reversal Learning in Schizophrenia. *Schizophrenia Bulletin, 42*(2), 484-493. doi:10.1093/schbul/sbv075

Culbreth, A. J., Westbrook, A., & Barch, D. M. (2016). Negative symptoms are associated with an increased subjective cost of cognitive effort. *Journal of Abnormal Psychology, 125*(4), 528-536. doi:10.1037/abn0000153

Culbreth, A. J., Westbrook, A., Daw, N. D., Botvinick, M., & Barch, D. M. (2016). Reduced model-based decision-making in schizophrenia. *Journal of Abnormal Psychology, 125*(6), 777-787. doi:10.1037/abn0000164

Culbreth, A. J., Westbrook, A., Xu, Z., Barch, D. M., & Waltz, J. A. (2016). Intact Ventral Striatal Prediction Error Signaling in Medicated Schizophrenia Patients. *Biological psychiatry. Cognitive neuroscience and neuroimaging, 1*(5), 474-483. doi:10.1016/j.bpsc.2016.07.007

Davis, K. L., Kahn, R. S., Ko, G., & Davidson, M. (1991). Dopamine in schizophrenia: a review and reconceptualization. *Am J Psychiatry, 148*(11), 1474-1486. doi:10.1176/ajp.148.11.1474

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*(12), 1704-1711.

den Ouden, Hanneke E. M., Daw, Nathaniel D., Fernandez, G., Elshout, Joris A., Rijpkema, M., Hoogman, M., . . . Cools, R. (2013). Dissociable Effects of Dopamine and Serotonin on Reversal Learning. *Neuron, 80*(4), 1090-1100. doi:https://doi.org/10.1016/j.neuron.2013.08.030

Depue, R. A., & Fowles, D. C. (1973). Electrodermal activity as an index of arousal in schizophrenics. *Psychological Bulletin, 79*(4), 233-238. doi:10.1037/h0034010

Deserno, L., Boehme, R., Heinz, A., & Schlagenhauf, F. (2013). Reinforcement Learning and Dopamine in Schizophrenia: Dimensions of Symptoms or Specific Features of a Disease Group? *Frontiers in Psychiatry, 4*(172). doi:10.3389/fpsyt.2013.00172

Deserno, L., Boehme, R., Mathys, C., Katthagen, T., Kaminski, J., Stephan, K. E., . . . Schlagenhauf, F. (2020). Volatility Estimates Increase Choice Switching and Relate to Prefrontal Activity in Schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 5*(2), 173-183. doi:https://doi.org/10.1016/j.bpsc.2019.10.007

Deserno, L., Heinz, A., & Schlagenhauf, F. (2017). Computational approaches to schizophrenia: A perspective on negative symptoms. *Schizophrenia Research, 186*, 46-54. doi:https://doi.org/10.1016/j.schres.2016.10.004

Deserno, L., Schlagenhauf, F., & Heinz, A. (2016). Striatal dopamine, reward, and decision making in schizophrenia. *Dialogues in clinical neuroscience, 18*(1), 77-89. doi:10.31887/DCNS.2016.18.1/ldeserno

Desmarais, J. E., Beauclair, L., & Margolese, H. C. (2012). Anticholinergics in the era of atypical antipsychotics: short-term or long-term treatment? *Journal of Psychopharmacology, 26*(9), 1167-1174.

Dowd, E. C., Frank, M. J., Collins, A., Gold, J. M., & Barch, D. M. (2016). Probabilistic Reinforcement Learning in Patients With Schizophrenia: Relationships to Anhedonia and Avolition. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1*(5), 460-473. doi:https://doi.org/10.1016/j.bpsc.2016.05.005

Dudley, R., Shaftoe, D., Cavanagh, K., Spencer, H., Ormrod, J., Turkington, D., & Freeston, M. (2011). 'Jumping to conclusions' in first‐episode psychosis. *Early Intervention in Psychiatry, 5*(1), 50-56.

Dudley, R., Taylor, P., Wickham, S., & Hutton, P. (2016). Psychosis, Delusions and the "Jumping to Conclusions" Reasoning Bias: A Systematic Review and Meta-analysis. *Schizophr Bull, 42*(3), 652-665. doi:10.1093/schbul/sbv150

Dudley, R. E. J., John, C. H., Young, A. W., & Over, D. E. (1997). Normal and abnormal reasoning in people with delusions. *British Journal of Clinical Psychology, 36*(2), 243-258. doi:10.1111/j.2044-8260.1997.tb01410.x

Eckstein, M. K., Guerra-Carrillo, B., Singley, A. T. M., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental cognitive neuroscience, 25*, 69-91.

Efklides, A. (2009). The role of metacognitive experiences in the learning process. *Psicothema, 21*(1), 76-82.

Eisenacher, S., & Zink, M. (2017). The Importance of Metamemory Functioning to the Pathogenesis of Psychosis. *Frontiers in Psychology, 8*, 304. doi:10.3389/fpsyg.2017.00304

Eldar, E., Cohen, J. D., & Niv, Y. (2013). The effects of neural gain on attention and learning. *Nat Neurosci, 16*(8), 1146-1153. doi:10.1038/nn.3428

Ernst, M., & Paulus, M. P. (2005). Neurobiology of Decision Making: A Selective Review from a Neurocognitive and Clinical Perspective. *Biological Psychiatry, 58*(8), 597-604. doi:https://doi.org/10.1016/j.biopsych.2005.06.004

Evans, S. L., Averbeck, B. B., & Furl, N. (2015). Jumping to conclusions in schizophrenia. *Neuropsychiatric disease and treatment, 11*, 1615-1624. doi:10.2147/NDT.S56870

Falcone, M. A., Murray, R. M., Wiffen, B. D. R., O'Connor, J. A., Russo, M., Kolliakou, A., . . . Jolley, S. (2015). Jumping to Conclusions, Neuropsychological Functioning, and Delusional Beliefs in First Episode Psychosis. *Schizophrenia Bulletin, 41*(2), 411-418. doi:10.1093/schbul/sbu104

FeldmanHall, O., Glimcher, P., Baker, A. L., & Phelps, E. A. (2016). Emotion and decision-making under uncertainty: Physiological arousal predicts increased gambling during ambiguity but not risk. *Journal of Experimental Psychology: General, 145*(10), 1255.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive Attention and Metacognitive Regulation. *Consciousness and Cognition, 9*(2), 288-307. doi:https://doi.org/10.1006/ccog.2000.0447

Fine, C., Gardner, M., Craigie, J., & Gold, I. (2007). Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. *Cognitive Neuropsychiatry, 12*, 46-77.

Fioravanti, M., Bianchi, V., & Cinti, M. E. (2012). Cognitive deficits in schizophrenia: an updated metanalysis of the scientific evidence. *BMC psychiatry*, *12*, 64. doi:https://doi.org/10.1186/1471-244X-12-64

Fisher, S. L., & Ford, J. K. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Personnel Psychology, 51*(2), 397-420.

Fiszdon, J. M., Richardson, R., Greig, T., & Bell, M. D. (2007). A comparison of basic and social cognition between schizophrenia and schizoaffective disorder. *Schizophrenia Research, 91*(1-3), 117-121. doi:10.1016/j.schres.2006.12.012

Fitzgerald, P. J. (2014). Is elevated norepinephrine an etiological factor in some cases of schizophrenia? *Psychiatry Research, 215*(3), 497-504. doi:https://doi.org/10.1016/j.psychres.2014.01.011

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci, 10*(1), 48-58. doi:10.1038/nrn2536

Forbes, N. F., Carrick, L. A., McIntosh, A. M., & Lawrie, S. M. (2009). Working memory in schizophrenia: a meta-analysis. *Psychological Medicine, 39*(6), 889-905. doi:http://dx.doi.org/10.1017/S0033291708004558

Frank, M. J. (2008). Schizophrenia: A Computational Reinforcement Learning Perspective. *Schizophrenia Bulletin, 34*(6), 1008-1011. doi:10.1093/schbul/sbn123

Freeman, D., Garety, P., Kuipers, E., Colbert, S., Jolley, S., Fowler, D., . . . Bebbington, P. (2006). Delusions and decision-making style: Use of the Need for Closure Scale. *Behaviour research and therapy, 44*(8), 1147-1158. doi:https://doi.org/10.1016/j.brat.2005.09.002

Freeman, D., Startup, H., Dunn, G., Cernis, E., Wingham, G., Pugh, K., . . . Kingdon, D. (2014). Understanding jumping to conclusions in patients with persecutory delusions: working

memory and intolerance of uncertainty. *Psychol Med, 44*(14), 3017-3024. doi:10.1017/S0033291714000592
Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods, 45*(1), 83-97.
Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris, 100*(1), 70-87. doi:https://doi.org/10.1016/j.jphysparis.2006.10.001
Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry, 1*(2), 148-158. doi:https://doi.org/10.1016/S2215-0366(14)70275-5
Fujino, J., Hirose, K., Tei, S., Kawada, R., Tsurumi, K., Matsukawa, N., . . . Takahashi, H. (2016). Ambiguity aversion in schizophrenia: An fMRI study of decision-making under risk and ambiguity. *Schizophrenia Research, 178*(1), 94-101. doi:https://doi.org/10.1016/j.schres.2016.09.006
Garety, P. A., & Freeman, D. (1999). Cognitive approaches to delusions: a critical review of theories and evidence. *The British journal of clinical psychology*, *38*(2), 113–154. https://doi.org/10.1348/014466599162700
Garety, P., & Freeman, D. (2013). The past and future of delusions research: From the inexplicable to the treatable. *British Journal of Psychiatry, 203*(5), 327-333. doi:10.1192/bjp.bp.113.126953
Garety, P. A., Hemsley, D. R., & Wessely, S. (1991). Reasoning in deluded schizophrenic and paranoid patients. Biases in performance on a probabilistic inference task. *The Journal of nervous and mental disease*, *179*(4), 194–201. https://doi.org/10.1097/00005053-199104000-00003
Garety, P., Joyce, E., Jolley, S., Emsley, R., Waller, H., Kuipers, E., . . . Freeman, D. (2013). Neuropsychological functioning and jumping to conclusions in delusions. *Schizophrenia Research, 150*(2-3), 570-574. doi:10.1016/j.schres.2013.08.035
Garety, P. A., Bebbington, P., Fowler, D., Freeman, D., & Kuipers, E. (2007). Implications for neurobiological research of cognitive models of psychosis: a theoretical paper. *Psychological Medicine, 37*(10), 1377-1391. doi:10.1017/S003329170700013X
Garety, P. A., Kuipers, E., Fowler, D., Freeman, D., & Bebbington, P. E. (2001). A cognitive model of the positive symptoms of psychosis. *Psychological Medicine, 31*(2), 189-195. doi:10.1017/S0033291701003312
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*: CRC press.
Gill K, & Grace, A. (2016). The Role of Neurotransmitters in Schizophrenia. In S.C. Schulz, M.F. Green, & K.J. Nelson (Eds.), *Schizophrenia and Psychotic Spectrum Disorders* (pp. 153-184). Oxford University Press. doi:10.1093/med/9780199378067.003.0010
Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cognitive, Affective, & Behavioral Neuroscience, 10(2), 252–269. doi:10.3758/CABN.10.2.252
Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2008). Determining a Role for Ventromedial Prefrontal Cortex in Encoding Action-Based Value Signals During Reward-Related Decision Making. *Cerebral Cortex, 19*(2), 483-495. doi:10.1093/cercor/bhn098
Gold, J. M., Kool, W., Botvinick, M. M., Hubzin, L., August, S., & Waltz, J. A. (2015). Cognitive effort avoidance and detection in people with schizophrenia. *Cognitive, Affective, & Behavioral Neuroscience, 15*(1), 145-154. doi:10.3758/s13415-014-0308-5
Gold, J. M., Strauss, G. P., Waltz, J. A., Robinson, B. M., Brown, J. K., & Frank, M. J. (2013). Negative symptoms of schizophrenia are associated with abnormal effort-cost computations. *Biological Psychiatry, 74*(2), 130-136. doi:10.1016/j.biopsych.2012.12.022
Gold, J. M., Waltz, J. A., Matveeva, T. M., Kasanova, Z., Strauss, G. P., Herbener, E. S., . . . Frank, M. J. (2012). Negative Symptoms and the Failure to Represent the Expected Reward Value of Actions: Behavioral and Computational Modeling Evidence. *Archives of General Psychiatry, 69*(2), 129-138. doi:10.1001/archgenpsychiatry.2011.1269
Gold, J. M., Waltz, J. A., Prentice, K. J., Morris, S. E., & Heerey, E. A. (2008). Reward Processing in Schizophrenia: A Deficit in the Representation of Value. *Schizophrenia Bulletin, 34*(5), 835-847. doi:10.1093/schbul/sbn068

Gold, J. M., Wilk, C. M., McMahon, R. P., Buchanan, R. W., & Luck, S. J. (2003). Working memory for visual features and conjunctions in schizophrenia. *J Abnorm Psychol, 112*(1), 61-71.
Gorissen, M., Sanz, J. C., & Schmand, B. (2005). Effort and cognition in schizophrenia patients. *Schizophrenia Research, 78*(2), 199-208. doi:https://doi.org/10.1016/j.schres.2005.02.016
Granholm, E., Ruiz, I., Gallegos-Rodriguez, Y., Holden, J., & Link, P. C. (2016). Pupillary Responses as a Biomarker of Diminished Effort Associated With Defeatist Attitudes and Negative Symptoms in Schizophrenia. *Biological psychiatry*, *80*(8), 581–588. doi:https://doi.org/10.1016/j.biopsych.2015.08.037
Granholm, E., Verney, S. P., Perivoliotis, D., & Miura, T. (2006). Effortful Cognitive Resource Allocation and Negative Symptom Severity in Chronic Schizophrenia. *Schizophrenia Bulletin, 33*(3), 831-842. doi:10.1093/schbul/sbl040
Green, M. F., Horan, W. P., Barch, D. M., & Gold, J. M. (2015). Effort-Based Decision Making: A Novel Approach for Assessing Motivation in Schizophrenia. *Schizophrenia Bulletin, 41*(5), 1035-1044. doi:10.1093/schbul/sbv071
Hämmerer, D., Schwartenbeck, P., Gallagher, M., FitzGerald, T. H. B., Düzel, E., & Dolan, R. J. (2019). Older adults fail to form stable task representations during model-based reversal inference. *Neurobiology of Aging, 74*, 90-100. doi:https://doi.org/10.1016/j.neurobiolaging.2018.10.009
Han, S., Huettel, S. A., Raposo, A., Adcock, R. A., & Dobbins, I. G. (2010). Functional significance of striatal responses during episodic decisions: recovery or goal attainment? *J Neurosci, 30*(13), 4767-4775. doi:10.1523/jneurosci.3077-09.2010
Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139-183): North-Holland.
Hartmann-Riemer, M. N., Aschenbrenner, S., Bossert, M., Westermann, C., Seifritz, E., Tobler, P. N., . . . Kaiser, S. (2017). Deficits in reinforcement learning but no link to apathy in patients with schizophrenia. *Scientific Reports, 7*(1), 40352. doi:10.1038/srep40352
Hauser, T. U., Allen, M., Purg, N., Moutoussis, M., Rees, G., & Dolan, R. J. (2017). Noradrenaline blockade specifically enhances metacognitive performance. *Elife, 6*. doi:10.7554/eLife.24901
Hegelstad, W. T. V., Kreis, I., Tjelmeland, H., & Pfuhl, G. (2020). Psychosis and Psychotic-Like Symptoms Affect Cognitive Abilities but Not Motivation in a Foraging Task. *Frontiers in Psychology, 11*, 1632-1632. doi:10.3389/fpsyg.2020.01632
Heinrichs, R. W., Ammari, N., McDermid Vaz, S., & Miles, A. A. (2008). Are schizophrenia and schizoaffective disorder neuropsychologically distinguishable? *Schizophrenia Research, 99*(1-3), 149-154. doi:10.1016/j.schres.2007.10.007
Heinz, A., & Schlagenhauf, F. (2010). Dopaminergic Dysfunction in Schizophrenia: Salience Attribution Revisited. *Schizophrenia Bulletin, 36*(3), 472-485. doi:10.1093/schbul/sbq031
Hill, S. K., Reilly, J. L., Keefe, R. S., Gold, J. M., Bishop, J. R., Gershon, E. S., . . . Sweeney, J. A. (2013). Neuropsychological impairments in schizophrenia and psychotic bipolar disorder: findings from the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP) study. *Am J Psychiatry, 170*(11), 1275-1284. doi:10.1176/appi.ajp.2013.12101298
Horan, W. P., Blanchard, J. J., Clark, L. A., & Green, M. F. (2008). Affective traits in schizophrenia and schizotypy. *Schizophrenia Bulletin, 34*(5), 856-874. doi:10.1093/schbul/sbn083
Horan, W. P., Braff, D. L., Nuechterlein, K. H., Sugar, C. A., Cadenhead, K. S., Calkins, M. E., . . . Green, M. F. (2008). Verbal working memory impairments in individuals with schizophrenia and their first-degree relatives: Findings from the Consortium on the Genetics of Schizophrenia. *Schizophrenia Research, 103*(1), 218-228. doi:https://doi.org/10.1016/j.schres.2008.02.014
Horga, G., Schatz, K. C., Abi-Dargham, A., & Peterson, B. S. (2014). Deficits in predictive coding underlie hallucinations in schizophrenia. *J Neurosci, 34*(24), 8072-8082. doi:10.1523/JNEUROSCI.0200-14.2014
Howells, F. M., Stein, D. J., & Russell, V. A. (2010). Perceived mental effort correlates with changes in tonic arousal during attentional tasks. *Behavioral and Brain Functions, 6*(1), 39. doi:10.1186/1744-9081-6-39

Howes, O. D., & Kapur, S. (2009). The Dopamine Hypothesis of Schizophrenia: Version III—The Final Common Pathway. *Schizophrenia Bulletin, 35*(3), 549-562. doi:10.1093/schbul/sbp006

Huq, S. F., Garety, P. A., & Hemsley, D. R. (1988) Probabilistic judgements in deluded and non-deluded subjects, *The Quarterly Journal of Experimental Psychology Section A, 40:4*, 801-812, doi: 10.1080/14640748808402300

Jäger, M., Haack, S., Becker, T., & Frasch, K. (2011). Schizoaffective disorder–an ongoing challenge for psychiatric nosology. *European Psychiatry, 26*(3), 159-165. doi:https://doi.org/10.1016/j.eurpsy.2010.03.010

Danion, J. M., Gokalsing, E., Robert, P., Massin-Krauss, M., & Bacon, E. (2001). Defective relationship between subjective experience and behavior in schizophrenia. *American Journal of Psychiatry*, *158*(12), 2064-2066. doi:10.1176/appi.ajp.158.12.2064

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off: evidence for the adaptive gain theory. Journal of Cognitive Neuroscience, 23(7), 1587–1596. doi: 10.1162/jocn.2010.21548

Joyce, E. M., & Roiser, J. P. (2007). Cognitive heterogeneity in schizophrenia. *Current Opinion in Psychiatry, 20*(3), 268-272. doi:10.1097/YCO.0b013e3280ba4975

Kane, J., Leucht, S., Carpenter, D., & Docherty, J. (2003). Expert Consensus Guidelines for Optimizing Pharmacologic Treatment of Psychotic Disorders. *The Journal of clinical psychiatry, 64 Suppl 12*, 5-19.

Kaplan, C. M., Saha, D., Molina, J. L., Hockeimer, W. D., Postell, E. M., Apud, J. A., . . . Tan, H. Y. (2016). Estimating changing contexts in schizophrenia. *Brain, 139*(7), 2082-2095. doi:10.1093/brain/aww095

Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry, 160*(1), 13-23. doi:10.1176/appi.ajp.160.1.13

Karvelis, P., Seitz, A. R., Lawrie, S. M., & Seriès, P. (2018). Autistic traits, but not schizotypy, predict increased weighting of sensory information in Bayesian visual integration. *Elife, 7*, e34115. doi:10.7554/eLife.34115

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. . *Schizophrenia Bulletin, 13*(2), 261-276.

Kempf, L., Hussain, N., & Potash, J. B. (2005). Mood disorder with psychotic features, schizoaffective disorder, and schizophrenia with mood features: trouble at the borders. *Int Rev Psychiatry, 17*(1), 9-19. doi:10.1080/09540260500064959

Khan, A., Lindenmayer, J.-P., Opler, M., Yavorsky, C., Rothman, B., & Lucic, L. (2013). A new Integrated Negative Symptom structure of the Positive and Negative Syndrome Scale (PANSS) in schizophrenia using item response analysis. *Schizophrenia Research, 150*(1), 185-196. doi:https://doi.org/10.1016/j.schres.2013.07.007

Kircher, T. T. J., Koch, K., Stottmeister, F., & Durst, V. (2007). Metacognition and Reflexivity in Patients with Schizophrenia. *Psychopathology, 40*(4), 254-260. doi:10.1159/000101730

Knoblich, G., Stottmeister, F., & Kircher, T. (2004). Self-monitoring in patients with schizophrenia. *Psychological Medicine, 34*(8), 1561-1569. doi:10.1017/S0033291704002454

Koch, K., Schachtzabel, C., Wagner, G., Schikora, J., Schultz, C., Reichenbach, J. R., . . . Schlösser, R. G. (2010). Altered activation in association with reward-related trial-and-error learning in patients with schizophrenia. *NeuroImage, 50*(1), 223-232. doi:10.1016/j.neuroimage.2009.12.031

Koren, D., Seidman, L. J., Poyurovsky, M., Goldsmith, M., Viksman, P., Zichel, S., & Klein, E. (2004). The neuropsychological basis of insight in first-episode schizophrenia: a pilot metacognitive study. *Schizophrenia Research, 70*(2), 195-202. doi:https://doi.org/10.1016/j.schres.2004.02.004

Kornetsky, C., & Mirsky, A. F. (1965). On certain psychopharmacological and physiological differences between schizophrenic and normal persons. *Psychopharmacologia, 8*(5), 309-318. doi:10.1007/BF00453508

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nat Neurosci, 21*(9), 1148-1160. doi:10.1038/s41593-018-0210-5

Krug, A., Cabanis, M., Pyka, M., Pauly, K., Kellermann, T., Walter, H., . . . Kircher, T. (2014). Attenuated prefrontal activation during decision-making under uncertainty in schizophrenia: A multi-center fMRI study. *Schizophrenia Research, 152*(1), 176-183. doi:https://doi.org/10.1016/j.schres.2013.11.007

Kuperberg, G., & Heckers, S. (2000). Schizophrenia and cognitive function. *Current opinion in neurobiology, 10*(2), 205-210. doi:https://doi.org/10.1016/S0959-4388(00)00068-4

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on psychological science, 7*(1), 18-27.

Lake, C. R., & Hurwitz, N. (2007). Schizoaffective disorder merges schizophrenia and bipolar disorders as one disease--there is no schizoaffective disorder. *Curr Opin Psychiatry, 20*(4), 365-379. doi:10.1097/YCO.0b013e3281a305ab

Langdon, R., Ward, P. B., & Coltheart, M. (2010). Reasoning anomalies associated with delusions in schizophrenia. *Schizophr Bull, 36*(2), 321-330. doi:10.1093/schbul/sbn069

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nat Neurosci, 20*(9), 1293-1299. doi:10.1038/nn.4615

Lee, J., & Park, S. (2005). Working Memory Impairments in Schizophrenia: A Meta-Analysis. *Journal of Abnormal Psychology, 114*(4), 599-611. doi:10.1037/0021-843X.114.4.599

Lee, Y., Kim, Y.-T., Seo, E., Park, O., Jeong, S.-H., Kim, S. H., & Lee, S.-J. (2007). Dissociation of emotional decision-making from cognitive decision-making in chronic schizophrenia. *Psychiatry Research, 152*(2), 113-120. doi:https://doi.org/10.1016/j.psychres.2006.02.001

Lehrl, S., Triebig, G., & Fischer, B. (1995). Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurologica Scandinavica, 91*(5), 335-345. doi:10.1111/j.1600-0404.1995.tb07018.x

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*: SAGE publications.

Li, C. T., Lai, W. S., Liu, C. M., & Hsu, Y. F. (2014). Inferring reward prediction errors in patients with schizophrenia: a dynamic reward task for reinforcement learning. *Front Psychol, 5*, 1282. doi:10.3389/fpsyg.2014.01282

Lincoln, T. M., Salzmann, S., Ziegler, M., & Westermann, S. (2011). When does jumping-to-conclusions reach its peak? The interaction of vulnerability and situation-characteristics in social reasoning. *J Behav Ther Exp Psychiatry, 42*(2), 185-191. doi:10.1016/j.jbtep.2010.09.005

Lindenmayer, J. P., Bernstein-Hyman, R., Grochowski, S., & Bark, N. (1995). Psychopathology of Schizophrenia: Initial Validation of a 5-Factor Model. *Psychopathology, 28*(1), 22-31. doi:10.1159/000284896

Logue, S. F., & Gould, T. J. (2014). The neural and genetic basis of executive function: attention, cognitive flexibility, and response inhibition. *Pharmacol Biochem Behav, 123*, 45-54. doi:10.1016/j.pbb.2013.08.007

Lysaker, P. H., Bryson, G. J., Lancaster, R. S., Evans, J. D., & Bell, M. D. (2003). Insight in schizophrenia: associations with executive function and coping style. *Schizophrenia Research, 59*(1), 41-47. doi:https://doi.org/10.1016/S0920-9964(01)00383-8

Lysaker, P. H., & Dimaggio, G. (2014). Metacognitive Capacities for Reflection in Schizophrenia: Implications for Developing Treatments. *Schizophrenia Bulletin, 40*(3), 487-491. doi:10.1093/schbul/sbu038

Lysaker, P. H., Erickson, M., Ringer, J., Buck, K. D., Semerari, A., Carcione, A., & Dimaggio, G. (2011). Metacognition in schizophrenia: The relationship of mastery to coping, insight, self-esteem, social anxiety, and various facets of neurocognition. *British Journal of Clinical Psychology, 50*(4), 412-424. doi:10.1111/j.2044-8260.2010.02003.x

Lysaker, P. H., Gagen, E., Wright, A., Vohs, J. L., Kukla, M., Yanos, P. T., & Hasson-Ohayon, I. (2019). Metacognitive Deficits Predict Impaired Insight in Schizophrenia Across Symptom Profiles: A Latent Class Analysis. *Schizophr Bull, 45*(1), 48-56. doi:10.1093/schbul/sby142

Lysaker, P. H., Vohs, J., Hamm, J. A., Kukla, M., Minor, K. S., de Jong, S., . . . Dimaggio, G. (2014). Deficits in metacognitive capacity distinguish patients with schizophrenia from those with prolonged medical adversity. *Journal of Psychiatric Research, 55*, 126-132. doi:https://doi.org/10.1016/j.jpsychires.2014.04.011

Lysaker, P. H., Vohs, J., Hillis, J. D., Kukla, M., Popolo, R., Salvatore, G., & Dimaggio, G. (2013). Poor insight into schizophrenia: contributing factors, consequences and emerging treatment approaches. *Expert Review of Neurotherapeutics, 13*(7), 785-793. doi:10.1586/14737175.2013.811150

Lysaker, P. H., Vohs, J. L., Ballard, R., Fogley, R., Salvatore, G., Popolo, R., & Dimaggio, G. (2013). Metacognition, self-reflection and recovery in schizophrenia. *Future Neurology, 8*(1), 103-115. doi:10.2217/fnl.12.78

Madre, M., Canales-Rodríguez, E. J., Ortiz-Gil, J., Murru, A., Torrent, C., Bramon, E., . . . Amann, B. L. (2016). Neuropsychological and neuroimaging underpinnings of schizoaffective disorder: a systematic review. *Acta Psychiatr Scand, 134*(1), 16-30. doi:10.1111/acps.12564

Maia, T. V., & Frank, M. J. (2017). An Integrative Perspective on the Role of Dopamine in Schizophrenia. *Biological Psychiatry, 81*(1), 52-66. doi:https://doi.org/10.1016/j.biopsych.2016.05.021

Maier, W. (2006). Do schizoaffective disorders exist at all? *Acta Psychiatrica Scandinavica, 113*(5), 369-371. doi:https://doi.org/10.1111/j.1600-0447.2006.00763.x

Malaspina, D., Owen, M. J., Heckers, S., Tandon, R., Bustillo, J., Schultz, S., . . . Carpenter, W. (2013). Schizoaffective Disorder in the DSM-5. *Schizophrenia Research, 150*(1), 21-25. doi:https://doi.org/10.1016/j.schres.2013.04.026

Martino, D. J., Bucay, D., Butman, J. T., & Allegri, R. F. (2007). Neuropsychological frontal impairments and negative symptoms in schizophrenia. *Psychiatry Research, 152*(2), 121-128. doi:https://doi.org/10.1016/j.psychres.2006.03.002

Mathot, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition, 1*(1), 16. doi:http://doi.org/10.5334/joc.18

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods, 50*(1), 94-106. doi:10.3758/s13428-017-1007-2

McKay, R., Langdon, R., & Coltheart, M. (2006). Need for closure, jumping to conclusions, and decisiveness in delusion-prone individuals. *J Nerv Ment Dis, 194*(6), 422-426. doi:10.1097/01.nmd.0000221353.44132.25

McKay, R., Langdon, R., & Coltheart, M. (2007). Jumping to delusions? Paranoia, probabilistic reasoning, and need for closure. *Cognitive Neuropsychiatry, 12*(4), 362-376. doi:10.1080/13546800701203769

Menon, M., Mizrahi, R., & Kapur, S. (2008). 'Jumping to conclusions' and delusions in psychosis: Relationship and response to treatment. *Schizophrenia Research, 98*(1), 225-231. doi:https://doi.org/10.1016/j.schres.2007.08.021

Menon, M., Pomarol-Clotet, E., McKenna, P. J., & McCarthy, R. A. (2006). Probabilistic reasoning in schizophrenia: A comparison of the performance of deluded and nondeluded schizophrenic patients and exploration of possible cognitive underpinnings. *Cognitive Neuropsychiatry, 11*(6), 521-536. doi:10.1080/13546800544000046

Merrin, J., Kinderman, P., & Bentall, R. P. (2007). 'Jumping to Conclusions' and Attributional Style in Persecutory Delusions. *Cognitive Therapy and Research, 31*(6), 741-758. doi:10.1007/s10608-006-9097-7

Mishara, A. L., & Sterzer, P. (2015). Phenomenology is Bayesian in its application to delusions. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, *14*(2), 185–186. https://doi.org/10.1002/wps.20213

Moritz, S., Ferahli, S., & Naber, D. (2004). Memory and attention performance in psychiatric patients: lack of correspondence between clinician-rated and patient-rated functioning with neuropsychological test results. *Journal of the International Neuropsychological Society*, *10*(4), 623-633.

Moritz, S., & Woodward, T. S. (2005). Jumping to conclusions in delusional and non-delusional schizophrenic patients. *British Journal of Clinical Psychology*, *44*(2), 193-207.

Moritz, S., Woodward, T. S., & Chen, E. (2006). Investigation of metamemory dysfunctions in first-episode schizophrenia. *Schizophrenia Research*, *81*(2-3), 247-252.

Moritz, S., Woodward, T. S., Cuttler, C., Whitman, J. C., & Watson, J. M. (2004). False memories in schizophrenia. *Neuropsychology*, *18*(2), 276-283.

Moritz, S., Woodward, T. S., & Rodriguez-Raecke, R. (2006). Patients with schizophrenia do not produce more false memories than controls but are more confident in them. *Psychological medicine*, *36*(5), 659-667.
Moritz, S., Woodward, T. S., Whitman, J. C., & Cuttler, C. (2005). Confidence in errors as a possible basis for delusions in schizophrenia. *The Journal of nervous and mental disease*, *193*(1), 9-16.
Moritz, S., Klein, J. P., Desler, T., Lill, H., Gallinat, J., & Schneider, B. C. (2017). Neurocognitive deficits in schizophrenia. Are we making mountains out of molehills? *Psychol Med, 47*(15), 2602-2612. doi:10.1017/s0033291717000939
Moritz, S., Stöckert, K., Hauschildt, M., Lill, H., Jelinek, L., Beblo, T., . . . Arlt, S. (2017). Are we exaggerating neuropsychological impairment in depression? Reopening a closed chapter. *Expert Review of Neurotherapeutics, 17*(8), 839-846. doi:https://doi.org/10.1080/14737175.2017.1347040
Moritz, S., Veckenstedt, R., Hottenrott, B., Woodward, T. S., Randjbar, S., & Lincoln, T. M. (2010). Different sides of the same coin? Intercorrelations of cognitive biases in schizophrenia. *Cogn Neuropsychiatry, 15*(4), 406-421. doi:10.1080/13546800903399993
Moritz, S., Veckenstedt, R., Randjbar, S., Hottenrott, B., Woodward, T. S., von Eckstaedt, F. V., . . . Lincoln, T. M. (2009). Decision making under uncertainty and mood induction: further evidence for liberal acceptance in schizophrenia. *Psychol Med, 39*(11), 1821-1829. doi:10.1017/S0033291709005923
Moritz, S., & Woodward, T. S. (2006). The contribution of metamemory deficits to schizophrenia. *J Abnorm Psychol, 115*(1), 15-25. doi:10.1037/0021-843X.15.1.15
Moritz, S., & Woodward, T. S. (2007). Metacognitive training in schizophrenia: from basic research to knowledge translation and intervention. *Current Opinion in Psychiatry, 20*(6), 619-625. doi:10.1097/YCO.0b013e3282f0b8ed
Moritz, S., Woodward, T. S., Jelinek, L., & Klinge, R. (2008). Memory and metamemory in schizophrenia: A liberal acceptance account of psychosis. *Psychological Medicine, 38*, 825-832.
Moritz, S., Woodward, T. S., & Lambert, M. (2007). Under what circumstances do patients with schizophrenia jump to conclusions? A liberal acceptance account. *Br J Clin Psychol, 46*(Pt 2), 127-137. doi:10.1348/014466506X129862
Moutoussis, M., Bentall, R. P., El-Deredy, W., & Dayan, P. (2011). Bayesian modelling of Jumping-to-Conclusions bias in delusional patients. *Cogn Neuropsychiatry, 16*(5), 422-447. doi:10.1080/13546805.2010.548678
Murray, G. K., Cheng, F., Clark, L., Barnett, J. H., Blackwell, A. D., Fletcher, P. C., . . . Jones, P. B. (2008). Reinforcement and reversal learning in first-episode psychosis. *Schizophrenia Bulletin, 34*(5), 848-855. doi:10.1093/schbul/sbn078
Mushtaq, F., Bland, A., & Schaefer, A. (2011). Uncertainty and Cognitive Control. *Frontiers in Psychology, 2*(249). doi:10.3389/fpsyg.2011.00249
Naicker, P., Anoopkumar-Dukie, S., Grant, G. D., Neumann, D. L., & Kavanagh, J. J. (2016). Central cholinergic pathway involvement in the regulation of pupil diameter, blink rate and cognitive function. *Neuroscience, 334*, 180-190. doi:https://doi.org/10.1016/j.neuroscience.2016.08.009
Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat Neurosci, 15*(7), 1040-1046. doi:10.1038/nn.3130
Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment. *The Journal of Neuroscience, 30*(37), 12366-12378. doi:10.1523/jneurosci.0822-10.2010
Norman, E., Pfuhl, G., Sæle, R. G., Svartdal, F., Låg, T., & Dahl, T. I. (2019). Metacognition in Psychology. *Review of General Psychology, 23*(4), 403-424. doi:10.1177/1089268019883821
Nuechterlein, K. H., & Dawson, M. E. (1984). A Heuristic Vulnerability/Stress Model of Schizophrenic Episodes. *Schizophrenia Bulletin, 10*(2), 300-312. doi:10.1093/schbul/10.2.300
Ogino, S., Miyamoto, S., Miyake, N., & Yamaguchi, N. (2014). Benefits and limits of anticholinergic use in schizophrenia: Focusing on its effect on cognitive function. *Psychiatry and Clinical Neurosciences, 68*(1), 37-49. doi:https://doi.org/10.1111/pcn.12088

Pantelis, C., Barber, F. Z., Barnes, T. R., Nelson, H. E., Owen, A. M., & Robbins, T. W. (1999). Comparison of set-shifting ability in patients with chronic schizophrenia and frontal lobe damage. *Schizophr Res, 37*(3), 251-270. doi:10.1016/s0920-9964(98)00156-x

Pantelis, C., Wood, S. J., Proffitt, T. M., Testa, R., Mahony, K., Brewer, W. J., . . . McGorry, P. D. (2009). Attentional set-shifting ability in first-episode and established schizophrenia: Relationship to working memory. *Schizophr Res, 112*(1-3), 104-113. doi:10.1016/j.schres.2009.03.039

Payzan-LeNestour, E., & Bossaerts, P. (2011). Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comput Biol, 7*(1), e1001048. doi:10.1371/journal.pcbi.1001048

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87*(6), 532-552. doi:10.1037/0033-295X.87.6.532

Peters, E., & Garety, P. (2006). Cognitive functioning in delusions: a longitudinal analysis. *Behaviour research and therapy, 44*(4), 481-514.

Picardi, A., Viroli, C., Tarsitani, L., Miglio, R., de Girolamo, G., Dell'Acqua, G., & Biondi, M. (2012). Heterogeneity and symptom structure of schizophrenia. *Psychiatry Research, 198*(3), 386-394. doi:https://doi.org/10.1016/j.psychres.2011.12.051

Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science, 357*(6351), 596-600. doi:10.1126/science.aan3458

Preuschoff, K., t Hart, B. M., & Einhauser, W. (2011). Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. *Front Neurosci, 5*, 115. doi:10.3389/fnins.2011.00115

Ramsøy, T. Z., Friis-Olivarius, M., Jacobsen, C., Jensen, S. B., & Skov, M. (2012). Effects of perceptual uncertainty on arousal and preference across different visual domains. *Journal of Neuroscience, Psychology, and Economics, 5*(4), 212.

Rausch, F., Mier, D., Eifler, S., Esslinger, C., Schilling, C., Schirmbeck, F., . . . Zink, M. (2014). Reduced activation in ventral striatum and ventral tegmental area during probabilistic decision-making in schizophrenia. *Schizophrenia Research, 156*(2), 143-149. doi:https://doi.org/10.1016/j.schres.2014.04.020

R Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna: R Core Team.

Reddy, L. F., Horan, W. P., Barch, D. M., Buchanan, R. W., Gold, J. M., Marder, S. R., . . . Green, M. F. (2018). Understanding the Association Between Negative Symptoms and Performance on Effort-Based Decision-Making Tasks: The Importance of Defeatist Performance Beliefs. *Schizophr Bull, 44*(6), 1217-1226. doi:10.1093/schbul/sbx156

Reddy, L. F., Waltz, J. A., Green, M. F., Wynn, J. K., & Horan, W. P. (2016). Probabilistic Reversal Learning in Schizophrenia: Stability of Deficits and Potential Causal Mechanisms. *Schizophr Bull, 42*(4), 942-951. doi:10.1093/schbul/sbv226

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton Century Crofts.

Ritter, L. M., Meador-Woodruff, J. H., & Dalack, G. W. (2004). Neurocognitive measures of prefrontal cortical dysfunction in schizophrenia. *Schizophrenia Research, 68*(1), 65-73. doi:https://doi.org/10.1016/S0920-9964(03)00086-0

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, 41*(5), 1183-1191. doi:10.1093/schbul/sbu187

Samuels, E. R., & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part II: physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. *Current neuropharmacology, 6*(3), 254-285. doi:10.2174/157015908785777193

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nat Rev Neurosci, 10*(3), 211-223. doi:10.1038/nrn2573

Satterthwaite, T. D., Ruparel, K., Loughead, J., Elliott, M. A., Gerraty, R. T., Calkins, M. E., . . . Wolf, D. H. (2012). Being right is its own reward: load and performance related ventral striatum activation to correct responses during a working memory task in youth. *NeuroImage, 61*(3), 723-729. doi:10.1016/j.neuroimage.2012.03.060

Schaefer, J., Giangrande, E., Weinberger, D. R., & Dickinson, D. (2013). The global cognitive impairment in schizophrenia: Consistent over decades and around the world. *Schizophrenia Research, 150*(1), 42-50. doi:https://doi.org/10.1016/j.schres.2013.07.009

Schlagenhauf, F., Huys, Q. J. M., Deserno, L., Rapp, M. A., Beck, A., Heinze, H.-J., . . . Heinz, A. (2014). Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *NeuroImage, 89*, 171-180. doi:https://doi.org/10.1016/j.neuroimage.2013.11.034

Schultz, W. (2002). Getting Formal with Dopamine and Reward. *Neuron, 36*(2), 241-263. doi:https://doi.org/10.1016/S0896-6273(02)00967-4

Semerari, A., Carcione, A., Dimaggio, G., Falcone, M., Nicolò, G., Procacci, M., & Alleva, G. (2003). How to evaluate metacognitive functioning in psychotherapy? The metacognition assessment scale and its applications. *Clinical Psychology & Psychotherapy, 10*(4), 238-261.

Sevy, S., Burdick, K. E., Visweswaraiah, H., Abdelmessih, S., Lukin, M., Yechiam, E., & Bechara, A. (2007). Iowa Gambling Task in schizophrenia: A review and new data in patients with schizophrenia and co-occurring cannabis use disorders. *Schizophrenia Research, 92*(1), 74-84. doi:https://doi.org/10.1016/j.schres.2007.01.005

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences, 18*(4), 186-193. doi:https://doi.org/10.1016/j.tics.2014.01.006

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of clinical psychiatry*.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a Rational and Mechanistic Account of Mental Effort. *Annual Review of Neuroscience, 40*(1), 99-124. doi:10.1146/annurev-neuro-072116-031526

Shimamura, A. P. (2000). Toward a Cognitive Neuroscience of Metacognition. *Consciousness and Cognition, 9*(2), 313-323. doi:https://doi.org/10.1006/ccog.2000.0450

Shurman, B., Horan, W. P., & Nuechterlein, K. H. (2005). Schizophrenia patients demonstrate a distinctive pattern of decision-making impairment on the Iowa Gambling Task. *Schizophrenia Research, 72*(2), 215-224. doi:https://doi.org/10.1016/j.schres.2004.03.020

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science, 5*(6), 679-692.

Speechley, W. J., Whitman, J. C., & Woodward, T. S. (2010). The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. *Journal of Psychiatry & Neuroscience, 35*(1), 7-17. doi:10.1503/jpn.090025

Stahl, S. M., & Stahl, S. M. (2013). *Stahl's essential psychopharmacology: neuroscientific basis and practical applications*: Cambridge university press.

Steinhauer, S., & Zubin, J. (1982). Vulnerability to schizophrenia: Information processing in the pupil and event-related potential. In *Biological markers in psychiatry and neurology* (pp. 371-385). Pergamon.

Steinhauer, S. R., Hakerem, G., & Spring, B. J. (1979). The pupillary response as a potential indicator of vulnerability to schizophrenia. *Psychopharmacology bulletin*, *15*(1), 44-45.

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., . . . Corlett, P. R. (2018). The Predictive Coding Account of Psychosis. *Biol Psychiatry*. doi:10.1016/j.biopsych.2018.05.015

Sterzer, P., Voss, M., Schlagenhauf, F., & Heinz, A. (2018). Decision-making in schizophrenia: A predictive-coding perspective. *NeuroImage*. doi:https://doi.org/10.1016/j.neuroimage.2018.05.074

Stopper, C. M., & Floresco, S. B. (2014). Dopaminergic Circuitry and Risk/Reward Decision Making: Implications for Schizophrenia. *Schizophrenia Bulletin, 41*(1), 9-14. doi:10.1093/schbul/sbu165

Strauss, G. P., Frank, M. J., Waltz, J. A., Kasanova, Z., Herbener, E. S., & Gold, J. M. (2011). Deficits in positive reinforcement learning and uncertainty-driven exploration are associated with distinct aspects of negative symptoms in schizophrenia. *Biol Psychiatry, 69*(5), 424-431. doi:10.1016/j.biopsych.2010.10.015

Strauss, G. P., Waltz, J. A., & Gold, J. M. (2014). A review of reward processing and motivational impairment in schizophrenia. *Schizophr Bull, 40 Suppl 2*, S107-116. doi:10.1093/schbul/sbt197

Struglia, F., Stratta, P., Gianfelice, D., Pacifico, R., Riccardi, I., & Rossi, A. (2011). Decision-making impairment in schizophrenia: Relationships with positive symptomatology. *Neuroscience Letters, 502*(2), 80-83. doi:https://doi.org/10.1016/j.neulet.2011.07.017

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning* (Vol. 135): MIT press Cambridge.

Takeda, T., Nakataki, M., Ohta, M., Hamatani, S., Matsuura, K., & Ohmori, T. (2018). Effect of cognitive function on jumping to conclusion in patients with schizophrenia. *Schizophrenia Research: Cognition, 12*, 50-55. doi:10.1016/j.scog.2018.04.002

Tek, C., Gold, J., Blaxton, T., Wilk, C., McMahon, R. P., & Buchanan, R. W. (2002). Visual Perceptual and Working Memory Impairments in Schizophrenia. *Archives of General Psychiatry, 59*(2), 146-153. doi:10.1001/archpsyc.59.2.146

Torniainen, M., Suvisaari, J., Partonen, T., Castaneda, A. E., Kuha, A., Suokas, J., . . . Tuulio-Henriksson, A. (2012). Cognitive impairments in schizophrenia and schizoaffective disorder: relationship with clinical characteristics. *J Nerv Ment Dis, 200*(4), 316-322. doi:10.1097/NMD.0b013e31824cb359

Townsend, L. A., Malla, A. K., & Norman, R. M. (2001). Cognitive functioning in stabilized first-episode psychosis patients. *Psychiatry Res, 104*(2), 119-131. doi:10.1016/s0165-1781(01)00302-x

Tricomi, E., & Fiez, J. A. (2008). Feedback signals in the caudate reflect goal achievement on a declarative memory task. *NeuroImage, 41*(3), 1154-1167. doi:10.1016/j.neuroimage.2008.02.066

Ullsperger, M., & von Cramon, D. Y. (2003). Error monitoring using external feedback: specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *J Neurosci, 23*(10), 4308-4314. doi:10.1523/jneurosci.23-10-04308.2003

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nat Commun, 8*, 14637. doi:10.1038/ncomms14637

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychol Rev, 121*(4), 649-675. doi:10.1037/a0037665

van der Gaag, M., Hoffman, T., Remijsen, M., Hijman, R., de Haan, L., van Meijel, B., . . . Wiersma, D. (2006). The five-factor model of the Positive and Negative Syndrome Scale II: A ten-fold cross-validation of a revised model. *Schizophrenia Research, 85*(1), 280-287. doi:https://doi.org/10.1016/j.schres.2006.03.021

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review, 25*(6), 2005-2015. doi:10.3758/s13423-018-1432-y

van Kammen, D. P., & Kelley, M. (1991). Dopamine and norepinephrine activity in schizophrenia: An intergrative perspective. *Schizophrenia Research, 4*(2), 173-191. doi:https://doi.org/10.1016/0920-9964(91)90032-M

van Schalkwyk, G. I., Volkmar, F. R., & Corlett, P. R. (2017). A predictive coding account of psychotic symptoms in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*(5), 1323-1340. doi:10.1007/s10803-017-3065-9

Vos, T., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., . . . Murray, C. J. L. (2017). Global, regional, and national incidence, prevalence, and years lived with disability

for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet, 390*(10100), 1211-1259. doi:10.1016/S0140-6736(17)32154-2

Vrugt, A., & Oort, F. J. (2008). Metacognition, achievement goals, study strategies and academic achievement: pathways to achievement. *Metacognition and Learning, 3*(2), 123-146. doi:10.1007/s11409-008-9022-4

Waltz, J. A., Frank, M. J., Robinson, B. M., & Gold, J. M. (2007). Selective Reinforcement Learning Deficits in Schizophrenia Support Predictions from Computational Models of Striatal-Cortical Dysfunction. *Biological Psychiatry, 62*(7), 756-764. doi:https://doi.org/10.1016/j.biopsych.2006.09.042

Waltz, J. A., Frank, M. J., Wiecki, T. V., & Gold, J. M. (2011). Altered probabilistic learning and response biases in schizophrenia: Behavioral evidence and neurocomputational modeling. *Neuropsychology, 25*(1), 86-97. doi:10.1037/a0020882

Waltz, J. A., & Gold, J. M. (2007). Probabilistic reversal learning impairments in schizophrenia: Further evidence of orbitofrontal dysfunction. *Schizophrenia Research, 93*(1), 296-303. doi:https://doi.org/10.1016/j.schres.2007.03.010

Waltz, J. A., Kasanova, Z., Ross, T. J., Salmeron, B. J., McMahon, R. P., Gold, J. M., & Stein, E. A. (2013). The Roles of Reward, Default, and Executive Control Networks in Set-Shifting Impairments in Schizophrenia. *PLoS One, 8*(2), e57257. doi:10.1371/journal.pone.0057257

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson, 22*(498), 816-827.

Wolf, D. H., Gerraty, R., Satterthwaite, T. D., Loughead, J., Campellone, T., Elliott, M. A., . . . Gur, R. E. (2011). Striatal intrinsic reinforcement signals during recognition memory: relationship to response bias and dysregulation in schizophrenia. *Front Behav Neurosci, 5*, 81. doi:10.3389/fnbeh.2011.00081

Wolf, D. H., Satterthwaite, T. D., Kantrowitz, J. J., Katchmar, N., Vandekar, L., Elliott, M. A., & Ruparel, K. (2014). Amotivation in schizophrenia: integrated assessment with behavioral, clinical, and imaging measures. *Schizophr Bull, 40*(6), 1328-1337. doi:10.1093/schbul/sbu026

Woodward, T. S., Mizrahi, R., Menon, M., & Christensen, B. K. (2009). Correspondences between theory of mind, jumping to conclusions, neuropsychological measures and the symptoms of schizophrenia. *Psychiatry Res, 170*(2-3), 119-123. doi:10.1016/j.psychres.2008.10.018

Worthy, D. A., & Todd Maddox, W. (2014). A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W.K. Estes. *Journal of Mathematical Psychology, 59*, 41-49. doi:https://doi.org/10.1016/j.jmp.2013.10.001

Yamamoto, K., & Hornykiewicz, O. (2004). Proposal for a noradrenaline hypothesis of schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 28*(5), 913-922. doi:https://doi.org/10.1016/j.pnpbp.2004.05.033

Yılmaz, A., Simsek, F., & Gonul, A. S. (2012). Reduced reward-related probability learning in schizophrenia patients. *Neuropsychiatric disease and treatment, 8*, 27-34. doi:10.2147/NDT.S26243

Yu, A. J., & Dayan, P. (2005). Uncertainty, Neuromodulation, and Attention. *Neuron, 46*(4), 681-692. doi:https://doi.org/10.1016/j.neuron.2005.04.026

Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J., & Lamm, C. (2020). Using reinforcement learning models in social neuroscience: frameworks, pitfalls and suggestions of best practices. *Social Cognitive and Affective Neuroscience*. doi:10.1093/scan/nsaa089

Paper I

Overestimation of volatility in schizophrenia and autism? A comparative study using a probabilistic reasoning task

Isabel Kreis[1]*, Robert Biegler[2], Håkon Tjelmeland[3], Matthias Mittner[1], Solveig Klæbo Reitan[4,5], & Gerit Pfuhl[1]

[1] Department of Psychology, Faculty of Health Sciences, UiT – The Arctic University of Norway, Tromsø, Norway

[2] Department of Psychology, Faculty of Social and Educational Sciences, Norwegian University of Science and Technology, Trondheim, Norway

[3] Department of Mathematical Sciences, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, Trondheim, Norway

[4] Department of Mental Health, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway

[5] Department of Mental Health, St Olav's University hospital, Trondheim, Norway

*Corresponding author

E-mail: isabel.v.kreis@uit.no (IK)

Abstract

Background and Objectives: A plethora of studies has investigated and compared social cognition in autism and schizophrenia ever since both conditions were first described in conjunction more than a century ago. Recent computational theories have proposed similar mechanistic explanations for various symptoms beyond social cognition. They are grounded in the idea of a general misestimation of uncertainty but so far, almost no studies have directly compared both conditions regarding uncertainty processing. The current study aimed to do so with a particular focus on estimation of volatility, i.e. the probability for the environment to change.

Methods: A probabilistic decision-making task and a visual working (meta-)memory task were administered to a sample of 86 participants (19 with a diagnosis of high-functioning autism, 21 with a diagnosis of schizophrenia, and 46 neurotypically developing individuals).

Results: While persons with schizophrenia showed lower visual working memory accuracy than neurotypical individuals, no significant group differences were found for metamemory or any of the probabilistic decision-making task variables. Nevertheless, exploratory analyses suggest that there may be an overestimation of volatility in subgroups of participants with autism and schizophrenia. Correlations revealed relationships between different variables reflecting (mis)estimation of uncertainty, visual working memory accuracy and metamemory.

Limitations: Limitations include the comparably small sample sizes of the autism and the schizophrenia group as well as the lack of cognitive ability and clinical symptom measures.

Conclusions: The results of the current study provide partial support for the notion of a general uncertainty misestimation account of autism and schizophrenia.

Keywords: psychosis; uncertainty; visual working memory; metacognition; Bayesian reasoning; computational psychiatry

Introduction

More than a century ago, the terms ‘autistic’ and ‘autism’ were coined to describe the social withdrawal observed in individuals with schizophrenia (SCZ) and a childhood form of SCZ, respectively [1]. While SCZ and autism spectrum disorders (ASD) are defined as distinct entities today [2], a substantial amount of research has investigated the shared characteristics of both conditions. Findings suggest an association from both a genetic [3, 4] and a cognitive-behavioral perspective, particularly within the social domain. Comparative and parallel studies have documented similarly impaired social cognitive abilities in SCZ and ASD relative to neurotypically developing (NT) individuals [5]. This concerns various subdomains, including theory of mind, i.e. the ability to infer others’ mental states [6, 7], eye gaze on faces [8], trustworthiness judgements and emotion identification [9]. In fact, a recent systematic review concluded that apart from emotion recognition there seem to be no clear and consistent differences between ASD and SCZ in terms of social cognitive performance [10].

While social cognition has been studied extensively, only few studies compare the two conditions in other cognitive domains [1]. However, results of separately conducted studies suggest similar decision-making impairments in non-social situations [11-13]. One decision-making bias that has extensively been investigated in persons with SCZ is the so-called “Jumping-to-Conclusions” (JTC) bias. It is usually assessed by the beads task in which beads are sampled from one out of two possible containers (e.g. bags) containing unlike amounts of differently colored beads. Based on the sampled beads, participants have to indicate what they believe to be the bag of origin [14]. Different versions of the task exist: in draws-to-decision versions, participants are free to sample as many beads as they want until they decide on the bag of origin. Here, the JTC bias is characterized by premature decisions, i.e. a decision on the

bag of origin after very few beads have been sampled. In graded estimates versions of the task participants indicate their certainty about the bag of origin after each bead. Here, reasoning biases include high (initial) certainty and over-adjustment of the reported estimates, meaning radical belief alterations in response to objectively only modest disconfirmatory evidence [15, 16]. Those biases are similar to the 'classical' JTC bias in that they all concern drastic decision-making in light of little evidence. While typically studied in SCZ, the JTC bias has also been found in ASD [17]. Conversely, Brosnan and colleagues reported that persons with ASD gathered more beads before making a decision [18].

Of the few studies directly comparing non-social decision-making in ASD and SCZ, Zhang and colleagues [19] found similar impairments in decision-making under different kinds of uncertainty, suggesting that both conditions may be characterized by misestimation of uncertainties. Such misestimations could also explain the aberrant behavior observed in the aforementioned beads task, where performance relies on Bayesian inference [20]. This perspective fits well with computational theories proposing similar mechanistic explanations based on misestimation of uncertainty for various symptoms of ASD and SCZ [21-23]. According to these theories, symptoms might be the result of (implicit) uncertainty misestimation on different levels in Bayesian belief hierarchies of the brain [24]. One 'level' concerns beliefs about the environment's volatility, i.e. the probability for the environment to change. The results of various studies indicate that both persons with ASD and persons with SCZ overestimate volatility, i.e. they seem to perceive the world as less stable. For example, they exhibit more (maladaptive) switching behavior in reversal learning tasks than NT individuals (ASD: [12, 25-27]; SCZ: [28-31]). Surprising events are thus attributed to a change in the overall stochastic structure of the environment [32] rather than to known uncertainties on lower levels, i.e. the expected uncertainty that arises naturally since some events are more likely than others in a stable but stochastic environment. Hence, new events will become more salient

as they might signal a relevant change in the environment when subjective volatility is high. Consequentially, beliefs are updated more drastically. This fits well with the over-adjustment of beliefs observed in the beads task, which in turn has been attributed to a "hypersalience" of new evidence [20].

Aberrant representation of uncertainties has also been described as a 'failure of metacognition' [24]. Metacognition can refer to both conscious reflective thought processes and automatic monitoring of one's own thoughts and cognitions [33]. Metacognitive performance is often determined by comparing self-reports and confidence ratings to actual performance [34]. Interestingly, impaired metacognition has been found in both SCZ [35, 36] and ASD [37, 38]. Further, previous studies have revealed a relationship between the JTC bias, corresponding decision confidence and metacognitive deficits in SCZ [36, 39, 40], but to what extent metacognition relates to higher level uncertainty estimation such as volatility remains to be elucidated.

A general misestimation of uncertainties could thus explain various cognitive-behavioral findings in both ASD and SCZ but it remains unclear if and to what extent both groups differ from each other when compared directly. This study aimed to investigate this question with a focus on volatility processing in a modified beads task and its relationship to belief updating, metacognition and working memory, to account for the potential role of general cognitive capacity.

Materials and methods

Persons with SCZ were contacted through a clinician at St. Olavs Hospital, Trondheim University Hospital, Norway, while persons with ASD were recruited through the patient interest group *Autismeforeningen* and, like NT control participants, through fliers and social media posts. Participants had to meet the following inclusion criteria: (1) 18 to 60 years of age,

(2) no current suicide intent, (3) no substance dependence, (4) IQ above 80, (5) a primary diagnosis from the schizophrenia spectrum (SCZ group) or high-functioning autism/Asperger (ASD group) or no psychiatric diagnosis at all (NT group). All participants in the SCZ group were inpatients who had previously been diagnosed according to the ICD-10 research criteria [41] in a consensus meeting assessing clinical reports with at least two senior psychiatrists or psychologists present, of which at least one had personally examined the patient. Diagnoses were confirmed by clinicians upon inclusion in the study. All participants in the ASD group reported prior diagnoses by independent clinicians. Where available, their diagnoses were confirmed through clinical records and their employer (a business exclusively employing persons with a confirmed ASD diagnosis). For three participants with ASD, no such confirmation was available. For all participants, written informed consent was obtained prior to the study. The study was approved by the Central-Norwegian regional committee for medical and health research ethics (REC Central; reference no.: 2014/1648). In total, 92 participants were recruited, whereof six were excluded since they did not complete enough (≥ 80%) trials of the administered tasks. A subset of the participants filled out additional questionnaires but those results are not reported here.

Measures

Beads task

To measure probabilistic decision-making and subjective volatility, a modified version of the beads task was administered (see Fig 1). Two virtual bags were displayed on screen, containing 80 black and 20 white beads and the converse. Five sequences of 20 beads each were presented to the participants. At the beginning of each sequence, one of the two bags was chosen at random (p = 0.5). Each sequence was then generated based on the probabilities for the different colors to be drawn (p = 0.2 and p = 0.8) and a fixed probability for the bag of origin

to change (v = 0.04 for each bead, amounting to a ca. 50% chance to observe a bag change in one sequence). This change probability introduced volatility to the task. Participants were informed about this by written instructions stating: "The chance for the bags to change is small enough that in ca. half of the sequences all 20 beads are coming from the same bag and in ca. half of the sequences the bag of origin changes." During the instruction, the experimenter emphasized the probabilistic nature of this description and explained that more or fewer bag changes are possible. To support understanding, five practice trials (i.e. five sampled beads) were completed before the main task.

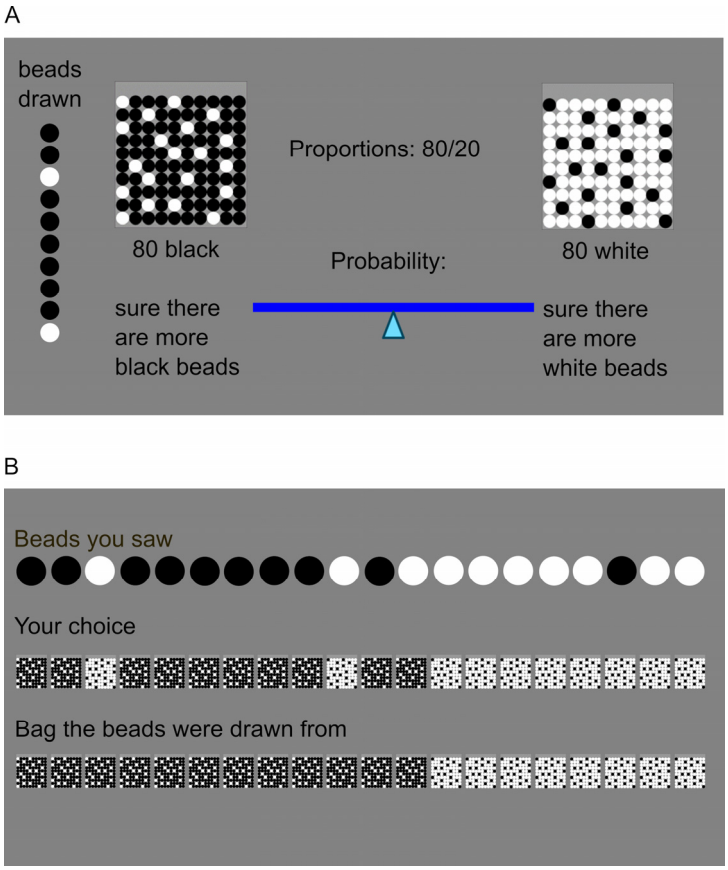


Fig 1. Schematic representation of the beads task. (A) Example of the 10th trial of one sequence. Two bags are displayed which contain either 80 black and 20 white beads, or the converse. Beads are drawn sequentially with replacement. Each of the five sequences consists of 20 drawn beads and the result of each draw, i.e. the color of the bead, remains displayed on the left side of the screen. Within 10 seconds, participants have to indicate their certainty about the bag of origin. They do so by dragging the marker on a visual scale either to the left or the right side. This slider is reset to the center after every trial. (B) At the end of each sequence, feedback about beads seen, choices made and the actual bag of origin of all beads is provided.

After each bead, participants had 10 seconds to indicate their certainty about the bag of origin. They did so by dragging the marker on a visual analogue scale ranging from 0 = "absolutely sure it comes from the bag with more black beads", to 1 = "absolutely sure it comes from the bag with more white beads" where 101 different steps on the scale were mapped to probabilities. After each sequence, participants received a visual feedback about the beads they had seen, the bags they had chosen, and the true bag of origin for each bead (see Fig 1). This feedback thus provided a demonstration of how the instructed probabilities could manifest in color changes.

During the task, the instructions, the two bags, and the currently drawn sequence remained on screen. Initial certainty was measured as the average of all indicated probabilities for the first bead of each sequence. Higher values indicate a more JTC-like behavior [14]. Disconfirmatory belief updating was measured as the change in probability rating in favor of a given color whenever this color differed from the color of the two or more preceding beads. The total size of changes was first averaged across the number of occurrences of such events per sequence, and subsequently across sequences. Here, higher values reflect the formerly described over-adjustment behavior [20]. Participants' perception of the probability for the bags to change, i.e. subjective volatility, was derived from the probabilities participants indicated for each trial n out of $N = 20$ in each sequence k out of $K = 5$. In an ideal Bayesian model, those probabilities should be based on all observed draws until the current trial n, as well as the assumed volatility v. A participant's probability rating $\tilde{p}_{k,n}$ should consequently be their guess of the theoretical probability $P\left(x_{k,n}\middle|z_{k,1}, \dots, z_{k,n}, v\right)$, where x is the bag of origin ($x_{k,n}$ = 0 if bag A, $x_{k,n}$ = 1 if bag B) and z is the color of the drawn bead ($z_{k,n}$ = 0 if white, $z_{k,n}$ = 1 if black), with n and k denoting current number of trial and sequence, respectively, and v the probability for a bag change to occur. Volatility v was estimated by finding the parameter value

that would minimize the difference between the set of theoretical probabilities and the participant's estimated probabilities $\tilde{p}_{k,n}$ (in the least-squares sense). Correlations between observed and predicted probabilities of this 'volatility model' indicated model fit and were moderate to high for the majority of participants, but close to zero for three of them ($n_{ASD} = 1$, $n_{SCZ} = 2$; see Fig 2). Across the sample, model fit correlated negatively with estimated volatility (ρ = -.70, $p < .001$), indicating that weaker model fit was associated with higher volatility estimates. For details on the calculation of the theoretical probabilities and parameter estimation, see model description in S1 File. Note that due to the probabilistic nature of the task, the sequences displayed differed between participants. This has the benefit that any observed effects on the group level are independent of the particular sequence chosen. In contrast, administering the same fixed sequence to all participants might introduce particular sequence-dependent biases, which hinders the generalization of any potential results.

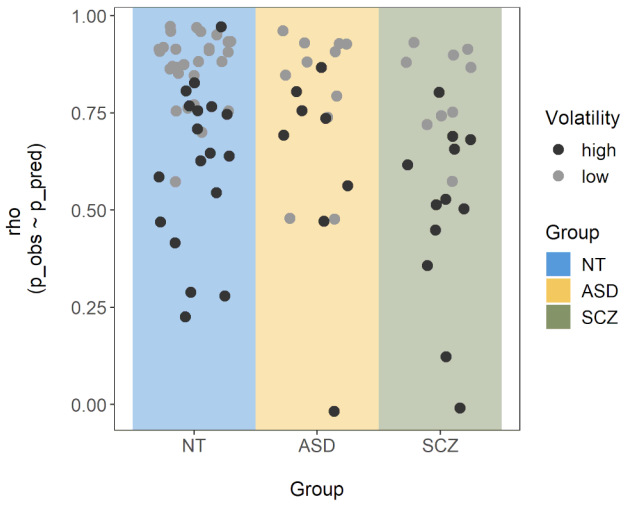


Fig 2. Correlations between observed and predicted probabilities across the task. For each participant, a Spearman correlation was calculated between the participant's subjective probability ratings and those predicted by the volatility model with the best fitting volatility parameter. Single points represent the corresponding correlation coefficient (ρ) for each participant and are colored by size of the corresponding volatility estimate (with clusters of high and low volatility based on a bimodality analysis reported in the results section).

Visual working memory task

A visual working memory task developed based on previously published paradigms [42-43] and a variant of the paradigm used by ten Velden Hegelstad and colleagues [44], was administered to measure both visual working memory and implicit metamemory as a proxy for metacognition. An implicit measure was chosen since uncertainty may be encoded without awareness and not accessible to explicit reports [45]. Working memory accuracy was included as a measure to test whether it was related to uncertainty estimation overall and to control for potential differences in cognitive capacity when interpreting group differences on the beads task variables.

A target shape was presented for one second and then had to be selected from an array of similar shapes (see Fig 3). In this array, thirty shapes that varied along continuous quantitative dimensions were displayed in a circular arrangement corresponding to their continuous modification, i.e. shorter angular distance on the circle meant higher resemblance. The shapes were generated by drawing lines in a polar plot using the following formula:

$$\begin{aligned}\mathrm{amplitude}(\mathrm{phase}) &= 10 + \mathrm{amplitude2} * \cos(\mathrm{frequency2} * \mathrm{phase} + \mathrm{shape}) \\ &+ \mathrm{amplitude1} * (\sin(\mathrm{phase}) + 1) * \sin(\mathrm{frequency1} * \mathrm{phase} + \mathrm{phase1})\end{aligned}$$

where "phase" describes the angle relative to the reference direction (upwards) and "amplitude" the length (radius) of the vector. By varying the "shape" parameter in steps of 12 from 0 to 348 degrees, 30 continuously modified shapes were generated (see S3 Fig for a full display of all shapes). The 30 shapes to choose from remained the same across trials and participants, whereas the target shape was selected randomly for each participant on each trial out of the pool of these 30 shapes.

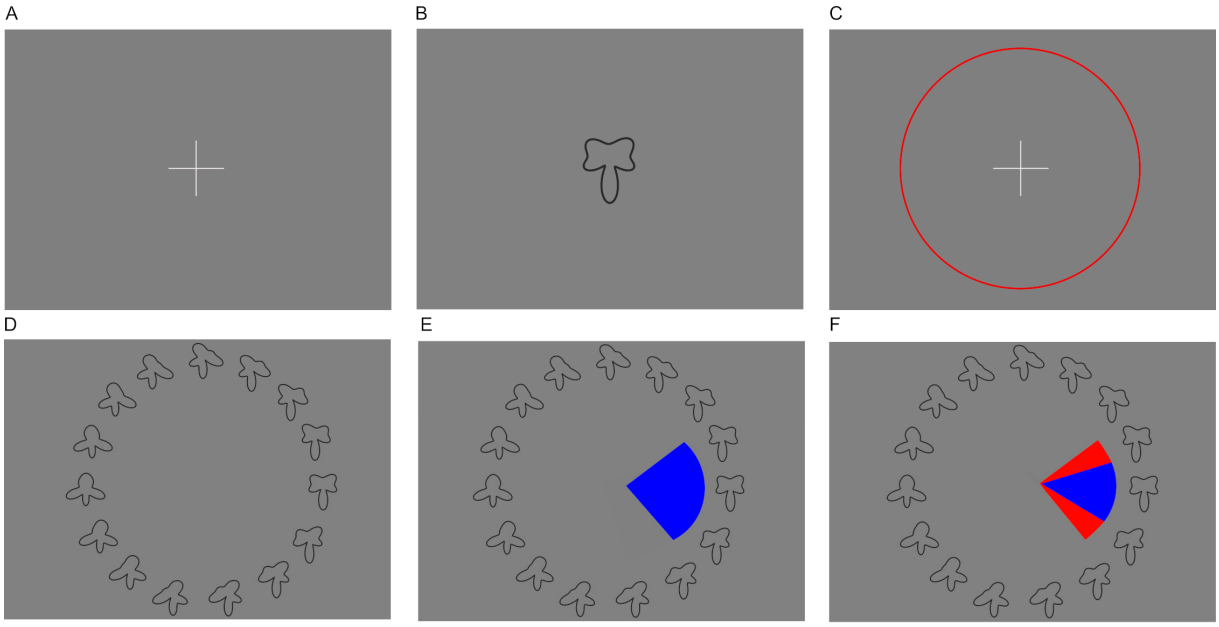


Fig 3. Schematic representation of the visual working memory task. (A) A trial starts with the display of a fixation cross and participants initiate the presentation of the sample shape by clicking on it. (B) The sample shape is then presented for one second, followed by a fixation cross (C). Clicking on it initiates the recall phase (D) in which 30 shapes are presented in a circular arrangement (note that in the example above, only 15 of the 30 shapes, enlarged, are shown for better visibility). (E) The participants now click onto the shape that most resembles what they remember, and set a capture area surrounding it. (F) They receive feedback by being shown the same shape as during the sample phase, correctly placed in the array of shapes. If it is included in the capture area they selected (as in the example above), the excessive part of that capture area is highlighted in red.

After selection of the target from the array of shapes, participants also set a capture area reflecting their uncertainty about how accurately they had selected the correct shape. They were instructed to set this area big enough to be sure the target shape was included but not bigger than necessary. To demotivate them from capturing the whole circle array at all times, they were rewarded with eight points when their capture area included the target shape, and punished by point subtraction proportional to the size of overshoot when making it too large, though points did not translate to any real reward after task completion. Visual feedback was provided after each trial (see Fig 3). Three practice trials and 30 test trials were administered. Participants had the option to skip trials if they had completely forgotten the sample shape. Trials with extremely large (>350 degrees) or small (<4 degrees) capture areas were excluded from analysis

as they might indicate trials where participants accidentally failed to use the option to skip a trial and tried to adjust for that by not setting an appropriate capture area. Visual working memory accuracy was measured as average error, i.e. the average angular distance of the selected shape from the target shape over all trials, with lower values reflecting higher accuracy. Implicit metamemory was assessed by the proportion of all trials where the capture area included the target shape ('hits'), with lower values indicating overestimation of actual accuracy.

Procedure

On the day of the assessment, participants were briefed regarding the background of the study and signed the consent form. They then first completed the visual working memory task, followed by the beads task. Duration of each task was ca. 15 minutes, depending on participants' speed of responding. A short break was introduced between both tasks if required. Task order was not counter-balanced and due to the low similarity and short length of both tasks, no carry-over effects were expected. Demographics were collected on a paper sheet.

Analysis

One-way ANOVAs were conducted and their residuals tested for normality. Only disconfirmatory belief updating and estimated volatility violated the normality assumption. While disconfirmatory belief updating was log-transformed, estimated volatility followed a bimodal distribution and could not be transformed. Results were therefore verified with Kruskal-Wallis tests and for volatility, exploratory analyses using Gaussian Mixture models were conducted. Significant ANOVA F-Tests were followed by Tukey's Honest Significant Difference tests and effect sizes are reported as η^2. Significant Kruskal-Wallis tests were followed by Bonferroni corrected Dunn's Tests and effect sizes are reported as ε^2. Age and sex

did not differ significantly between the groups, and were not controlled for but see analyses in S2 File for group comparisons after propensity matching for both. Level of education differed significantly between the groups but could not be controlled for independently of the diagnosis as the lowest level included more than half of all patients with SCZ but only one participant from the ASD and the NT group. To gauge whether it could be a confounder for any group differences on the task related variables, significant results were followed up by education level comparisons within the NT group. Spearman correlations were chosen to investigate the relationship between the variables of interest across the whole sample. All confirmatory testing was conducted with a significance level of 0.05, one-sided where specified, using the R programming language (R version 3.5.1 [46]).

Results

Demographic variables are summarized in Table 1.

Table 1. Sample demographics per group (total sample size = 86).

	ASD (*n* = 19)			SCZ (*n* = 21)			NT (*n* = 46)			
	n	*M* (*SD*)	*Md* (*IQR*)	*n*	*M* (*SD*)	Md (*IQR*)	*n*	*M* (*SD*)	*Md* (*IQR*)	*p*
Sex (m/f)	11/8			17/4			25/21			.11
Education ("1"/"2"/"3")	1/8/10			12/3/2[a]			1/13/32			<.001
Antipsychotic medication										
Amisulpride				1						
Aripiprazol				5[b]						
Clozapine				2						
Olanzapine				6						
Quetiapine				2						
Risperidone				1						
None				4						
Age		30.32 (8.85)	26 (12)		25.67 (4.74)	26 (7)		28.41 (7.64)	25 (9.75)	.14

Sample sizes (*n*), counts, means (*M*; with standard deviations *SD*) and medians (*Md*; with inter-quartile ranges *IQR*) are displayed. Education was recorded in Norwegian school system categories corresponding to completion of 1 = secondary school (up to age 16), 2 = 6th form college (up to age 19), 3 = higher education (Bachelor, Master, PhD); p-values for group comparisons are provided only for the demographical variables sex and education (Chi-squared tests) as well as Age (ANOVA).

[a] missing data from 4 patients

[b] thereof two with additional Quetiapine treatment

In the beads task, sequences of beads were drawn randomly for each participant, but group comparisons indicated that on average, all groups experienced approximately the same amount of color changes per sequence, $F(2,83) = 2.53$, $\eta^2 = 0.06$, $p = .09$, with $M_{ASD} = 6.38$, $M_{SCZ} = 6.30$, and $M_{NT} = 5.79$ [nonparametric analysis: $\chi^2(2) = 4.88$, $\varepsilon^2 = 0.06$, $p = .09$]. Similarly, the average number of (hidden) bag changes per sequence did not differ by group, $F(2,83) = 1.19$, $\eta^2 = 0.03$, $p = .31$, with $M_{ASD} = 0.65$, $M_{SCZ} = 0.75$, and $M_{NT} = 0.78$ [$\chi^2(2) = 2.22$, $\varepsilon^2 = 0.03$, $p = .33$].

Behaviorally, there were no significant group differences in any of the beads task variables: initial certainty, $F(2,83) = 0.09$, $\eta^2 < 0.01$, $p = .91$ [$\chi^2(2) = 0.04$, $\varepsilon^2 < 0.001$, $p = .98$] (see Fig 4A); estimated volatility, $F(2,83) = 1.92$, $\eta^2 = 0.04$, $p = .15$ [$\chi^2(2) = 3.30$, $\varepsilon^2 = 0.04$, $p = .19$] (see Fig 4C); and log transformed disconfirmatory belief updating, $F(2,83) = 1.24$, $\eta^2 = 0.03$, $p = .30$ [not log transformed for the non-parametric test: $\chi^2(2) = 3.16$, $\varepsilon^2 = 0.04$, $p = .21$ (see Fig 4B)]. Average volatility estimates were higher than the instructed value of 0.04 in all groups (see Table 2). Three one-sided one-sample Wilcoxon signed-rank tests confirmed that this was significant for the ASD ($Md = 0.14$, $V = 167$, $p < .01$), the SCZ ($Md = 0.40$, $V = 216$, $p < .001$), and the NT ($Md = 0.11$, $V = 918$, $p < .001$) group. Model fit (correlations between predicted and observed probabilities, see Fig 2) differed significantly between groups, $\chi^2(2) = 6.70$, $\varepsilon^2 = 0.08$, $p = .04$, with $Md_{ASD} = 0.79$, $Md_{SCZ} = 0.68$, and $Md_{NT} = 0.84$. Post-hoc comparisons revealed a significant difference between the NT and the SCZ group, $z = 2.59$, $p_{adj} = .03$, but not between the ASD and the NT, $z = -0.78$, $p_{adj} > .99$, or the ASD and the SCZ group, $z = 1.48$, $p_{adj} = .42$. To control for potential learning effects over the course of the task, volatility was additionally estimated separately for the first two and the last two sequences of beads. This revealed a slight decrease in volatility towards the end of the task, possibly related to learning

effects in response to the visually provided feedback. However, this volatility change did not differ between groups (see S4 File for details).

Table 2. Descriptive summary statistics of the two tasks per group, effect size and p-value from the conducted ANOVAs (total sample size = 86).

	ASD (n = 19)		SCZ (n = 21)		NT (n = 46)			
	M (*SD*)	*Md* (*IQR*)	*M* (*SD*)	*Md* (*IQR*)	*M* (*SD*)	*Md* (*IQR*)	η^2	*p*
Beads task	0.73 (0.16)	0.70 (0.25)	0.72 (0.13)	0.71 (0.17)	0.71 (0.12)	0.70 (0.18)	<0.01	.90
initial certainty								
disconfirmatory belief updating[a]	0.26 (0.20)	0.17 (0.24)	0.33 (0.19)	0.28 (0.18)	0.28 (0.19)	0.21 (0.17)	0.03[b]	.42
estimated volatility	0.30 (0.30)	0.14 (0.56)	0.37 (0.31)	0.40 (0.56)	0.23 (0.24)	0.11 (0.41)	0.04	.16
VWM task	0.57 (0.16)	0.54 (0.23)	0.52 (0.13)	0.50 (0.15)	0.58 (0.12)	0.57 (0.14)	0.04	.18
proportion hits								
error	28.23 (11.07)	26.31 (16.84)	35.72 (11.58)	35.67 (12.77)	24.93 (9.18)	23.88 (10.65)	0.15	<.001

M = mean, *SD* = standard deviation, *Md* = median, *IQR* = interquartile range, VWM = visual working memory

[a] descriptive data not log-transformed but based on original scale

[b] effect size based on log transformed data

In the visual working memory task, a non-parametric group comparison of number of skipped trials revealed no significant group differences, $\chi^2(2) = 1.99$, $\varepsilon^2 = 0.02$, $p = .37$, with $Md_{ASD} = 0.00$, $Md_{SCZ} = 0.00$, and $Md_{NT} = 0.00$. A similar comparison for number of trials where the capture area was out of range (i.e. <4 or >350 degrees), also demonstrated no significant differences between groups, $\chi^2(2) = 0.59$, $\varepsilon^2 = 0.01$, $p = .75$, with $Md_{ASD} = 0.00$, $Md_{SCZ} = 1.00$, and $Md_{NT} = 0.00$. There was a significant effect of group on average error (i.e. memory inaccuracy), $F(2,83) = 8.03$, $\eta^2 = 0.16$, $p < .001$ [$\chi^2(2) = 12.91$, $\varepsilon^2 = 0.15$, $p < .01$] (see Table 2 and Fig 4D). Post-hoc comparisons revealed that the average error in the SCZ group ($M = 35.72$, $SD = 11.58$) was significantly larger than in the NT group ($M = 24.93$, $SD = 9.18$), $p_{adj} < .001$ [nonparametric analysis: $z = -3.59$, $p_{adj} < .001$], and numerically but not significantly larger compared to the ASD group ($M = 28.23$, $SD = 11.07$), $p_{adj} = .06$ [$z = -1.97$, $p_{adj} = .15$]. ASD and the NT group did not differ, $p_{adj} = .47$ [$z = 1.18$, $p_{adj} = .71$]. Within the NT group, level of

education was unrelated to memory inaccuracy, $F(2,43) = 1.39$, $\eta^2 = 0.06$, $p = .26$ [$\chi^2(2) = 2.11$, $\varepsilon^2 = 0.05$, $p = .35$]. For proportion of hits (i.e. metamemory), no significant group differences were found, $F(2,83) = 1.73$, $\eta^2 = 0.04$, $p = .18$ [$\chi^2(2) = 3.29$, $\varepsilon^2 = 0.04$, $p = .19$] (see Fig 4E). To control for potential effects of response times in the visual working memory task, additional analyses were conducted. These revealed that the SCZ group responded faster on average, but that independent of group membership longer response times were associated with larger errors (see S5 File for details).

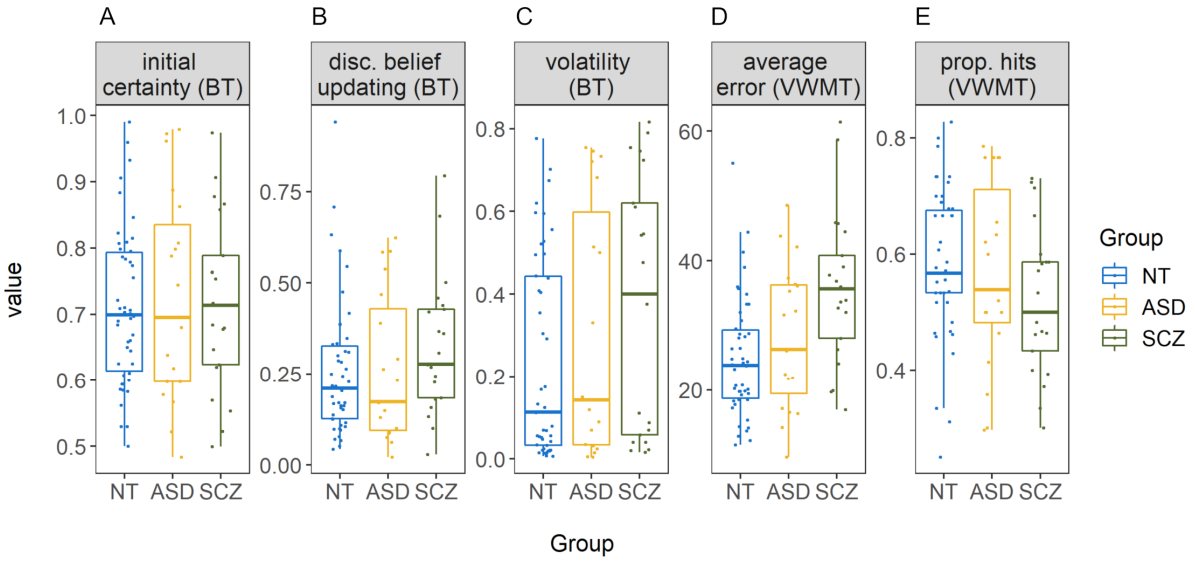


Fig 4. Boxplots per group for all main variables. Beads task (BT) variables are initial certainty (A), untransformed disconfirmatory belief updating (B) and estimated volatility (C). Variables from the visual working memory task (VWMT) include error (D) and proportion of hits (E). NT = neurotypically developing individuals, ASD = individuals with autism spectrum disorder, SCZ = individuals with schizophrenia. All BT variables are expressed as probabilities, average error is expressed in degrees, and proportion of hits is the proportion of trials where the capture area included the target. Points represent single participants.

Disconfirmatory belief updating correlated with both initial certainty ($\rho = .48$, $p < .001$) and estimated volatility ($\rho = .62$, $p < .001$, see Fig 5), but there was no significant relationship between initial certainty and estimated volatility ($\rho = .09$, $p = .39$). There was a strong correlation between proportion of hits (metamemory) and average error (memory inaccuracy)

as measured by the visual working memory task (ρ = -.59, p < .001). Across tasks, average error was positively correlated with disconfirmatory belief updating (ρ = .33, p < .01) and with estimated volatility (ρ = .40, p < .001), but not initial certainty (ρ = .04, p = .68). Proportion of hits was not related to initial certainty (ρ = .12, p = .28) or disconfirmatory belief updating (ρ = -.09, p = .42) but was negatively associated with estimated volatility (ρ = -.24, p = .03). Thus, participants with better metamemory as measured by the visual working memory task also tended to estimate the volatility within the beads task more appropriately, with lower values approaching the true volatility that was introduced by the task design.

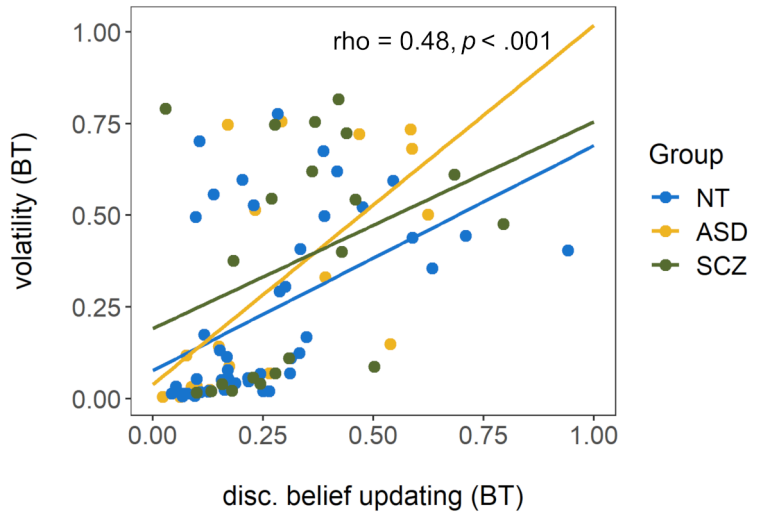


Fig 5. Scatterplot of disconfirmatory belief updating (untransformed) and estimated volatility from the beads task (BT). Rho and p display the results of a Spearman correlation conducted across the total sample. Regression lines are fitted for each group for illustrative purposes only. NT = neurotypically developing individuals, ASD = individuals with autism spectrum disorder, SCZ = individuals with schizophrenia.

As visible in Fig 4C, estimated volatility ν followed a bimodal distribution, suggesting one high- and one low-volatility cluster. This structure may have masked potential between-group effects in traditional and non-parametric tests. In an exploratory approach, the bimodality of this variable was therefore modeled using Gaussian mixture models in conjunction with a Bayesian estimation method. That approach allowed for the extraction of posterior probability

distributions to find the most likely values of the estimated coefficients given the data [47]. The model can be written as

$$p(y|\mu_1, \mu_2, \sigma_1, \sigma_2, \theta) = \theta \text{Normal}(y|\mu_1, \sigma_1) + (1 - \sigma)\text{Normal}(y|\mu_2, \sigma_2),$$

where (μ_1, σ_1) and (μ_2, σ_2) are the parameters of the first and second cluster, respectively, and θ is the mixing proportion indicating the relative proportion of subjects who belonged to the first vs. the second cluster. Volatility values (formerly ν) are labeled as y to emphasize the fact that they are treated as data in this estimation context. Prior distributions were specified to be weakly informative [48] with the standard-deviations

$$\sigma \sim \text{LogNormal}(0,0.1),$$

and the means

$$\mu \sim \text{Normal}(0,1).$$

Hamiltonian Monte-Carlo (HMC) methods were applied and implemented in the Stan software [49] using the RStan interface [50]. All models were fitted using four independent chains with 2000 iterations per chain where the first 1000 steps were discarded as warm-up samples. The Gelman-Rubin diagnostic $\hat{R}$ [51] was used to ensure convergence and all $\hat{R} < 1.01$. Results are reported in terms of the posterior mean value and the 95% highest-density intervals (HDI) which cover the area in which the true parameter value is located with probability 95% given the model structure. In order to detect group-level effects, parameters μ_1, μ_2 and θ were modeled separately per group and the resulting models were compared using leave-one-out cross-validation (LOOCV [52]). Concretely, a sequence of models of increasing complexity was designed and the leave-one-out information criterion (LOOIC) was calculated for each (see Table 3). This criterion can be interpreted similarly as the AIC and BIC criteria (lower values indicate better fit) but is appropriate for Bayesian models.

Table 3. Model comparison.

Rank	Free variables between groups	LOOIC	SE(LOOIC)	ΔLOOIC	SE(ΔLOOIC)
1	μ_2	-70.20	17.05	–	–
2	none	-67.81	17.10	2.39	2.91
3	μ_1 and μ_2	-66.86	16.59	3.34	1.37
4	μ_1,μ_2 and θ	-65.98	16.49	4.22	1.72
5	μ_2 and θ	-36.46	13.35	33.74	4.48

As can be seen in Table 3, the model which allowed the mean of the high-volatility cluster (μ_2) to vary between groups performed best in comparison to the baseline-model in which no group-differences were modeled. The model successfully identified two separate clusters, one that was very close to the optimal volatility value of $\nu_{\text{optimal}} = 0.04$ with a cluster mean of $\mu_1 = 0.05$, HDI = [0.04,0.07] and small variance ($\sigma_1 = 0.04$, HDI = [0.03,0.06]) and one that was centered at $\mu_2 = 0.51$, HDI = [0.43,0.59] ($\sigma_2 = 0.17$, HDI = [0.12,0.22]) reflecting well the bimodal nature of the distribution. Further, the size of the two clusters was very similar with approximately 53% of the subjects belonging to the first (close-to-optimal) cluster, $\theta = 0.53$, HDI = [0.42,0.65]. Model-fit was excellent as determined by the posterior predictive distributions for all groups in Fig 6A - 6C.

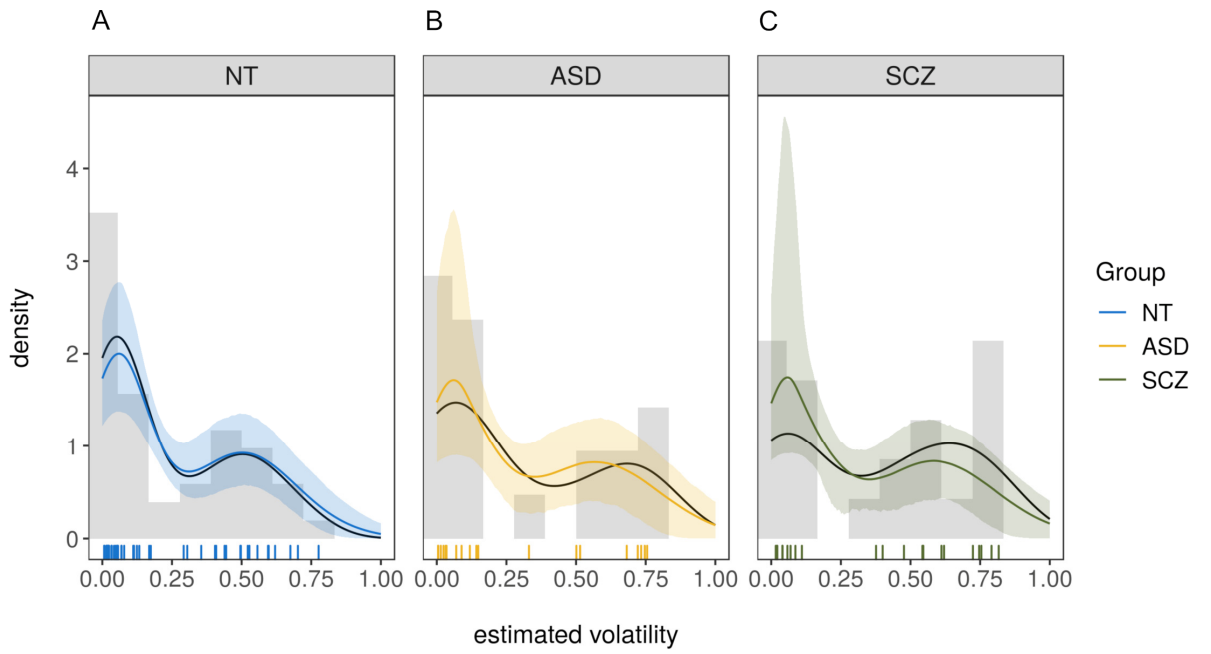


Fig 6. Posterior-predictive distributions of the winning model for all three groups. (A) for neurotypically developing individuals (NT), (B) for individuals with autism spectrum disorder (ASD), (C) for individuals with schizophrenia (SCZ). Colored lines are posterior means of the posterior predictive distributions, shaded areas are the 5% and the 95% percentile. Black lines are the actual data. Vertical lines are the estimated volatility values for each participant based on the Bayesian volatility model as described in section 'Measures: Beads task'. NT = neurotypically developing individuals, ASD = individuals with autism spectrum disorder, SCZ = individuals with schizophrenia.

Both the ASD and the SCZ group had a slightly elevated mean in the high-volatility cluster. For the ASD group, the effect was $b_{\mathrm{ASD}} = 0.07$, HDI = [-0.05,0.19] with a probability of a truly higher volatility in this cluster compared to the NT group of 89 %. For the SCZ group, the effect was $b_{\mathrm{SCZ}} = 0.09$, HDI = [-0.01,0.20], with a probability of a truly higher volatility than the NT group of 95 %. To check for the effect of education, a median split on estimated volatility was conducted. Within the NT group, level of education was unrelated to volatility ratings being above or below and equal to the median, $\chi^2(2) = 0.04$, $p = .98$. Further, volatility estimates within the median-split groups did not differ by education (below and equal to median: $F(1,23) = 2.94$, $\eta^2 = 0.11$, $p = .10$ [$\chi^2(1) = 2.95$, $\varepsilon^2 = 0.12$, $p = .09$]; above: $F(2,18) = 0.04$, $\eta^2 < 0.01$, $p = .96$ [$\chi^2(2) = 0.54$, $\varepsilon^2 = 0.03$, $p = .76$]).

Discussion

This study investigated probabilistic decision-making and visual (meta-)memory in persons with schizophrenia (SCZ group), persons with high-functioning autism (ASD group) and neurotypically developing individuals without any psychiatric diagnosis (NT group) to explore if and to what extent groups differed in processing of probabilistic information and subsequent estimation of uncertainty. Unexpectedly, none of the groups differed significantly on any of the probabilistic reasoning measures. Relative to NT individuals, neither participants with SCZ nor persons with ASD showed significantly higher or lower certainty when making their first probability rating, when integrating new evidence with previous beliefs or when interpreting the volatility of the task environment. Similarly, none of the groups differed regarding their (un)certainty about their own visual memory performance (metamemory). However, participants with SCZ showed lower visual working memory accuracy than participants of the NT group.

While the absence of a difference between ASD and SCZ group in subjectively perceived volatility is not unexpected in light of the literature that found overestimation of volatility in both groups (e.g. [26, 28]), it is surprising that neither clinical group differed from the NT group. However, additional analyses revealed two clusters of participants: those who estimated volatility in a near-optimal manner and those who strongly overestimated volatility. Within the second cluster, volatility was higher in individuals with ASD and SCZ compared to the NT group, confirming in part the aforementioned findings of volatility overestimation in those clinical groups. The bimodal distribution itself might indicate qualitatively different processing modes [53]. Such processing modes could be related to the different decision-making strategies proposed in the reinforcement learning literature: a model-based mode, which relies on a cognitive representation of state transitions and a complex model of the task overall, and a model-free mode, which is more habitual and driven by trial-and-error feedback [54].

Participants in the low volatility cluster might be more prone to model-based strategies, whereas participants in the high volatility cluster may be more sensitive to trial-wise fluctuations of colors. While both modes are in theory available to all individuals, the choice of one strategy over the other can vary depending on the task at hand and available cognitive resources (e.g. [55]). Notably, volatility was higher for persons with lower working memory accuracy in the current study.

Nevertheless, the absence of overall group differences in the main analyses seems at odds with studies reporting a general overestimation of volatility in individuals with ASD or SCZ. Crucially, most previous studies did not inform their participants about the actual size of the change probability. Instead, it had to be inferred from exposure to the learning environment (e.g. [29, 31]). In contrast, the current study attempted to induce the same prior belief in all groups by providing explicit instructions about the degree of the task environment's volatility. While one possible explanation of the bimodal volatility distribution is the aforementioned choice of processing mode, another explanation may be individual differences in understanding of the instructions. It is possible that individuals in the high volatility cluster misunderstood the instructions and assumed bags would change with a probability of 0.5 per bead rather than per sequence. Interestingly, miscomprehension of task instructions has been suggested as an explanation for the Jumping-to-Conclusions (JTC) bias in other versions of the beads task [15]. Similarly, misunderstanding of probabilities has been found to explain the JTC bias, possibly caused by reduced general cognitive abilities [56]. In order to clarify the effect of explicit information about volatility on behavior, future studies should contrast conditions where volatility is explicitly announced against conditions where it is not. Further, the role of working memory and other cognitive ability measures in this context should be elaborated, as they may link to the understanding of probabilities and (mis)comprehension of task instructions.

Importantly, while the volatility-estimating model fitted the data of the majority of participants well, model fit was significantly weaker for the SCZ group. Furthermore, weaker model fit was associated with increased volatility estimates across the sample. This may reflect the aforementioned deviation from task instructions or a different choice of processing mode in participants with high volatility values, causing an increased deviation from the behavior the theoretical model would predict. Nevertheless, the estimated volatility values were still those that fitted the observed behavior best, even if not perfectly, and model fit was still reasonable for the majority of participants with high volatility estimates (see Fig 2).

The absence of differences in other, directly observable JTC related variables was surprising. While it was unclear what to expect for participants with ASD given the few and contradictory findings (see [17, 18]), over-adjustment in response to disconfirmatory evidence has been reported for patients with SCZ [20]. This inconsistency with previous findings may in part be related to the choice of method. For SCZ, group differences seem to be less consistent in graded estimates versions of the beads task [16]. Further, the explicit introduction of volatility in the current study may have contributed to the absence of group differences. Similar beliefs about the task's volatility across groups could cause similar belief updating, as over-adjustment (i.e. increased disconfirmatory belief updating) is likely related to overestimation of volatility: In an environment that is constantly changing, the newest observations seem most reliable and therefore deserve greater attention. This interpretation is supported by the positive correlation between disconfirmatory belief updating and estimated volatility. Hence, introducing volatility explicitly in the current study may have eliminated the difference between persons who typically overestimate volatility (persons with ASD and SCZ) and those that do not (NT group). Importantly, the volatility parameter of the model used in this study is estimated based on all trial-wise deviations of participants' probabilistic estimates from an ideal Bayesian observer. The model rests on the assumption that these deviations are mainly caused by a misestimation

of the true volatility. Yet, other causes for such deviations are conceivable, even if unlikely. As such, estimated volatility might be affected by "noisy" decision-making (see S4 File for additional analyses that address this question). Nonetheless, the positive correlation with disconfirmatory belief updating seems to substantiate the idea that estimated volatility reflects at least in part a belief about the probability for the bags to change, i.e. subjectively perceived volatility of the environment.

The lack of group differences in metamemory could be the result of measuring it implicitly as opposed to former studies that used explicit self-reports (e.g. [36, 37]). It has been suggested that implicit metacognition relies on a different cognitive system than explicit metacognition and is only minimally dependent on working memory [57]. These findings are also in line with recent reports of intact implicit metacognition in SCZ [44] and metacognitive efficiency in first episode psychosis [58]. Interestingly, metamemory was negatively related to estimated volatility. This suggests, that both misestimation of subjective cognitive capacity and overestimation of environmental uncertainty (such as volatility) might be affected by similar mechanisms, potentially driven by higher-level uncertainty calculations in the belief hierarchy of the human mind, and is in line with the conceptualization of aberrant representation of uncertainties as a 'failure of metacognition' [24]. Notably, average metamemory scores were rather low, with proportion of hits of 50% to ca. 60% for each group. On the one hand, this might indicate an overall tendency of participants to overestimate the accuracy with which they had remembered and correctly identified the target shape. On the other, this may in part be due to difficulties in perceptually differentiating between the shape stimuli overall, suggesting that the task in that regard might have been slightly too demanding.

The finding of lower visual working memory accuracy only for participants with SCZ relative to the NT group was little surprising. Working memory deficits are well established in SCZ (e.g. [59, 60]) but not in high-functioning autism, where findings are less consistent and

performance, particularly in the visual domain, is often unimpaired (e.g. [61, 62]). Lower visual working memory accuracy was related to disconfirmatory belief updating across the whole sample. This fits well with findings that linked the JTC bias to memory performance [40, 59].

Limitations of the current study include the rather small sample sizes for the ASD and the SCZ group. The power of this study might have been too low to detect actual group differences in some of the measures. This is particularly the case for estimated volatility, where descriptive statistics and additional modelling suggest higher values in parts of the SCZ and the ASD group. The study would further have profited from the inclusion of additional cognitive ability measures. It remains unclear to what extent differences in cognitive ability may have attributed to differences in probability estimation and task comprehension. This similarly concerns the findings for visual working memory and JTC, both of which have been linked to general cognitive ability [63, 64]. While possibly related, educational degree was not controlled for in the analyses, as differences in educational levels were so large, that their effects could not be assessed independently of clinical diagnoses. However, within the NT group, education was unrelated to the main variables of interest, though it is noteworthy that the lowest educational level was underrepresented in this group. Groups were not matched by education prior to data analysis as this has been criticized for possibly leading to the selection of an atypical, high-achieving SCZ sample [65].

Further, this comparative approach was purely diagnosis-based and there was no differentiation between patients by symptoms. However, recent studies have not found any correlations between severity of psychopathological symptoms and volatility estimation [32] or aberrant switching behavior [28, 30] in SCZ. For ASD, the relationship is less clear with some studies reporting no relationship between ASD-typical symptoms and volatility-related behavior [66], some suggesting a relationship with few of the behavioral variables [25], and some not investigating any correlations along those lines [12, 27]. It is unclear whether linear

relationships should even be expected in a cross-sectional design as some of the symptoms (e.g. delusions in SCZ, rigid behavior in ASD) may constitute a secondary coping mechanism in response to prior volatility overestimation [21, 22]. Regarding the often investigated relationship between JTC like behavior and delusions, results are similarly inconsistent [16, 67], but point towards an absence of this relationship for certainty and responses to contradictory evidence [67]. Furthermore, type or dose of medication were not controlled for in the current study. Antipsychotic medication might worsen or improve cognitive capacity. However, some of the study's main variables were similar to those investigated in the JTC bias literature and previous findings actually indicated that JTC is not influenced by antipsychotic drugs (e.g. [68, 69]). Finally, the SCZ group was recruited amongst the most severely ill patients (inpatient care) and a majority were males. It is therefore unclear how well the results can be generalized.

To summarize, this study demonstrates reduced visual working memory accuracy of SCZ patients compared to NT controls. Further, the findings did not reveal any group differences for metamemory but suggest higher overestimation of volatility among some participants with autism and schizophrenia. This partially supports the conceptualization of uncertainty misestimation based approaches to phenomenology of these conditions. Nevertheless, despite similarities in social and non-social cognitive performance, both conditions' symptomatology is heterogeneous in nature and while overlap of some clinical symptoms exists, many of them are rather particular for one of the conditions, respectively (e.g. rigid behavior in ASD, delusions or hallucinations in SCZ). It remains unclear how, if present, similar underlying mechanisms can account for that and future studies should investigate this more closely, linking subjective volatility estimation to clinical symptoms and cognitive ability in a longitudinal design.

Acknowledgements

We thank all participants for their time and effort as well as Kristin Sandvik, Mari Arnesen and Jonas Kymre for assistance with data collection. The publication charges for this article have been funded by a grant from the publication fund of UiT – The Arctic University of Norway.

Additional information

Anonymized raw and processed data as well as the supporting information (S1 – S5) are available in an Open Science Framework repository: DOI 10.17605/OSF.IO/UCA5E.

References

1. Eack SM, Bahorik AL, McKnight SAF, Hogarty SS, Greenwald DP, Newhill CE, et al. Commonalities in social and non-social cognitive impairments in adults with autism spectrum disorder and schizophrenia. Schizophrenia Research. 2013;148(1):24-8. doi: https://doi.org/10.1016/j.schres.2013.05.013.
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. 5 ed. Washington DC: American Psychiatric Association; 2013.
3. Burbach JP, van der Zwaag B. Contact in the genetics of autism and schizophrenia. Trends Neurosci. 2009;32(2):69-72. Epub 2009/01/13. doi: 10.1016/j.tins.2008.11.002. PubMed PMID: 19135727.
4. Craddock N, Owen MJ. The Kraepelinian dichotomy - going, going... but still not gone. The British journal of psychiatry : The journal of mental science. 2010;196(2):92-5. doi: 10.1192/bjp.bp.109.073429. PubMed PMID: 20118450.
5. Pinkham AE, Morrison KE, Penn DL, Harvey PD, Kelsven S, Ludwig K, et al. Comprehensive comparison of social cognitive performance in autism spectrum disorder and schizophrenia. Psychological Medicine. 2019:1-9. Epub 2019/10/02. doi: 10.1017/S0033291719002708.
6. Lysaker PH, Olesek KL, Warman DM, Martin JM, Salzman AK, Nicolò G, et al. Metacognition in schizophrenia: Correlates and stability of deficits in theory of mind and self-reflectivity. Psychiatry Research. 2011;190:18-22.
7. Williams D, Happé F. Representing intentions in self and other: Studies of autism and typical development. Developmental Science. 2010;13(2):307-19. doi: 10.1111/j.1467-7687.2009.00885.x.
8. Sasson N, Tsuchiya N, Hurley R, Couture SM, Penn DL, Adolphs R, et al. Orienting to social stimuli differentiates social cognitive impairment in autism and schizophrenia. Neuropsychologia. 2007;45(11):2580-8. doi: https://doi.org/10.1016/j.neuropsychologia.2007.03.009.
9. Couture SM, Penn DL, Losh M, Adolphs R, Hurley R, Piven J. Comparison of social cognitive functioning in schizophrenia and high functioning autism: More convergence than divergence. Psychol Med. 2010;40(4):569-79. Epub 2009/08/13. doi: 10.1017/S003329170999078X. PubMed PMID: 19671209; PubMed Central PMCID: PMCPMC2827676.
10. Fernandes JM, Cajão R, Lopes R, Jerónimo R, Barahona-Corrêa JB. Social Cognition in Schizophrenia and Autism Spectrum Disorders: A Systematic Review and Meta-Analysis of Direct Comparisons. Frontiers in Psychiatry. 2018;9(504). doi: 10.3389/fpsyt.2018.00504.
11. Brown EC, Hack SM, Gold JM, Carpenter WT, Fischer BA, Prentice KP, et al. Integrating frequency and magnitude information in decision-making in schizophrenia: An account of patient performance on the Iowa Gambling Task. Journal of Psychiatric Research. 2015;66-67:16-23. doi: https://doi.org/10.1016/j.jpsychires.2015.04.007.
12. Mussey JL, Travers BG, Klinger LG, Klinger MR. Decision-Making Skills in ASD: Performance on the Iowa Gambling Task. Autism Research. 2015;8(1):105-14. doi: 10.1002/aur.1429.
13. Raffard S, Capdevielle D, Gely-Nargeot M-C, Attal J, Baillard A, Del-Monte J, et al. Insight is not associated with insensitivity to future consequences in schizophrenia. Psychiatry Research. 2011;187(1):307-9. doi: https://doi.org/10.1016/j.psychres.2010.11.020.
14. Huq SF, Garety PA, Hemsley DR. Probabilistic judgements in deluded and non-deluded subjects. The Quarterly Journal of Experimental Psychology Section A. 1988;40(4):801-12. doi: 10.1080/14640748808402300.

15. Balzan RP, Delfabbro PH, Galletly CA, Woodward TS. Over-adjustment or miscomprehension? A re-examination of the jumping to conclusions bias. Aust N Z J Psychiatry. 2012;46(6):532-40. Epub 2012/06/09. doi: 10.1177/0004867411435291. PubMed PMID: 22679205.
16. Moritz S, Woodward TS. Jumping to conclusions in delusional and non-delusional schizophrenic patients. British Journal of Clinical Psychology. 2005;44(2):193-207. doi: http://dx.doi.org/10.1348/014466505X35678. PubMed PMID: 2005-09366-004.
17. Jansch C, Hare DJ. An investigation of the "jumping to conclusions" data-gathering bias and paranoid thoughts in Asperger syndrome. J Autism Dev Disord. 2014;44(1):111-9. Epub 2013/08/13. doi: 10.1007/s10803-013-1855-2. PubMed PMID: 23933997.
18. Brosnan M, Chapman E, Ashwin C. Adolescents with autism spectrum disorder show a circumspect reasoning bias rather than 'jumping-to-conclusions'. J Autism Dev Disord. 2014;44(3):513-20. Epub 2013/09/05. doi: 10.1007/s10803-013-1897-5. PubMed PMID: 24002414.
19. Zhang L, Tang J, Dong Y, Ji Y, Tao R, Liang Z, et al. Similarities and Differences in Decision-Making Impairments between Autism Spectrum Disorder and Schizophrenia. Frontiers in Behavioral Neuroscience. 2015;9(259). doi: 10.3389/fnbeh.2015.00259.
20. Speechley WJ, Whitman JC, Woodward TS. The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. Journal of Psychiatry & Neuroscience. 2010;35(1):7-17. doi: 10.1503/jpn.090025.
21. Fletcher PC, Frith CD. Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci. 2009;10(1):48-58. Epub 2008/12/04. doi: 10.1038/nrn2536. PubMed PMID: 19050712.
22. Van de Cruys S, Evers K, Van der Hallen R, Van Eylen L, Boets B, de-Wit L, et al. Precise minds in uncertain worlds: Predictive coding in autism. Psychol Rev. 2014;121(4):649-75. Epub 2014/10/28. doi: 10.1037/a0037665. PubMed PMID: 25347312.
23. van Schalkwyk GI, Volkmar FR, Corlett PR. A predictive coding account of psychotic symptoms in autism spectrum disorder. Journal of Autism and Developmental Disorders. 2017;47(5):1323-40. doi: http://dx.doi.org/10.1007/s10803-017-3065-9 28185044. PubMed PMID: 2017-06616-001.
24. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. Front Psychiatry. 2013;4:47. Epub 2013/06/12. doi: 10.3389/fpsyt.2013.00047. PubMed PMID: 23750138; PubMed Central PMCID: PMCPMC3667557.
25. D'Cruz A-M, Ragozzino ME, Mosconi MW, Shrestha S, Cook EH, Sweeney JA. Reduced behavioral flexibility in autism spectrum disorders. Neuropsychology. 2013;27(2):152-60. doi: 10.1037/a0031721.
26. Lawson RP, Mathys C, Rees G. Adults with autism overestimate the volatility of the sensory environment. Nat Neurosci. 2017;20(9):1293-9. Epub 2017/08/02. doi: 10.1038/nn.4615. PubMed PMID: 28758996; PubMed Central PMCID: PMCPMC5578436.
27. Solomon M, Smith AC, Frank MJ, Ly S, Carter CS. Probabilistic reinforcement learning in adults with autism spectrum disorders. Autism Research. 2011;4(2):109-20. doi: 10.1002/aur.177.
28. Culbreth AJ, Gold JM, Cools R, Barch DM. Impaired Activation in Cognitive Control Regions Predicts Reversal Learning in Schizophrenia. Schizophrenia Bulletin. 2016;42(2):484-93. doi: 10.1093/schbul/sbv075.
29. Kaplan CM, Saha D, Molina JL, Hockeimer WD, Postell EM, Apud JA, et al. Estimating changing contexts in schizophrenia. Brain. 2016;139(7):2082-95. doi: 10.1093/brain/aww095.

30. Li CT, Lai WS, Liu CM, Hsu YF. Inferring reward prediction errors in patients with schizophrenia: A dynamic reward task for reinforcement learning. Front Psychol. 2014;5:1282. Epub 2014/11/27. doi: 10.3389/fpsyg.2014.01282. PubMed PMID: 25426091; PubMed Central PMCID: PMCPMC4227479.
31. Schlagenhauf F, Huys QJM, Deserno L, Rapp MA, Beck A, Heinze H-J, et al. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. NeuroImage. 2014;89:171-80. doi: https://doi.org/10.1016/j.neuroimage.2013.11.034.
32. Deserno L, Boehme R, Mathys C, Katthagen T, Kaminski J, Stephan KE, et al. Volatility Estimates Increase Choice Switching and Relate to Prefrontal Activity in Schizophrenia. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2020;5(2):173-83. doi: https://doi.org/10.1016/j.bpsc.2019.10.007.
33. Flavell JH. Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist. 1979;34(10):906-11. doi: http://dx.doi.org/10.1037/0003-066X.34.10.906. PubMed PMID: 1980-09388-001.
34. Norman E, Pfuhl G, Sæle RG, Svartdal F, Låg T, Dahl TI. Metacognition in Psychology. Review of General Psychology. 2019;23(4):403-24. doi: 10.1177/1089268019883821.
35. Lysaker PH, Dimaggio G. Metacognitive capacities for reflection in schizophrenia: Implications for developing treatments. Schizophrenia bulletin. 2014;40(3):487-91. Epub 2014/03/17. doi: 10.1093/schbul/sbu038. PubMed PMID: 24636965.
36. Moritz S, Woodward TS, Jelinek L, Klinge R. Memory and metamemory in schizophrenia: A liberal acceptance account of psychosis. Psychological Medicine. 2008;38:825-32.
37. Grainger C, Williams DM, Lind SE. Metacognition, metamemory, and mindreading in high-functioning adults with autism spectrum disorder. J Abnorm Psychol. 2014;123(3):650-9. Epub 2014/06/24. doi: 10.1037/a0036531. PubMed PMID: 24955572.
38. Wilkinson DA, Best CA, Minshew NJ, Strauss MS. Memory Awareness for Faces in Individuals with Autism. Journal of Autism and Developmental Disorders. 2010;40(11):1371-7. doi: 10.1007/s10803-010-0995-x.
39. Buck KD, Warman DM, Huddy V, Lysaker PH. The Relationship of Metacognition with Jumping to Conclusions among Persons with Schizophrenia Spectrum Disorders. Psychopathology. 2012;45(5):271-5. doi: 10.1159/000330892.
40. Takeda T, Nakataki M, Ohta M, Hamatani S, Matsuura K, Ohmori T. Effect of cognitive function on jumping to conclusion in patients with schizophrenia. Schizophrenia Research: Cognition. 2018;12:50-5. doi: 10.1016/j.scog.2018.04.002. PubMed PMID: PMC6007052.
41. World Health Organization. The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. Geneva: World Health Organization; 1992.
42. Pfuhl G, Barrera LB, Living M, Biegler R. Do adjustments in search behavior depend on the precision of spatial memory? Learn Behav. 2013;41(1):77-93. Epub 2012/08/14. doi: 10.3758/s13420-012-0087-1. PubMed PMID: 22886715.
43. Zhang W, Luck SJ. Discrete fixed-resolution representations in visual working memory. Nature. 2008;453(7192):233-5. Epub 2008/04/04. doi: 10.1038/nature06860. PubMed PMID: 18385672; PubMed Central PMCID: PMCPMC2588137.
44. ten Velden Hegelstad W, Kreis I, Tjelmeland H, Pfuhl G. Psychosis and psychotic-like symptoms affect cognitive abilities but not motivation in a foraging task. Frontiers in Psychology. 2020;11. doi: 10.3389/fpsyg.2020.01632

45. Knoblich G, Stottmeister F, Kircher T. Self-monitoring in patients with schizophrenia. Psychological Medicine. 2004;34(8):1561-9. Epub 2004/11/04. doi: 10.1017/S0033291704002454.
46. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.
47. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis: CRC press; 2013.
48. Chung Y, Gelman A, Rabe-Hesketh S, Liu J, Dorie V. Weakly informative prior for point estimation of covariance matrices in hierarchical models. Journal of Educational and Behavioral Statistics. 2015;40(2):136-57.
49. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A probabilistic programming language. Journal of statistical software. 2017;76(1).
50. Stan Development Team. RStan: The R interface to Stan. R package version 2019. http://mc-stan.org/.
51. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical science. 1992;7(4):457-72.
52. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing. 2017;27(5):1413-32.
53. Freeman JB, Dale R. Assessing bimodality to detect the presence of a dual cognitive process. Behavior research methods. 2013;45(1):83-97.
54. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature neuroscience. 2005;8(12):1704-11.
55. Otto AR, Gershman SJ, Markman AB, Daw ND. The curse of planning: Dissecting multiple reinforcement-learning systems by taxing the central executive. Psychological science. 2013;24(5):751-61.
56. Klein HS, Pinkham AE. Examining reasoning biases in schizophrenia using a modified “Jumping to Conclusions” probabilistic reasoning task. Psychiatry research. 2018;270:180-6.
57. Shea N, Boldt A, Bang D, Yeung N, Heyes C, Frith CD. Supra-personal cognitive control and metacognition. Trends in Cognitive Sciences. 2014;18(4):186-93. doi: https://doi.org/10.1016/j.tics.2014.01.006.
58. Wright A, Nelson B, Fowler D, Greenwood K. Perceptual biases and metacognition and their association with anomalous self experiences in first episode psychosis. Consciousness and Cognition. 2020;77:102847. doi: https://doi.org/10.1016/j.concog.2019.102847.
59. Freeman D, Startup H, Dunn G, Cernis E, Wingham G, Pugh K, et al. Understanding jumping to conclusions in patients with persecutory delusions: Working memory and intolerance of uncertainty. Psychol Med. 2014;44(14):3017-24. Epub 2014/07/30. doi: 10.1017/S0033291714000592. PubMed PMID: 25066636.
60. Horan WP, Braff DL, Nuechterlein KH, Sugar CA, Cadenhead KS, Calkins ME, et al. Verbal working memory impairments in individuals with schizophrenia and their first-degree relatives: Findings from the Consortium on the Genetics of Schizophrenia. Schizophrenia Research. 2008;103(1):218-28. doi: https://doi.org/10.1016/j.schres.2008.02.014.
61. Jiang YV, Palm BE, DeBolt MC, Goh YS. High-precision visual long-term memory in children with high-functioning autism. Journal of Abnormal Psychology. 2015;124(2):447-56. doi: 10.1037/abn0000022.
62. Koshino H, Carpenter PA, Minshew NJ, Cherkassky VL, Keller TA, Just MA. Functional connectivity in an fMRI working memory task in high-functioning autism. NeuroImage. 2005;24(3):810-21. doi: https://doi.org/10.1016/j.neuroimage.2004.09.028.

63. Bentall RP, Rowse G, Shryane N, Kinderman P, Howard R, Blackwood N, et al. The Cognitive and Affective Structure of Paranoid Delusions: A Transdiagnostic Investigation of Patients With Schizophrenia Spectrum Disorders and Depression. Archives of General Psychiatry. 2009;66(3):236-47. doi: 10.1001/archgenpsychiatry.2009.1.
64. Conway ARA, Kane MJ, Engle RW. Working memory capacity and its relation to general intelligence. Trends in Cognitive Sciences. 2003;7(12):547-52. doi: https://doi.org/10.1016/j.tics.2003.10.005.
65. Resnick SM. Matching for education in studies of schizophrenia. Archives of General Psychiatry. 1992;49(3):246.
66. Goris, J., Silvetti, M., Verguts, T., Wiersema, J. R., Brass, M., & Braem, S. (2020). Autistic traits are related to worse performance in a volatile reward learning task despite adaptive learning rates. *Autism*. doi: https://doi.org/10.1177/1362361320962237.
67. Fine C, Gardner M, Craigie J, Gold I. Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. Cognitive Neuropsychiatry. 2007;12:46-77.
68. Andreou C, Moritz S, Veith K, Veckenstedt R, Naber D. Dopaminergic modulation of probabilistic reasoning and overconfidence in errors: A double-blind study. Schizophrenia Bulletin. 2014;40:558-65.
69. Menon M, Mizrahi R, Kapur S. 'Jumping to conclusions' and delusions in psychosis: Relationship and response to treatment. Schizophrenia Research. 2008;98(1):225-31. doi: https://doi.org/10.1016/j.schres.2007.08.021.

Supporting information

S1 File. Mathematical model of the Bayesian observer.

S2 File. Results of the group comparisons after propensity matching by age and sex.

S3 Fig. Illustration of all stimuli used in the visual working memory task. Constitutes an exemplary representation of the circle of stimuli in which the target location had to be indicated on each trial. Stimuli are arranged according to their continuous modification. This pool of stimuli and the order of their arrangement were consistent across trials and participants.

S4 File. Additional analyses of beads task variables.

S5 File. Response time analysis of the visual working memory task.

Supplementary material S1 - mathematical model

1 Game rules

- There are two bags, bag A and bag B. The bags are filled with white and black balls. In bag A the fraction of white balls is p, whereas in bag B the fraction of white balls $1-p$.
- The game starts by the administrator drawing one of the bags at random. Let $x_1 = 0$ if the result is bag A and $x_1 = 1$ otherwise. Thus,

$$\mathrm{P}(x_1 = 0) = \mathrm{P}(x_1 = 1) = \frac{1}{2}.$$

- If $x_1 = 0$, the administrator is sampling a ball from bag A at random, and if $x_1 = 1$ the administrator is sampling a ball from bag B at random. The ball sampled is shown to the player and put back into the same bag as it was sampled from. Let $z_1 = 0$ if the draw results in a white ball, and $z_1 = 1$ otherwise. Thus,

$$\mathrm{P}(z_1|x_1) = p^{I(z_1=x_1)}(1-p)^{1-I(z_1=x_1)},$$

  where $I(\cdot)$ equals 1 if the argument is true and zero otherwise.

- For $i = 2, \ldots, n$ sequentially:
  - The administrator puts $x_i = 1 - x_{i-1}$ or $x_i = x_{i-1}$ with probabilities v and $1-v$, respectively.
  - If $x_i = 0$, the administrator is sampling a ball from bag A at random, and if $x_i = 1$ the administrator is sampling a ball from bag B at random. The ball sampled is shown to the player and put back into the same bag as it was sampled from. Let $z_i = 0$ if the draw results in a white ball, and $z_i = 1$ otherwise. Thus,

$$\mathrm{P}(z_i|x_i) = p^{I(z_i=x_i)}(1-p)^{1-I(z_i=x_i)}.$$

- After each ball is shown to the player, the player should
  - say from which bag (s)he thinks the last ball is coming, and
  - give an estimate on the probability that the last ball came from bag A.

2 Wanted results

In this note we discuss how to obtain the following

- Assuming the value of v to be known, compute the ideal Bayesian probability for the last ball to come from bag A, i.e. compute

$$P(x_n|z_1,\ldots,z_n,v) \tag{1}$$

  for each value of n.

- Assuming the value of v to be unknown, use the given probability estimates given by the player to estimate the value of v assumed by the player.

3 Computing $P(x_n|z_1,\ldots,z_n,v)$

To find $P(x_n|z_1,\ldots,z_n,v)$, one must first study $P(x_1,\ldots,x_n,z_1,\ldots,z_n|v)$. From the game rules it follows that

$$\begin{aligned}\mathrm{P}(x_1,\ldots,x_n,z_1,\ldots,z_n|v) &= \mathrm{P}(x_1,\ldots,x_n|v)\cdot\mathrm{P}(z_1,\ldots,z_n|x_1,\ldots,x_n)\\ &= \frac{1}{2}\prod_{i=2}^{n}\left[v^{1-I(x_i=x_{i-1})}(1-v)^{I(x_i=x_{i-1})}\right]\prod_{i=1}^{n}\left[p^{I(z_i=x_i)}(1-p)^{1-I(z_i=x_i)}\right].\end{aligned} \tag{2}$$

We have

$$\begin{aligned}P(x_n|z_1,\ldots,z_n,v) &= \frac{P(x_n,z_1,\ldots,z_n|v)}{P(z_1,\ldots,z_n|v)}\\ &\propto P(x_n,z_1,\ldots,z_n|v)\\ &= \sum_{x_1}\cdots\sum_{x_{n-1}}P(x_1,\ldots,x_n,z_1,\ldots,z_n|v),\end{aligned} \tag{3}$$

where the proportionality is as a function of x_n. To find $P(x_n|z_1,\ldots,z_n,v)$ we therefore need to evaluate the $n-1$ sums in (3) for each possible value of x_n and thereafter scale the result so that the values sum to one. For small values of n direct evaluation of the $n-1$ sums in (3) is computationally feasible, but for larger values of n the Markov structure present in (2) must be utilised to get a computationally efficient procedure. In the following we assume $n \geq 3$. The joint distribution in (2) can then be factorised into

$$\mathrm{P}(x_1,\ldots,x_n|v,z_1,\ldots,z_n) \propto h_{1,2}(x_1,x_2)\cdot h_{2,3}(x_2,x_3)\cdot\ldots\cdot h_{n-1,n}(x_{n-1},x_n), \tag{4}$$

where

$$h_{1,2}(x_1,x_2) = \frac{1}{2}v^{1-I(x_2=x_1)}(1-v)^{I(x_2=x_1)}p^{I(z_1=x_1)}(1-p)^{1-I(z_1=x_1)},$$

$$h_{i-1,i}(x_{i-1},x_i) = v^{1-I(x_i=x_{i-1})}(1-v)^{I(x_i=x_{i-1})}p^{I(z_{i-1}=x_{i-1})}(1-p)^{1-I(z_{i-1}=x_{i-1})}$$

for $i=3,\ldots,n-1$, and

$$h_{n-1,n}(x_{n-1},x_n) = v^{1-I(x_n=x_{n-1})}(1-v)^{I(x_n=x_{n-1})}p^{I(z_{n-1}=x_{n-1})}(1-p)^{1-I(z_{n-1}=x_{n-1})}$$

$$\cdot p^{I(z_n=x_n)}(1-p)^{1-I(z_n=x_n)}.$$

One should note that all the $h_{i-1,i}(x_{i-1},x_i)$ functions also depends on the value of v and the values $z_1,\ldots,z_n$ even if this dependence is not explicitly represented in the notation. Defining

$$g_2(x_2)=\sum_{x_1} h_{1,2}(x_1,x_2) \tag{5}$$

and

$$g_i(x_i)=\sum_{x_{i-1}} g_{i-1}(x_{i-1})h_{i-1,i}(x_{i-1},x_i) \tag{6}$$

for $i=3,\ldots,n$, we get that $g_n(x_n)$ equals the right hand side of (3). Thus,

$$P(x_n|z_1,\ldots,z_n,v)=\frac{g_n(x_n)}{\sum_x g_n(x)}. \tag{7}$$

To evaluate $P(x_n|z_1,\ldots,z_n,v)$ for each possible value of x_n can thereby be done in the following steps.

1. For each $i=2,\ldots,n$, evaluate $h_{i-1,i}(x_{i-1},x_i)$ for each possible combination of values for x_{i-1} and x_i. As the possible values for each of x_{i-1} and x_i is zero and one, four values must be computed for each value of i.
2. Using (5), compute $g_2(x_2)$ for $x_2=0$ and for $x_2=1$.
3. For $i=3,\ldots,n$ in turn, use (6) to compute $g_i(x_i)$ for $x_i=0$ and for $x_i=1$.
4. Using (7), compute $P(x_n|z_1,\ldots,z_n,v)$ for $x_n=0$ and for $x_n=1$.

4 Estimate the value of v used by the player

We now assume the player is using a value of the parameter v when deciding on the probability estimates. We let K denote the number of games or rounds the player is playing, and assume that the player sees N balls in each play. We let $\widetilde{p}_{k,n}$ denote the probability estimate specified by the player after seeing ball number n in play number k. One should note that $\widetilde{p}_{n,k}$ is the players guess on the probability $P(x_{k,n}|z_{k,1},\ldots,z_{k,n},v)$, where $x_{k,n}$ and $z_{k,i}$ corresponds to x_n and z_i, respectively, in Section 3, but where we have now added an index k to distinguish the K rounds played. As the theoretical probability $P(x_{k,n}|z_{k,1},\ldots,z_{k,n},v)$ is a function of v, one can formally estimate the value of v used by the player by finding the value that makes the set of theoretical probabilities $P(x_{k,n}|z_{k,1},\ldots,z_{k,n},v)$ as close as possible to the probability estimates $\widetilde{p}_{k,n}$. More precisely, we suggest to estimate v by minimising the sum of squares of the differences between the probability estimate $\widetilde{p}_{k,n}$ specified by the player and the corresponding theoretical probability $P(x_{k,n}|z_{k,1},\ldots,z_{k,n},v)$. Thus, we define the estimate as

$$\widehat{v}=\operatorname*{argmin}_v\left[\sum_{k=1}^{K}\sum_{n=1}^{N}\left(\widetilde{p}_{k,n}-P(x_{k,n}|z_{k,1},\ldots,z_{k,n},v)\right)^2\right]. \tag{8}$$

The minimisation must be done by some numerical minimisation algorithm, within which the theoretical probabilities, $P(x_{k,n}|z_{k,1},\ldots,z_{k,n},v)$, for any value of v can be computed as discussed in Section 3.

Supplementary material S2 - matched group comparisons

To account for differences in age and sex between the groups, propensity matching was applied, resulting in separately matched subsamples of the neurotypically developing individuals (NT group) for the participants with autism spectrum disorders (ASD group) and the ones with schizophrenia (SCZ group), respectively, as well as a matched subsample of the SCZ group for comparison with the ASD group. After matching, none of the paired samples differed significantly on either age or sex.
To test for group differences, multiple T-Tests were conducted for every single variable of interest. None of the Levene's tests indicated significant heterogeneity of variance for any of the subsample's comparisons. Therefore, no Welch corrections were performed. Effect sizes were calculated as Cohen's *d*. Since estimated volatility and disconfirmatory belief updating were not normally distributed within most of the subgroups, Mann Whitney-U tests were conducted for non-parametric verification of the results. Those results, as well as the associated effect sizes calculated as Cliff's delta, are reported in brackets.

Group comparison for ASD vs. NT:

No significant difference between groups was found for error, $t(36) = 1.28$, $p = .21$, $d = 0.42$; proportion of hits, $t(36) = -0.86$, $p = .39$, $d = -0.28$; initial certainty, $t(36) = -0.44$, $p = .66$, $d = -0.14$; disconfirmatory belief updating, $t(36) = -0.47$, $p = .64$, $d = -0.15$ [U = 153, $p = .44$, Cliff's $d = -0.15$] and estimated volatility, $t(36) = 0.36$, $p = .72$, $d = 0.12$ [U = 186, $p = .89$, Cliff's $d = 0.03$].

Group comparison for SCZ vs. NT:

There was a significant effect of group on average error (i.e. memory inaccuracy) in the VWM task, $t(40) = -2.75$, $p = .01$, $d = -0.85$, indicating a less accurate visual-memory performance in participants with SCZ ($M = 35.72$, $SD = 11.58$) in comparison to healthy controls ($M = 26.34$, $SD = 10.54$).

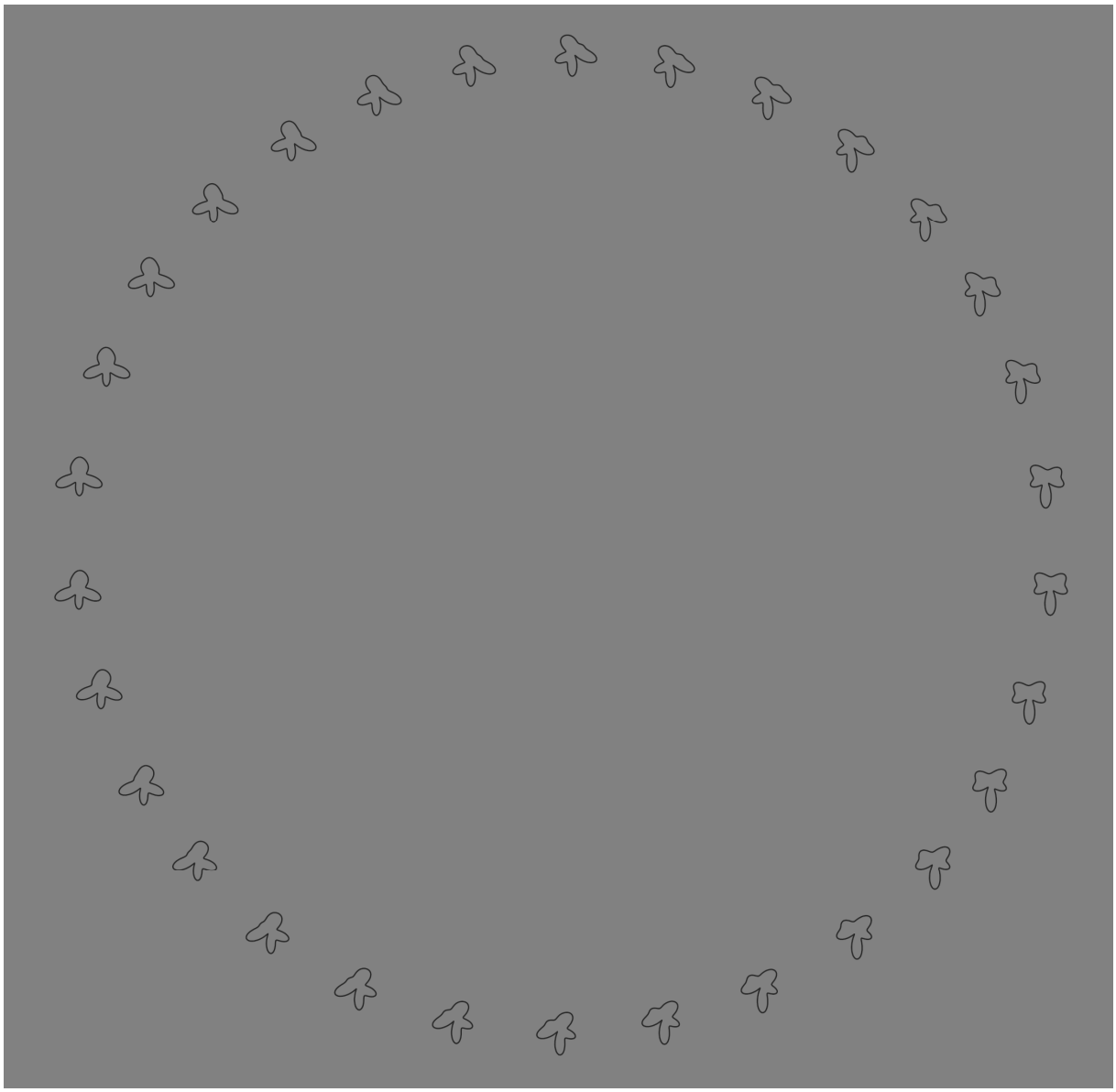
No significant difference between groups was found for proportion of hits, $t(40) = 0.60$, $p = .55$, $d = 0.19$; initial certainty, $t(40) = 0.61$, $p = .54$, $d = 0.19$; disconfirmatory belief updating, $t(40) = -0.56$, $p = .58$, $d = -0.17$ [U = 180, $p = .32$, Cliff's $d = -0.18$] and estimated volatility, $t(40) = -1.35$, $p = .18$, $d = -0.42$ [U = 171, $p = .22$, Cliff's $d = -0.22$].

Group comparison for ASD vs. SCZ:

The effect of group on average error (i.e. memory inaccuracy) in the VWM task was only marginally significant, $t(36) = -1.94$, $p = .06$, $d = -0.63$, with a (numerically) less accurate visual-memory performance in participants with SCZ ($M = 35.55$, $SD = 12.13$) in comparison to participants with ASD ($M = 28.23$, $SD = 11.07$).

No significant difference between groups was found for proportion of hits, $t(36) = 1.08$, $p = .29$, $d = 0.35$; initial certainty, $t(36) = 0.24$, $p = .81$, $d = 0.08$; disconfirmatory belief updating, $t(36) = -0.85$, $p = .40$, $d = -0.27$ [U = 141, $p = .26$, Cliff's $d = -0.22$] and estimated volatility, $t(36) = -0.83$, $p = .41$, $d = -0.27$ [U = 149, $p = .37$, Cliff's $d = -0.17$].

Supplementary material S3 – stimuli of the visual working memory task



S3 Fig. Illustration of all stimuli used in the visual working memory task. Constitutes an exemplary representation of the circle of stimuli in which the target location had to be indicated on each trial. Stimuli are arranged according to their continuous modification. This pool of stimuli and the order of their arrangement were consistent across trials and participants.

Supplementary material S4 – additional analyses of beads task variables

Random or 'noisy' decision-making and volatility

Estimation of subjective volatility via the ideal Bayesian model assumes that deviations between 'ideal' probabilistic responses and the probability ratings made by a participant are largely caused by a misestimation of the true volatility. However, other causes are conceivable. As such, estimated volatility might be affected by "noisy" or "random" decision-making. Notably, it is difficult to conceptually differentiate such "noise" from volatility, as volatility per se might be the cause driving "noise" or seemingly "random" choice behavior.

Nevertheless, to obtain an approximate estimate of "random" or "noisy" behavior in the beads task, an additional measure was constructed based on all those occurrences where when a bead was of the same color as the previous two, the belief was updated into the *opposite* direction, i.e. the belief in the currently presented colors was decreased.

Example: a participant sees three white beads in a row and indicates a probability for them to originate from the bag with more white beads as 0.7 and 0.8 for the first two trials. On the third trial, they then *decrease* their belief to 0.7 again when actually, given the evidence, they should keep *increasing* their belief certainty about the beads to originate from the bag with more white beads.

Such "random belief updating" was calculated as the mean change in belief across all occurrences of this kind for each sequence, averaged over number of sequences for each participant.

A non-parametric Kruskal-Wallis test (due to the high positive skewness in random belief updating) revealed no significant group difference, $\chi^2(2) = 3.32$, $p = 0.19$, $\varepsilon^2 = 0.04$.

Across groups, random belief updating was strongly and positively associated with volatility, $\rho = .63$, $p < .001$. While this might suggest that estimated volatility largely reflected noise or random behavior, it is important to consider that a conceptual distinction between both concepts may not fully be valid. After all, "random" belief changes may indeed be caused by an increased belief about the frequency with which the bag of origin is secretly changed (volatility), even in the absence of obvious evidence for an occurred change.

Importantly, volatility was also strongly related to disconfirmatory belief updating. Here, the conceptual relationship between both variables is slightly more obvious: in an unstable environment, disconfirmatory evidence might suggest an occurred change – so the larger one thinks the probability is for a change to occur, the more one will react to disconfirmatory evidence in terms of belief updating.

An additional analysis was conducted to gauge to what extent both random and disconfirmatory belief updating contributed to estimated volatility. Participants were divided into groups with high (above the median) or low (below or equal to the median) volatility estimates. A logistic regression was conducted on volatility group membership (0 = low, 1 = high), including main

effects of both random and disconfirmatory belief updating, both standardized. McFadden's R^2 of this model was .40, and the Odds Ratio was 10.12 for (standardized) random belief updating [CI 2.5%: 2.64, 97.5%: 53.23] and 1.92 for (standardized) disconfirmatory belief updating [CI 2.5%: 1.92, 97.5%: 10.62]. This demonstrates that even if random belief updating was interpreted as a pure measure of "noise" caused by different factors than an overestimation of volatility, when accounting for its contribution to volatility there remains a significant contribution of disconfirmatory belief updating, a variable which is clearly also conceptually related to volatility.

Volatility change throughout the task

Since feedback was provided after every completed sequence in the beads task, learning processes may have caused a decrease in subjective volatility over time. In the original volatility model, subjective volatility was estimated based on all sequences. To explore whether volatility estimates might have decreased over time, the model was refitted to the first two and the last two sequences, respectively. Volatility change was then calculated by subtracting volatility estimated for the first two sequences from volatility estimated for the last two sequences for each participant, with values below zero indicating a decrease of volatility towards the end of the task.

A one-sided one-sample Wilcoxon signed-rank test (due to the non-normality of the volatility change variable) on data of the whole sample confirmed that indeed, this change was significantly below zero across participants, $Md = -0.01$, $V = 1260$, $p < .01$.

To assess whether groups differed in terms of this volatility change, a Kruskal-Wallis test was applied. This did not reveal any significant group differences, $\chi^2(2) = 0.77$, $\varepsilon^2 = 0.06$, $p = .68$, indicating that groups learned similarly from feedback.

Supplementary material S5 – response time analysis (VWMT)

Response times in the visual working memory task (VWMT) were compared between groups. They were recorded as the time between a) appearance of the array from which the target shape had to be selected and b) finalization of the capture area. For each participant, invalid trials (i.e. skipped trials or trials where the capture area failed to meet the inclusion criteria) as well as response times larger than mean response time + 2.5*standard deviation or smaller than mean response time - 2.5*standard deviation were excluded. An ANOVA conducted on the log-transformed average response times showed a significant group effect, $F(2,83) = 7.99$, $\eta^2 = 0.16$, $p < .001$, with $M_{ASD} = 2.64$, $M_{SCZ} = 2.24$, and $M_{NT} = 2.47$ (not log-transformed means: $M_{ASD} = 14.76$, $M_{SCZ} = 9.87$, and $M_{NT} = 12.55$). Post-hoc comparisons revealed that response times in the SCZ group were significantly shorter as compared to both the NT, $p_{adj} = .02$, and the ASD group, $p_{adj} < .001$. Response times did not differ significantly between the ASD and the NT group, $p_{adj} = .15$.

Notably, shorter response times may on the one hand indicate that a reduced amount of time is allocated to the processing of the visually presented information. Accordingly, shorter response times in the SCZ group might explain the decreased recall performance. On the other hand, shorter response times may reflect increased certainty about the upcoming choice (see e.g. Rahnev et al., 2020) and as such might generally predict increased performance. To elucidate these questions, a linear mixed-effects model was specified, with trial-wise recall error (i.e. deviation from target) as outcome, and trial-wise log transformed response times, group, and their interaction as predictors, including a random intercept for participant number. This revealed a significant positive association between response times and error, meaning recall accuracy was lower on trials where response times were longer (see Table S5). However, this relationship did not differ significantly between NT and SCZ, or NT and ASD group. Taken together, these results suggest that the reduced memory accuracy observed in the SCZ group cannot be explained by their tendency to respond faster.

Table S5. Linear mixed-effects model results for error in the visual working memory task.

	b	*t*	*p*
log(RT)	6.95	3.06	< .01
ASD	-1.78	-0.17	.87
SCZ	21.47	2.39	.02
log(RT)*ASD	1.77	0.44	.66
log(RT)*SCZ	-4.00	-1.05	.30

Notes: log(RT) = log-transformed response time; baseline for group effects: neurotypical control group

References (suppl.)

Rahnev, D., Desender, K., Lee, A. L., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., ... & Bègue, I. (2020). The confidence database. *Nature human behaviour*, *4*(3), 317-325.

Paper II

Spared performance but increased uncertainty in schizophrenia: evidence from a probabilistic decision-making task

Isabel Kreis[a]*, Lei Zhang[a,b] Steffen Moritz[c], & Gerit Pfuhl[a]

[a]*Department of Psychology, Faculty of Health Sciences, UiT – The Arctic University of Norway, 9037, Tromsø, Norway,*

[b]*Social, Cognitive and Affective Neuroscience Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010, Vienna, Austria*

[c]*Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Martinistraße 52, 20246, Hamburg, Germany*

Correspondence: Isabel Kreis, Department of Psychology, UiT – The Arctic University of Norway, 9037 Tromsø, Norway, isabel.v.kreis@uit.no

Abstract

Aberrant attribution of salience to in fact little informative events might explain the emergence of positive symptoms in schizophrenia and has been linked to belief uncertainty. Uncertainty is thought to be encoded by neuromodulators, including norepinephrine. However, norepinephrinergic encoding of uncertainty, measured as task-related pupil dilation, has rarely been explored in schizophrenia. Here, we addressed this question by comparing individuals with a disorder from the schizophrenia spectrum to a non-psychiatric control group on behavioral and pupillometric measures in a probabilistic prediction task, where different levels of uncertainty were introduced. Behaviorally, patients performed similar to controls, but their belief uncertainty was higher, particularly when instability of the task environment was high, suggesting an increased sensitivity to this instability. Furthermore, while pupil dilation scaled positively with uncertainty in the control group, this was not the case for patients, suggesting aberrant neuromodulatory regulation of neural gain, which may hinder the reduction of uncertainty in the long run. Together, the findings point to abnormal uncertainty processing and norepinephrinergic signaling in schizophrenia, potentially informing future development of both psychopharmacological therapies and psychotherapeutic approaches that deal with the processing of uncertain information.

Keywords: pupillometry; feedback sensitivity; positive symptoms; Hidden Markov Model; probabilistic reversal learning

1. Introduction

Aberrant salience attribution to insignificant events has been suggested to explain various symptoms in schizophrenia, including positive symptoms such as delusions (Kapur, 2003) and cognitive biases such as 'jumping-to-conclusions', where patients typically make or alter decisions based on little evidence (Speechley et al., 2010). Recent theories propose that salience is affected by uncertainty (Adams et al., 2013; Broyd et al., 2017; Fletcher and Frith, 2009). Here, increased attribution of salience ('hypersalience') to external information may result from increased uncertainty surrounding cognitive representations in the mind's belief hierarchy. Consequentially, perception and belief updating are biased towards external information and sensory events as opposed to prior beliefs, explaining the experience of 'strange percepts' in a state of delusional mood (Adams et al., 2013). Delusions may then manifest as an attempt to give meaning to these 'strange percepts' (Fletcher and Frith, 2009). Increased belief uncertainty might further explain why patients with schizophrenia often exhibit maladaptive switching behavior in probabilistic reversal learning tasks (Culbreth et al., 2016a; Kaplan et al., 2016; Li et al., 2014; Murray et al., 2008; Schlagenhauf et al., 2014; Waltz et al., 2013). In these tasks, participants have to learn which choice option is more likely to result in a positive outcome and have to adapt their choices once the choice-outcome probability (risk) reverses. A positive outcome should encourage staying with the previous choice, whereas a negative outcome might either reflect the inherent risk, in which case it should be disregarded, or indicate a change in risk, hence encouraging a choice switch. Increased choice switching observed in schizophrenia often occurs in response to both positive and negative outcomes (Culbreth et al., 2016a; Deserno et al., 2020; Waltz et al., 2013), though some have reported a decreased sensitivity particularly to positive feedback (Li et al., 2014; Schlagenhauf et al., 2014). Patients' impaired performance in these tasks may reflect either inherent deficits to learn about choice-outcome probabilities (*risk*; Murray et al., 2008; Reddy et al., 2016; Weickert et al., 2010), or an overestimation of the probability for those contingencies to change (*volatility*) (Cole et al., 2020; Deserno et al.,

2020; Schlagenhauf et al., 2014), and possibly both (Waltz et al., 2013). A misrepresentation of these different types of uncertainties (risk and volatility) may hence cause patients with schizophrenia to attribute too much salience to a given outcome, resulting in increased switching between the different choice options even when it is not beneficial.

Mechanistically, hypersalience in schizophrenia has been linked to dysfunctional dopaminergic signaling (Heinz and Schlagenhauf, 2010), but the role of norepinephrine is less explored, despite its suggested association with uncertainty processing (Yu and Dayan, 2005). Norepinephrinergic activity in the locus coeruleus is reflected in pupil size (Joshi et al., 2016; Rajkowski et al., 1994; Samuels and Szabadi, 2008) and indeed, task-related pupil dilation responds to both outcome surprise and environmental volatility (Browning et al., 2015; Lawson et al., 2017; Nassar et al., 2012; Preuschoff et al., 2011), scales with the extent to which an outcome should evoke belief updating (Hämmerer et al., 2019), and signals fluctuations in neural gain and learning (Eldar et al., 2013). Early studies showed that pupil size scales less with the probabilities of presented stimuli in individuals with schizophrenia (Steinhauer and Zubin, 1982; Steinhauer et al., 1979), indicating a reduced adaptation of neural gain to uncertainty. However, it is unclear how this diminished pupil response would be affected by volatility. Furthermore, group differences regarding pupil responses, switching behavior and the extent to which they are affected by volatility, may depend on the particular risk conditions of the task. While the most commonly chosen choice-outcome probabilities are 0.20 and 0.80 (Culbreth et al., 2016a; Deserno et al., 2020; Waltz and Gold, 2007; Waltz et al., 2013), the differential effects of other risk conditions and their interaction with volatility remain to be explored.

To address the above questions, we compared individuals with a disorder from the schizophrenia spectrum to a non-psychiatric control group in a probabilistic prediction task where risk and volatility were manipulated independently. Using cognitive-computational

models, we estimated uncertainty related parameters and latent variables behind the observed behavior, and investigated their relationship with clinical symptoms, and pupil dilation.

2. Methods and Materials

Participants had to meet the following inclusion criteria: (1) 18 to 65 years old, (2) capacity for informed consent, (3) very good command of German, (4) IQ above 80, (5) normal or corrected-to-normal eyesight, (6) no history of neurological disorders, (7) no substance dependence, (8) no recreational drug consumption within one week prior to the assessment (excluding alcohol, nicotine, and caffeine), (9) a primary diagnosis of schizophrenia or schizoaffective disorder (SZ group; DSM-V, American Psychiatric Association, 2013) or no psychiatric diagnosis at all (HC group), verified with the Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998). The SZ group included in– and outpatients from the Department of Psychiatry and Psychotherapy of the University Medical Center Hamburg-Eppendorf (UKE), Germany, who were contacted directly or replied to announcements made on site. Control participants were recruited via student job websites and advertising leaflets. In total, 62 participants (SZ: n = 32, HC: n = 30) were recruited whereof one was excluded from all analyses because they failed to meet the inclusion criteria. The study was approved by the local ethics committee of psychologists at the UKE. All participants gave written informed consent prior to the study.

2.1 Measures

2.1.1 Probabilistic prediction task

To measure decision-making and belief updating under different risk and volatility conditions, a newly developed probabilistic prediction task was administered (Kreis et al., 2020b). On each trial, participants had to predict whether an upcoming Gabor patch would be tilted to the left or the right from the center (left-alt key for ‘left-tilted’, right-ctrl key for ‘right-tilted’; orientation ± 45°; see Fig. 1A). The probability for the left- or the right-tilted patch was unknown to the

participants and alternated between 85:15 (indicating outcome schedule, namely, 85% left-tilted and 15% right tilted) and 60:40 and the reverse (15:85, 40:60) after 20 (± 4) trials, constituting conditions of high (60:40/40:60) and low risk (85:15/15:85; Fig. 1B). Participants were instructed to track the probabilities and the changes as good as possible and to minimize the amount of prediction errors. In a first, volatile block of the task, changes between risk conditions were hidden, and in a second, cued block, changes were announced, constituting conditions of high (volatile) and low (cued) volatility, each spanning 160 trials (+12 and 18 practice trials, respectively). For the cued block, participants were advised to 'reset' their beliefs about the distribution of stimuli at every announced change point, and relearn the new underlying distribution through choice-outcome observations. While the order of the risk conditions was the same for both blocks and across participants to ensure the same reward structure across blocks, the identity of the majority Gabor patch was inverted (Fig. 1B). Since time points of changes were identical in both blocks but explicitly announced in the cued block, block order was not counterbalanced to prevent priming participants from detecting the hidden changes in the volatile block.

2.1.2 Working Memory Task: visual digit span task

To control for inter-individual differences in working memory capacity, a visual, computerized version of the digit span subtest of the Wechsler adult intelligence scale (WAIS-IV; Wechsler, 2008) was administered (for details see Kreis et al., 2020a). Working memory capacity was measured as the maximum amount of digits recalled in the correct order.

2.1.3 Clinical assessments and demographics

Demographic and clinical variables (see Table 1) were recorded during an interview. The MINI (Sheehan et al., 1998) was applied to confirm the self-reported information about the presence (SZ group) or absence (HC group) of clinical diagnoses. Within the SZ group, positive and

negative symptoms were assessed with the Positive and Negative Symptoms Scale (PANSS; Kay et al., 1987). Negative symptom scores were calculated as suggested by van der Gaag et al. (2006; subsequently PANSS-N_{vdGaag}). To estimate premorbid intelligence, the German multiple choice vocabulary test (WST; Lehrl et al., 1995) was administered.

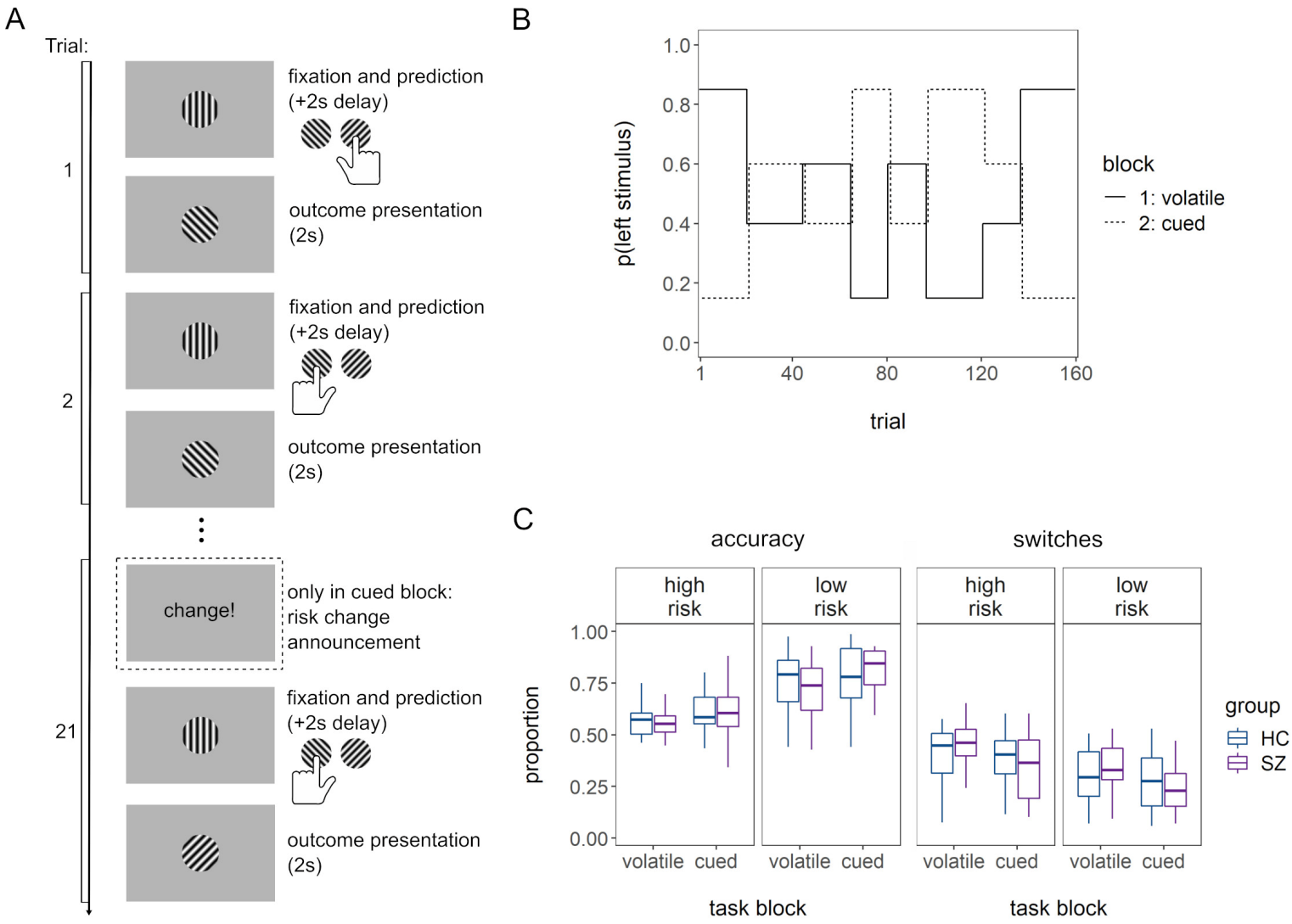


Fig. 1: Probabilistic prediction task. A) Trial structure: Example trials 1, 2 and 21 are displayed. Each trial started with the presentation of a vertically striped Gabor patch. Participants then had to predict via a button press whether the upcoming patch was going to be either left- or right-tilted from the center. After a fixed two-second delay, the outcome was presented and remained on screen for another two seconds. Then the vertical patch reappeared, prompting the next trial/prediction. Within the cued task block, changes in risk conditions were announced in the beginning of the respective trial (see B) through a 'change' message that appeared on screen. No further information was provided about the nature of the upcoming risk condition. Participants had to press 'enter' in response to that change message before they could continue with the task in order to guarantee that they perceived it. B) Task structure: the probabilities for

the left- (p(left)) and the right-tilted (1-p(left)) Gabor patch changed at fixed time points after 20 ± 4 trials. In the volatile block (solid line), these changes were hidden, and in the cued block (dashed line) they were announced (see A). Whereas the timing of change points and the order of the different risk conditions were identical across blocks (lines are only jittered for display), the identity of the respective majority stimulus within a block was inversed. C) Proportion of accurate predictions (prediction of current majority stimulus; left panels) and proportion of choice switches (prediction on trial $t + 1$ is different from prediction on trial t; right panels) for each group on trials of high and low risk within the volatile and the cued block of the task.

2.1.4 Pupil size

Pupil diameter was recorded from the left (in seven cases from the right) eye at a sampling rate of 500 Hz with an infrared video-based eye tracker (Eyelink 1000, SR Research) during the prediction task.

2.2 Procedure

First, demographic and clinical variables were recorded. Next, the volatile block of the prediction task was administered, followed by the working memory task, a brief decision-making task (not reported here) and the WST. Then, the cued block of the prediction task was completed. At the end of the session, the clinical assessment was conducted with the MINI and the PANSS.

2.3 Analysis

To test for the relevance of potential covariates, SZ and HC group were compared regarding age, education, premorbid verbal intelligence and working memory capacity, using non-parametric methods when variables were not normally distributed. To investigate the relationships between task conditions, group membership, behavioral performance, pupil dilation, and latent variables as extracted from cognitive-computational models, linear mixed-effects models were implemented. Their residuals were tested for normality and dependent variables were cube root or square transformed if normality was violated. Group-level

parameters (estimated using the hierarchical Bayesian approach) from the winning cognitive-computational model were compared between groups and task conditions by contrasting their posterior sampling distributions (Zhang et al., 2020). Associations between symptoms and cognitive-computational parameters were tested with Spearman correlations (ρ) under conditions of non-normality. Testing was conducted with a significance level of 0.05 using R (R version 3.5.1; R Core Team, 2018)).

2.3.1 Cognitive-computational modelling of behavior

To quantify latent cognitive processes, various cognitive-computational models were fitted to participants' predictions (i.e. 'left' or 'right') and observed outcomes (i.e. correct or incorrect) for the volatile and the cued block, respectively, and separately for the SZ and the HC group. The models included a win-stay-loose-shift model (Worthy and Todd Maddox, 2014), four different Reinforcement Learning models (den Ouden et al., 2013; Gläscher et al., 2008; Pearce and Hall, 1980; Rescorla and Wagner, 1972), and two variants of a Hidden Markov Model (HMM; Schlagenhauf et al., 2014) – all chosen to allow for the fact that participants might employ different strategies when solving the task (see Supplementary Material for details). For the cued block, additional variants of all models were specified that incorporated belief resets whenever a change in risk condition was announced.

Models were estimated using a Markov chain Monte Carlo (MCMC) within the hierarchical Bayesian framework (Ahn et al., 2017; Gelman et al., 2013). For both groups and both blocks, respectively, a variant of the HMM provided the best fit (see Supplementary Material for model comparison). The HMM, a Bayesian inference model, assumes a higher-order representation of the task structure that accounts for the instability of the task environment. Here, participants are expected to choose 'left' or 'right' depending on whether they believe to be in a left- ('majority stimulus is *left*') or right-tilted hidden state ('majority stimulus is *right*'). State beliefs are inferred and updated on each trial, depending on the history

of choice-outcome pairs as well as the estimated transition probability γ, which quantifies how the two hidden states are expected to change. Thus, γ indicates a participant's perceived volatility of the task environment. In the winning model (HMM_{RP}), positive (correct prediction) and negative (incorrect prediction) feedback sensitivity were allowed to differ since positive and negative feedback may affect participants' belief updating differently. For the cued block, the winning model included belief resets.

To obtain a measure that indicates to which extent a state belief should be updated on a given trial, Bayesian surprise was estimated as the Kullblack-Leibler divergence of the trial-wise state beliefs before ($P(S_{t_{pre}})$), and after an outcome observation ($P(S_{t_{post}})$), extracted from the HMM_{RP}:

$$D_{KL}(P(S_{t_{post}})||P(S_{t_{pre}})) = \sum_{i=1}^{2} P(S_{t_{pre}} = i) \log\left(\frac{P(S_{t_{post}}=i)}{P(S_{t_{pre}}=i)}\right) \qquad \textbf{\textit{Eq. 1}}$$

As a measure of trial-wise uncertainty regarding the hidden states, belief entropy, $H(S_t)$, was estimated based on the posterior for the different probabilities of the prediction to be correct. Hence, on a given trial this reflected a participant's uncertainty about the current task state:

$$H(S_t) = -\sum_{i=1}^{2} P(S_t = i) \log P(S_t = i) \qquad \textbf{\textit{Eq. 2}}$$

2.3.2 Pupil signal preprocessing

The pupil signal was corrected for eye blinks and other artefacts based on the signal's velocity and subsequent cubic-spline interpolation (Mathôt et al., 2018). Missing data of more than 1000 consecutive milliseconds were not interpolated but treated as missing in subsequent analyses. The corrected signal was smoothed with a 3 Hz low pass Butterworth filter and z-scored per task block and participant. The z-scored signal was baseline-corrected per trial through subtraction of the average signal of the 500 ms preceding outcome onset. Trials where more

than 50% of the signal were missing or interpolated were treated as missing in subsequent analyses.

2.3.3 Data exclusion

For one participant, all data for the cued block were treated as missing as they aborted before completion. Another participant was excluded from the computational model of the volatile block, as prior modeling attempts resulted in an inappropriate fit. All pupil data of a participant within a task block were treated as missing if more than 50% of trials were missing within that block (no. of exclusions in volatile block: $n_{HC} = 1$, $n_{SZ} = 5$; cued block: $n_{HC} = 1$, $n_{SZ} = 5$, three overlapping with volatile block).

3. Results

No significant group differences emerged in any of the demographic variables or working memory capacity (Table 1).

3.1 Behavioral performance: accuracy and choice switching

Frequency of predicting the current majority stimulus (accuracy; square transformed) was higher when volatility was low (i.e. in the cued task block; $b = 0.10$, $t = 4.25$, $p < .001$; Fig 1C) and within the low-risk condition ($b = -0.23$, $t = -9.51$, $p < .001$; Fig 1C). The interaction between volatility and risk was not significant ($b = -0.04$, $t = -1.05$, $p = .297$). When including group as a predictor, neither group membership ($b = -0.05$, $t = -1.03$, $p = 0.308$), nor any of the interactions between volatility and group ($b = 0.08$, $t = 1.61$, $p = .113$), risk and group ($b = 0.02$, $t = 0.41$, $p = .680$), or volatility, risk and group ($b = -0.02$, $t = -0.29$, $p = .773$) showed a significant effect.

Table 1

Sample demographics per group (total sample size = 61)[a]

	SZ (*n* = 31)			HC (*n* = 30)			
	n	*M* (*SD*)	*Md* (*IQR*)	*n*	*M* (*SD*)	*Md* (*IQR*)	*p*
Gender (m/f)	16/15			13/17			.696
Education ("1"/"2"/"3")	1/2/28			1/5/24			.454
Age		47.13 (11.43)	48 (15)		45.80 (11.64)	47 (16.75)	.740
WST		33.55 (3.43)	34 (3)		32.37 (4.55)	34 (6.25)	.436
WMC	29[b]	6.10 (1.47)	6 (2)	30	6.77 (1.36)	7 (1)	.071
PANSS							
Positive Scale		11.77 (4.18)	11 (6)				
Negative Scale[c]		12.71 (4.43)	12 (4.5)				
Total score		49.29 (14.13)	45 (15.50)				
Time since onset		19.39 (11.73)	19 (11.50)				
Inpatients/Outpatients	6/25						
Antipsychotic medication	26						
First generation	1						
Second generation	21						
Both	4						
Other psychotropic drugs	13						

Notes: Sample sizes (*n*), counts, means (*M*; with standard deviations *SD*) and medians (*Md*; with inter-quartile ranges *IQR*) are displayed. Education was recorded in German school system categories corresponding to completion of 1 = secondary school I (up to age 15), 2 = secondary school II (up to age 16), 3 = 6th form college (up to age 19); WST = German vocabulary test; WMC = working memory capacity; p-values for group comparisons are provided for gender and education (Chi-squared tests), age, WMC and the WST scores (Mann–Whitney U tests).

[a] A subgroup of this sample (n = 59) was previously described in Kreis et al. (2020a)

[b] WMC results are only available for 29 patients as technical errors caused incorrect scores for two of the patients

[c] PANSS-NvdGaag: Negative symptom scores calculated as suggested by van der Gaag et al. (2006)

Proportion of choice switches (prediction on trial $t + 1$ is different from prediction on trial t) was lower in the low volatility condition (b = -0.06, t = -3.14, p = .003) and higher on high-risk trials (b = 0.11, t = 8.80, p < .001). The interaction between volatility and risk was not significant (b = 0.00, t = 0.06, p = .950). Including group as a predictor revealed no significant effect for group (b = 0.04, t = 1.25, p = .217), or a risk by group interaction (b = 0.00, t = 0.17, p = .862). The interaction between volatility and group indicated that patients adapted the amount of

choice switching more between blocks (see Fig 1C). However, this effect was not significant (b = -0.06, t = -1.86, p = .068), and neither was the three-way interaction of volatility, risk and group (b = -0.03, t = -0.90, p = .372).

3.2 Cognitive-computational parameters

The HMM_{RP} (see 2.3.1) entailed three parameters: sensitivity to positive feedback (c), sensitivity to negative feedback (d), and participants' beliefs about the transition probability (γ). The corresponding group parameters per block are presented in Table 2.

Table 2
Group parameters of the HMM_{rp} per group and block

	volatile block				cued block			
	HC		SZ		HC		SZ	
	μ	*SD*	μ	*SD*	μ	*SD*	μ	*SD*
gamma (γ)	0.11	1.00	0.12	0.85	0.09	1.15	0.05	0.70
sensitivity to positive feedback (c)	0.96	1.71	0.84	1.88	0.92	1.20	0.91	1.19
sensitivity to negative feedback (d)	0.87	1.58	0.86	1.28	0.80	1.67	0.67	1.98

To test for effects of volatility condition, group and their interaction on these group parameters, posterior distribution comparisons were conducted and the 89% highest density intervals (HDI; see McElreath, 2020) of the differences between block, group and their respective difference were investigated. For γ, the comparison revealed credibly higher values (Table 2) in the volatile than in the cued task block, without indication of a main group effect or an interaction (Fig 2). For c, all HDIs included zero and for d, there was again only a credible effect of task block (Fig 2).

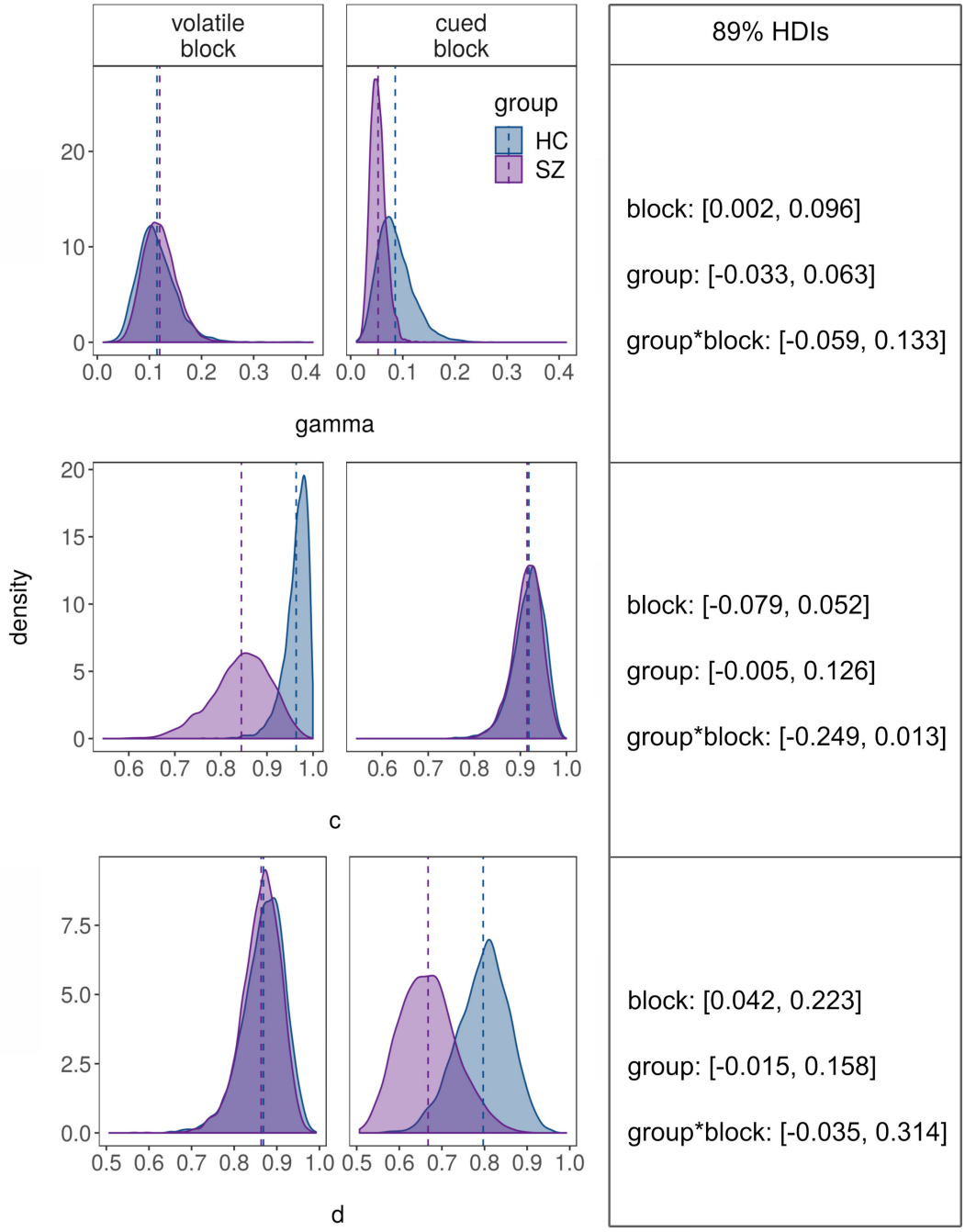


Fig. 2 Density plots displaying the posterior distribution of the HMM$_{RP}$ parameters fitted separately per group and per block. Dashed vertical lines indicate the posterior mean. Right column displays the 89% highest density intervals (HDIs) of the posterior distribution differences between block, group and their respective difference (group*block interaction).

We further explored how individual model parameters were related to positive or negative symptoms within the SZ group (Table 3). Here, severity of positive symptoms was associated

with a decreased sensitivity to positive feedback c under low volatility (ρ = -.40, p = .030), while all other results were not significant.

Table 3
Spearman correlations between individual HMM_{RP} parameters and symptoms

	volatile block				cued block			
	PANSS-P		PANSS-N_{vdGaag}		PANSS-P		PANSS-N_{vdGaag}	
	ρ	*p*	ρ	*p*	ρ	*p*	ρ	*p*
gamma (γ)	0.17	.374	-0.11	.575	-0.03	.858	-0.21	.262
sensitivity to positive feedback (c)	0.16	.394	0.16	.386	-0.40	.030	0.17	.375
sensitivity to negative feedback (d)	-0.12	.530	0.04	.831	-0.20	.300	-0.01	.948

3.3 Bayesian surprise, entropy and pupil dilation

The effect of pupil dilation, group, volatility, risk, and their interactions on trial-wise Bayesian surprise and belief entropy (uncertainty), both cube root transformed, were assessed with linear mixed effect models. Baseline pupil size variation did not differ between the groups within the volatile (U = 460, p = .167; Md_{HC} = 0.14, Md_{SZ} = .12; n = 55) or the cued task block (U = 396, p = .570; Md_{HC} = 0.15, Md_{SZ} = .13; n = 54). To control for potential effects of the anticholinergic load induced by daily dosage of the prescribed antipsychotics in the SZ group (Minzenberg et al., 2004; Naicker et al., 2016), benztropine mesylate equivalents, where available (n = 27), were calculated and correlated with baseline variation. This revealed no significant relationship in the volatile (ρ = 0.25, p = .269) or the cued block (ρ = 0.21, p = .358).

Uncertainty was higher on high-risk trials (Table 4) and within the SZ group (b = 0.12, t = 2.85, p = .006), though this seemed to be most pronounced during high volatility (b = -0.08, t = -2.01, p = .051). To test for group differences in psychophysiological responses to uncertainty, the interaction between group and maximum pupil dilation during outcome presentation as well as their three-way interaction with block were of main interest. Pupil dilation significantly predicted uncertainty on a given trial (b = 0.03, t = 5.66, p < .001). A

significant negative interaction between block and pupil dilation indicated that the positive association between pupil dilation and uncertainty was smaller during low volatility (b = -0.02, t = -3.86, p < .001). The positive relationship between pupil size and entropy was smaller in the SZ group (b = -0.02, t = -2.17, p = .030), meaning patients seemed to adapt their pupil size less in response to uncertainty (Fig 3).

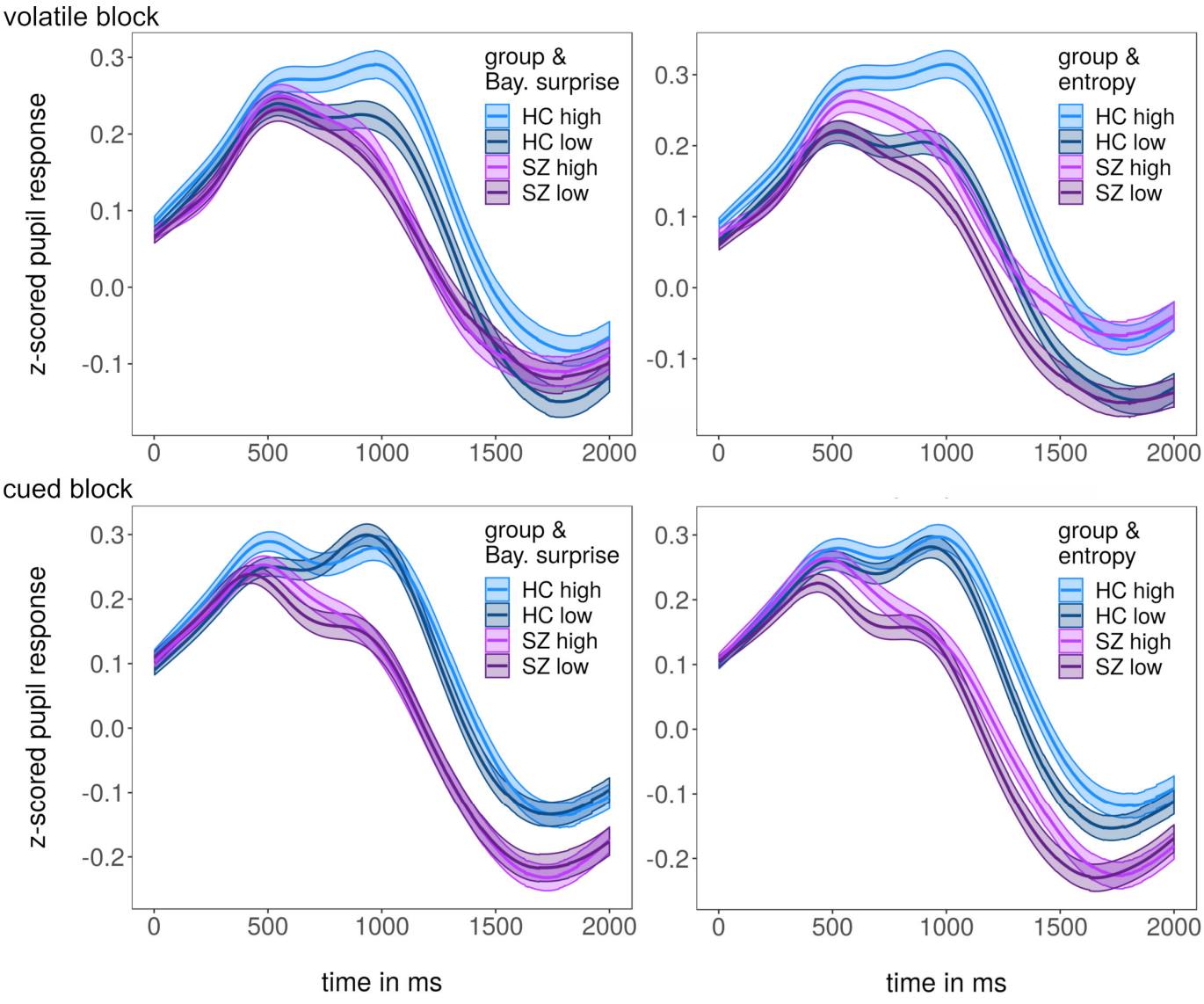


Fig. 3 Pupil responses to Bayesian surprise (left panels) and belief entropy (uncertainty; right panels) for each block. Trials of high and low values of the latent variables were categorized for each participant separately, with high values above and low values below or equal to the participant specific median within a block. HC = individuals without psychiatric disorder, SZ = individuals with disorder from the schizophrenia spectrum.

Table 4

Linear mixed-effects model results for entropy (uncertainty)

IV	*b*	*t*	*p*	R^2_M	R^2_C
				0.07	0.44
Block	0.04	1.61	.11		
Risk	0.08	5.37	<.001		
Group	0.12	2.85	.006		
PDmax	0.03	5.66	<.001		
Block*Risk	0.03	1.60	.112		
Block*Group	-0.08	-2.01	.051		
Block*PDmax	-0.02	-3.86	<.001		
Risk*Group	0.00	-0.02	.984		
Risk*PDmax	0.01	1.025	.305		
Group*PDmax	-0.02	-2.17	.030		
Block*Risk*Group	-0.02	-0.53	.597		
Block*Group*PDmax	0.01	1.44	.150		
Risk*Group*PDmax	0.00	0.37	.711		

Notes: Entropy= cube root transformed choice uncertainty (HMM_{RP});
IV = independent variable, Block = contrast of the second, cued task block to the first, volatile task block; Risk = contrast of the high- to the low-risk condition; Group = contrast of the SZ (schizophrenia) to the HC (controls) group; PDmax = maximum z-scored pupil dilation from baseline during presentation of the outcome stimulus on a given trial; R^2m = marginal R^2, i.e. proportion of variance explained by the fixed effects alone; R^2c = conditional R^2, i.e. proportion of variance explained by both the fixed and random effects (R^2m and R^2c based on Nakagawa and Schielzeth, 2013).

Bayesian surprise was significantly higher on high-risk trials but did not differ by volatility or between groups. Neither pupil dilation nor any of the associated interactions were significant predictors (Table 5).

Table 5

Linear mixed-effects model results for Bayesian surprise

IV	*b*	*t*	*p*	R^2_M	R^2_C
				0.05	0.36
Block	-0.05	-1.63	.109		
Risk	0.08	4.51	<.001		
Group	-0.02	-0.35	.727		
PDmax	0.01	1.49	.135		
Block*Risk	-0.00	-0.14	.892		
Block*Group	-0.03	-0.72	.477		
Block*PDmax	-0.01	-1.02	.307		
Risk*Group	0.01	0.29	.776		
Risk*PDmax	0.00	0.42	.676		
Group*PDmax	0.01	1.14	.253		

Block*Risk*Group	0.01	0.39	.700
Block*Group*PDmax	-0.01	-0.91	.362
Risk*Group*PDmax	-0.02	-1.65	.099

Notes: Bayesian surprise = cube root transformed belief updating (HMM_{RP});
IV = independent variable, Block = contrast of the second, cued task block to the first, volatile task block; Risk = contrast of the high- to the low-risk condition; Group = contrast of the SZ (schizophrenia) to the HC (controls) group; PDmax = maximum z-scored pupil dilation from baseline during presentation of the outcome stimulus on a given trial; R^2m = marginal R^2, i.e. proportion of variance explained by the fixed effects alone; R^2c = conditional R^2, i.e. proportion of variance explained by both the fixed and random effects (R^2m and R^2c based on Nakagawa and Schielzeth, 2013).

4. Discussion

Here, we investigated decision-making under uncertainty in a probabilistic prediction task where risk and volatility were independently manipulated to assess their effect on behavior in individuals with a diagnosis from the schizophrenia spectrum (SZ group) and non-psychiatric controls (HC group).

While task manipulation had the expected effects, with lower accuracy and more switches when risk or volatility were high, groups did not differ. This contrasts previous findings of impaired probabilistic learning and increased switching behavior in patients with schizophrenia and first-episode psychosis (Culbreth et al., 2016a; Deserno et al., 2020; Murray et al., 2008; Waltz et al., 2013) and may in part reflect task paradigm differences. In studies where a monetary reward is implemented, group differences may emerge due to differences in valuation processes (Chang et al., 2019; Culbreth et al., 2016b). Importantly, average accuracy was above chance level for all task conditions, indicating successful learning and effort investment even in the absence of an external reward. Another difference concerns the selected risk conditions: in most reversal learning tasks, only one risk condition is employed (Culbreth et al., 2016a; Deserno et al., 2020; Waltz et al., 2013). Here, risk conditions varied to test whether this moderates group differences. The low-risk condition (85:15) may have been easier to track, even for patients, whereas the high-risk condition (60:40) may have been so demanding that even the HC group experienced difficulties - both contributing to smaller group differences.

This study, however, is not the first to report intact probabilistic learning in schizophrenia. Reddy et al. (2016) found preserved initial and reversal learning in a substantial subgroup of patients. Meanwhile, deficits in an impaired subgroup were linked to decreased feedback sensitivity and diminished neurocognitive performance, e.g. lower working memory capacity. Similar to their sample, our sample contained a large proportion of outpatients. Furthermore, working memory capacity did not differ significantly between SZ and HC group and groups were matched on relevant demographic variables and premorbid verbal intelligence. The general neurocognitive 'fitness', the rather stable psychopathology, and the comparable demographics of our sample may thus explain the absence of behavioral differences. This highlights the importance of considering the heterogeneity of schizophrenia populations when drawing conclusions from and comparing results across single studies in this field (see also Moritz et al., 2020).

Similar to the behavioral results, the lack of group differences on the main parameters of the cognitive-computational model were at odds with previous reports of increased subjective volatility in patients with schizophrenia (Schlagenhauf et al., 2014) or at high risk for psychosis (Cole et al., 2020). However, in line with the results from Reddy et al. (2016) and replicating in part previous reports of a decreased sensitivity to positive feedback in schizophrenia (Schlagenhauf et al., 2014), there was a negative correlation between positive symptoms and positive feedback sensitivity within the SZ group. Interestingly, this was only true when volatility was minimal, suggesting that despite announced environmental changes, participants with a higher current severity of delusions and hallucinations did not perceive a positive feedback (i.e. a correct prediction) as a reliable indicator for their choice to be correct. In the volatile condition, this correlation might have been overshadowed since hidden changes increased feedback unreliability overall. Particularly during high volatility, however, uncertainty was higher in the SZ group, demonstrating some increased sensitivity to the

environment's volatility in patients, even though this did not translate into a significantly increased model-based volatility estimate. Moreover, patients showed a decreased adaption of pupil size to uncertainty. When uncertainty is high, especially in volatile environments, a given outcome should be highly salient as it serves as a teaching signal that could help to decrease prior uncertainty. Accordingly, pupil dilation should be larger if interpreted as an index of neural gain (Eldar et al., 2013). Therefore, the results point to a reduced ability to differentiate between high and low salient, or informative, outcomes in the SZ group, in line with the aberrant salience account. It seems surprising that no similar effect was found for Bayesian surprise, which indicates the extent to which a given outcome should evoke internal belief updates. Here, results suggest that both SZ and HC participants did not adapt their psychophysiological responses to the size of Bayesian surprise. Interestingly, a recent study found diminished pupil responses to Bayesian surprise in older adults (age > 60; Hämmerer et al., 2019). Hence, our finding may partly be related to the relatively high median age in our sample.

Taken together, our study demonstrates that under certain conditions, individuals with a diagnosis from the schizophrenia spectrum exhibit probabilistic decision-making similar to that of non-psychiatric controls, even though they are more uncertain, particularly when the task environment is volatile. Furthermore, positive symptom severity is related to an attenuated positive feedback sensitivity during probabilistic learning, possibly driven by the generally increased uncertainty. The failure to reliably adapt pupil responses to the degree of uncertainty indicates a failure to differentiate between more and less informative outcomes. This might explain why uncertainty remains generally higher in the patient group and is not reduced through learning. The findings thus corroborate hypotheses of aberrant norepinephrinergic signaling in schizophrenia (Fitzgerald, 2014; Mäki-Marttunen et al., 2020) and call for further investigation of the different implicated neuromodulatory systems and their interactions. Accumulated evidence from this field could inspire the development of psychopharmacological

treatments where adding norepinephrine transmission modulating agents might show beneficial effects in subgroups of patients (Fitzgerald, 2014). Furthermore, the study highlights the role of uncertainty processing in schizophrenia, a concept that is already addressed in metacognitive training interventions (Moritz and Woodward, 2007). The future development of therapeutic interventions of this kind may profit from further insights into the distinct effects of different kinds of uncertainty, such as risk and volatility, on belief formation and updating in schizophrenia.

Acknowledgements

We thank all participants who dedicated their time and effort. We also thank the supportive staff of both the outpatient and the inpatient clinic of the Department of Psychiatry and Psychotherapy of the UKE and the members of the Clinical Neuropsychology Working Group of the Department of Psychiatry and Psychotherapy of the UKE for opening their facilities and providing indispensable guidance during clinical assessments.

This work was supported by a Norwegian Research council grant under Grant FRIMEDBIO 262338, awarded to GP. LZ was partially supported by the Vienna Science and Technology Fund (WWTF VRG13-007) and the Austrian Science Fund (FWF P 32686).

Data availability

Raw and processed anonymized data are available in an Open Science Framework repository: DOI 10.17605/OSF.IO/AD65K.

References

Adams, R.A., Stephan, K.E., Brown, H.R., Frith, C.D., Friston, K.J., 2013. The computational anatomy of psychosis. Front Psychiatry 4, 47.

Ahn, W.Y., Haines, N., Zhang, L., 2017. Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. Comput Psychiatr 1, 24-57.

Browning, M., Behrens, T.E., Jocham, G., O'Reilly, J.X., Bishop, S.J., 2015. Anxious individuals have difficulty learning the causal statistics of aversive environments. Nat Neurosci 18(4), 590-596.

Broyd, A., Balzan, R.P., Woodward, T.S., Allen, P., 2017. Dopamine, cognitive biases and assessment of certainty: A neurocognitive model of delusions. Clinical Psychology Review 54, 96-106.

Chang, W.C., Westbrook, A., Strauss, G.P., Chu, A.O.K., Chong, C.S.Y., Siu, C.M.W., Chan, S.K.W., Lee, E.H.M., Hui, C.L.M., Suen, Y.M., Lo, T.L., Chen, E.Y.H., 2019. Abnormal cognitive effort allocation and its association with amotivation in first-episode psychosis. Psychological Medicine, 1-11.

Cole, D.M., Diaconescu, A.O., Pfeiffer, U.J., Brodersen, K.H., Mathys, C.D., Julkowski, D., Ruhrmann, S., Schilbach, L., Tittgemeyer, M., Vogeley, K., Stephan, K.E., 2020. Atypical processing of uncertainty in individuals at risk for psychosis. NeuroImage: Clinical 26, 102239.

Culbreth, A.J., Gold, J.M., Cools, R., Barch, D.M., 2016a. Impaired Activation in Cognitive Control Regions Predicts Reversal Learning in Schizophrenia. Schizophrenia Bulletin 42(2), 484-493.

Culbreth, A.J., Westbrook, A., Barch, D.M., 2016b. Negative symptoms are associated with an increased subjective cost of cognitive effort. Journal of Abnormal Psychology 125(4), 528-536.

den Ouden, Hanneke E.M., Daw, Nathaniel D., Fernandez, G., Elshout, Joris A., Rijpkema, M., Hoogman, M., Franke, B., Cools, R., 2013. Dissociable Effects of Dopamine and Serotonin on Reversal Learning. Neuron 80(4), 1090-1100.

Deserno, L., Boehme, R., Mathys, C., Katthagen, T., Kaminski, J., Stephan, K.E., Heinz, A., Schlagenhauf, F., 2020. Volatility Estimates Increase Choice Switching and Relate to Prefrontal Activity in Schizophrenia. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 5(2), 173-183.

Eldar, E., Cohen, J.D., Niv, Y., 2013. The effects of neural gain on attention and learning. Nat Neurosci 16(8), 1146-1153.

Fitzgerald, P.J., 2014. Is elevated norepinephrine an etiological factor in some cases of schizophrenia? Psychiatry Research 215(3), 497-504.

Fletcher, P.C., Frith, C.D., 2009. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 10(1), 48-58.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian data analysis. CRC press.

Gläscher, J., Hampton, A.N., O'Doherty, J.P., 2008. Determining a Role for Ventromedial Prefrontal Cortex in Encoding Action-Based Value Signals During Reward-Related Decision Making. Cerebral Cortex 19(2), 483-495.

Hämmerer, D., Schwartenbeck, P., Gallagher, M., FitzGerald, T.H.B., Düzel, E., Dolan, R.J., 2019. Older adults fail to form stable task representations during model-based reversal inference. Neurobiology of Aging 74, 90-100.

Heinz, A., Schlagenhauf, F., 2010. Dopaminergic Dysfunction in Schizophrenia: Salience Attribution Revisited. Schizophrenia Bulletin 36(3), 472-485.

Joshi, S., Li, Y., Kalwani, R.M., Gold, J.I., 2016. Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. Neuron 89(1), 221-234.
Kaplan, C.M., Saha, D., Molina, J.L., Hockeimer, W.D., Postell, E.M., Apud, J.A., Weinberger, D.R., Tan, H.Y., 2016. Estimating changing contexts in schizophrenia. Brain 139(7), 2082-2095.
Kapur, S., 2003. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am J Psychiatry 160(1), 13-23.
Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. . Schizophrenia bulletin 13(2), 261-276.
Kreis, I., Moritz, S., Pfuhl, G., 2020a. Objective Versus Subjective Effort in Schizophrenia. Frontiers in Psychology 11(1469).
Kreis, I., Zhang, L., Mittner, M., Syla, L., Lamm, C., Pfuhl, G., 2020b, July 24. Aberrant uncertainty processing is linked to psychotic-like experiences, autistic traits and reflected in pupil dilation. OSF Preprint, doi: https://doi.org/10.31219/osf.io/nc2rx.
Lawson, R.P., Mathys, C., Rees, G., 2017. Adults with autism overestimate the volatility of the sensory environment. Nat Neurosci 20(9), 1293-1299.
Lehrl, S., Triebig, G., Fischer, B., 1995. Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. Acta Neurologica Scandinavica 91(5), 335-345.
Li, C.T., Lai, W.S., Liu, C.M., Hsu, Y.F., 2014. Inferring reward prediction errors in patients with schizophrenia: a dynamic reward task for reinforcement learning. Front Psychol 5, 1282.
Mäki-Marttunen, V., Andreassen, O.A., Espeseth, T., 2020. The role of norepinephrine in the pathophysiology of schizophrenia. Neuroscience & Biobehavioral Reviews 118, 298-314.
Mathôt, S., Fabius, J., Van Heusden, E., Van der Stigchel, S., 2018. Safe and sensible preprocessing and baseline correction of pupil-size data. Behavior Research Methods 50(1), 94-106.
McElreath, R., 2020. Statistical rethinking: A Bayesian course with examples in R and Stan. CRC press.
Minzenberg, M.J., Poole, J.H., Benton, C., Vinogradov, S., 2004. Association of Anticholinergic Load With Impairment of Complex Attention and Memory in Schizophrenia. American Journal of Psychiatry 161(1), 116-124.
Moritz, S., Silverstein, S.M., Dietrichkeit, M., Gallinat, J., 2020. Neurocognitive deficits in schizophrenia are likely to be less severe and less related to the disorder than previously thought. World Psychiatry 19(2), 254-255.
Moritz, S., Woodward, T.S., 2007. Metacognitive training in schizophrenia: from basic research to knowledge translation and intervention. Current Opinion in Psychiatry 20(6), 619-625.
Murray, G.K., Cheng, F., Clark, L., Barnett, J.H., Blackwell, A.D., Fletcher, P.C., Robbins, T.W., Bullmore, E.T., Jones, P.B., 2008. Reinforcement and reversal learning in first-episode psychosis. Schizophrenia bulletin 34(5), 848-855.
Naicker, P., Anoopkumar-Dukie, S., Grant, G.D., Neumann, D.L., Kavanagh, J.J., 2016. Central cholinergic pathway involvement in the regulation of pupil diameter, blink rate and cognitive function. Neuroscience 334, 180-190.
Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in Ecology and Evolution 4(2), 133-142.
Nassar, M.R., Rumsey, K.M., Wilson, R.C., Parikh, K., Heasly, B., Gold, J.I., 2012. Rational regulation of learning dynamics by pupil-linked arousal systems. Nat Neurosci 15(7), 1040-1046.
Pearce, J.M., Hall, G., 1980. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review 87(6), 532-552.

Preuschoff, K., t Hart, B.M., Einhauser, W., 2011. Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. Front Neurosci 5, 115.
R Core Team, 2018. R: A language and environment for statistical computing. . R Foundation for Statistical Computing, Vienna, Austria.
Rajkowski, J., Kubiak, P., Aston-Jones, G., 1994. Locus coeruleus activity in monkey: Phasic and tonic changes are associated with altered vigilance. Brain Research Bulletin 35(5), 607-616.
Reddy, L.F., Waltz, J.A., Green, M.F., Wynn, J.K., Horan, W.P., 2016. Probabilistic Reversal Learning in Schizophrenia: Stability of Deficits and Potential Causal Mechanisms. Schizophr Bull 42(4), 942-951.
Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement., in: Black, A.H., Prokasy, W.F. (Eds.), Classical Conditioning II: Current Research and Theory. Appleton Century Crofts, New York, pp. 64–99.
Samuels, E.R., Szabadi, E., 2008. Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part II: physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. Current neuropharmacology 6(3), 254-285.
Schlagenhauf, F., Huys, Q.J.M., Deserno, L., Rapp, M.A., Beck, A., Heinze, H.-J., Dolan, R., Heinz, A., 2014. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. NeuroImage 89, 171-180.
Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. The Journal of clinical psychiatry.
Speechley, W.J., Whitman, J.C., Woodward, T.S., 2010. The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. Journal of Psychiatry & Neuroscience 35(1), 7-17.
Steinhauer, S., Zubin, J., 1982. Vulnerability to schziophrenia: Information processing in the pupil and event-related potential, in: Usdin, E., Hanin, I. (Eds.), Biological Markers in Psychiatry and Neurology. Pergamon, pp. 371-385.
Steinhauer, S.R., Hakerem, G., Spring, B.J., 1979. The pupillary response as a potential indicator of vulnerability to schizophrenia. Psychopharmacology Bulletin 15(1), 44-45.
van der Gaag, M., Hoffman, T., Remijsen, M., Hijman, R., de Haan, L., van Meijel, B., van Harten, P.N., Valmaggia, L., de Hert, M., Cuijpers, A., Wiersma, D., 2006. The five-factor model of the Positive and Negative Syndrome Scale II: A ten-fold cross-validation of a revised model. Schizophrenia Research 85(1), 280-287.
Waltz, J.A., Gold, J.M., 2007. Probabilistic reversal learning impairments in schizophrenia: Further evidence of orbitofrontal dysfunction. Schizophrenia Research 93(1), 296-303.
Waltz, J.A., Kasanova, Z., Ross, T.J., Salmeron, B.J., McMahon, R.P., Gold, J.M., Stein, E.A., 2013. The Roles of Reward, Default, and Executive Control Networks in Set-Shifting Impairments in Schizophrenia. PLOS ONE 8(2), e57257.
Weickert, T.W., Goldberg, T.E., Egan, M.F., Apud, J.A., Meeter, M., Myers, C.E., Gluck, M.A., Weinberger, D.R., 2010. Relative risk of probabilistic category learning deficits in patients with schizophrenia and their siblings. Biol Psychiatry 67(10), 948-955.
Worthy, D.A., Todd Maddox, W., 2014. A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W.K. Estes. Journal of Mathematical Psychology 59, 41-49.
Yu, A.J., Dayan, P., 2005. Uncertainty, Neuromodulation, and Attention. Neuron 46(4), 681-692.

Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J., Lamm, C., 2020. Using reinforcement learning models in social neuroscience: frameworks, pitfalls and suggestions of best practices. Social Cognitive and Affective Neuroscience. 15(6), 695-707.

Supplementary material: computational modelling

Choice of models

To quantify latent cognitive processes, various cognitive-computational models were fitted independently to participants' choices (i.e. predictions of 'left' or 'right') and observed outcomes (i.e. prediction correct or incorrect) for the volatile and the cued block of the prediction task, respectively, separately for individuals with a diagnosis from the schizophrenia spectrum (SZ group) and those without psychiatric diagnoses (HC group). The models included a simple win-stay-loose-shift model (WSLS), four different Reinforcement Learning models (RL), and two variants of a Hidden Markov Model (HMM) – all chosen to allow for the fact that participants might employ different strategies of solving the task. According to the WSLS model (Worthy and Todd Maddox, 2014) individuals would change their predictions each time they received a negative feedback on the previous trial, but stay with their previous prediction if it turned out to be correct. Note that in the prediction task, feedback was not explicit but reflected in whether the outcome on a given trial indeed corresponded to a participant's prediction. In contrast to the simplistic WSLS model, where choices (predictions) are merely based on the outcome of the previous trial, the RL models imply that prediction errors and choice values (i.e. values for left- and right-tilted) are integrated over a longer timescale. Four versions of RL models were tested, reflecting different assumptions of how this information is integrated. The Rescorla-Wagner model (RL_{RW}; Rescorla and Wagner, 1972) assumes value updating only for the action chosen (prediction of the left or the right-tilted Gabor patch) on a given trial based on a trial's prediction error weighted by a constant learning rate. In contrast, learning rates are allowed to differ for positive (prediction correct) vs. negative feedback (prediction incorrect) in the reward-punishment model (RL_{RP}; den Ouden et al., 2013), accounting for the fact that participants might learn differently from both types of feedback. In the counterfactual updating model (RL_{CF}; Gläscher et al., 2008), values for both available

actions are updated concurrently. This might account better for the anti-correlated task structure, where the probabilities for the left- and the right-tilted stimuli are inversely related. Given that participants' rate of learning might change during the course of a task block, an additional model based on Pearce and Hall (RL_{PH}; Pearce and Hall, 1980)) was fitted with an adaptive learning rate. Lastly, the HMM (Schlagenhauf et al., 2014), a Bayesian inference model, assumes a higher-order representation of the task structure that accounts for the instability of the task environment. Here, participants are expected to choose 'left' or 'right' depending on whether they believe to be in a hidden state where either the left- or the right-tilted stimulus constitutes the majority. States beliefs are inferred and updated on each trial, depending on the history of choice-outcome pairs as well as the estimated transition probability γ, which is the assumed probability for the two states of 'majority stimulus is left' and 'majority stimulus is right' to change. Thus, γ indicates a participant's perceived volatility of the task environment. To allow for the fact that positive (prediction correct) and negative (prediction incorrect) feedback may affect participant's belief updating differently, a model of the HMM where they were allowed to differ (HMM_{RP}) was contrasted against one where they were not (HMM).

All models are described in detail below. In addition, given that changes between risk conditions were announced in the cued task block, additional variants of all models were specified which incorporated choice value and state probability resets at each announced change point in this task block. In tables S1 – S4, these models are indicated by the suffix '_reset'.

Computational models

(I) Win-Stay-Loose-Shift model

The Win-Stay-Lose-Shift model (WSLS; Worthy and Todd Maddox, 2014) assumes that if participants were rewarded for their choice on a given trial (i.e. their prediction of either 'left'

or ‘right’ turned out to be correct), they continue to choose this option on the next trial (‘win-stay’). Similarly, if they were not rewarded (i.e. their prediction of either ‘left’ or ‘right’ turned out to be incorrect), they are expected to change their prediction and choose the other, previously non-selected option (‘lose-shift’). With the two possible actions *A* and *B*, the value of ‘staying’ with a choice after a ‘win’ is then calculated as:

$$V(A|A, win) = 1 \quad (1)$$

$$V(B|A, win) = -1 \quad (2)$$

Likewise, the value of ‘switching’ to the other choice option after a ‘loss’ is then:

$$V(B|A, loss) = 1 \quad (3)$$

$$V(A|A, loss) = -1 \quad (4)$$

(II) Standard Rescorla-Wagner model

In the Rescorla-Wagner reinforcement learning model ($RL_{RW;}$ Rescorla and Wagner, 1972), a constant learning rate drives the trial-wise value updates for the chosen option. For each trial t, the value V of the current choice is defined by the value and the prediction error δ (the difference between ‘reward’ and expected value) of the previous trial $t-1$, weighted by the learning rate α:

$$V_t = V_{t-1} + \alpha \times \delta \quad (5)$$

The prediction error, δ, is calculated as:

$$\delta = (R_{t-1} - V_{t-1}) \quad (6)$$

Importantly, ‘rewards’ in the prediction task were defined as correct predictions and assigned a value of +1. Since the experience of ‘rewards’ (i.e. correct predictions) and ‘punishments’ (i.e. incorrect predictions, assigned a value -1) might impact learning differently, a variant of the model was fitted where learning rates α for rewards and punishments were allowed to differ (model: $RL_{RP;}$ den Ouden et al., 2013).

(III) Counterfactual Reinforcement Learning model

Given the anti-correlated task structure, values for both the chosen and the unchosen option may be updated simultaneously. To account for that, a counterfactual updating model (RL_{CF}; Gläscher et al., 2008) was fitted to the data. The formula for the value update was the same as in (II), applied to both the chosen and the unchosen option, where a counterfactual prediction error was used for the unchosen option (uc):

$$\delta^{CF} = (-R_{t-1} - V_{t-1}^{uc}) \quad (7)$$

(IV) Pearce-Hall model

In the Pearce-hall model, the learning rate was dynamic (RL_{PH}; Pearce and Hall, 1980). Value updating was similar to the RL_{RW} model (see II), but the learning rate α varied across trials, depended on the previous prediction error:

$$\alpha_t = \gamma \times |(R_{t-1} - V_{t-1})| + (1 - \gamma) \times \alpha_{t-1} \quad (8)$$

(V) Hidden Markov Model

The Hidden Markov Model (HMM; Schlagenhauf et al., 2014) assumes that participants base their choices (i.e. predict either the left or the right Gabor patch) on their beliefs about the current task state, which in turn are modulated by the probability for those states to reverse (transition probability). Here, the different task states describe states where either the left stimulus ('state L') or the right patch is more common ('state R').

Participants are expected to infer the belief distribution over the different states from their observations of action-reward pairs (i.e. the combination of their prediction and the consequent 'reward', i.e. a correct or an incorrect prediction): $O_t = \{a_t, r_t\}$. A participant's estimation of such an action-outcome pair is then represented by the hidden state variable S_t.

In a transition matrix, the prior belief over the current state $P(S_t|S_{t-1})$ is calculated based on the posterior belief from the previous trial modulated by the transition probability γ, a free parameter between 0 and 1:

$$P(S_t|S_{t-1}) = \begin{pmatrix} 1-\gamma & \gamma \\ \gamma & 1-\gamma \end{pmatrix} \quad (9)$$

The probability of observing an outcome reflective of a given latent state depends further on the parameters c and d. Here, c is the probability with which a reward (i.e. a positive feedback in terms of a correct prediction) indicates that the true latent state indeed corresponds to the selected chosen option, i.e. the prediction made. Conversely, d is the probability with which a 'punishment' (i.e. a negative feedback in terms of an incorrect prediction) indicates that the true latent state does *not* correspond to the chosen option. The probability of observing a particular outcome given a particular state is then updated as:

$$P(O_t|S_t) = 0.5 \times \begin{pmatrix} c & 1-c \\ 1-d & d \end{pmatrix} \quad (10)$$

Here, c and d were free parameters, initialized to lie between 0.5 and 1. Similar to the RL_{RP} model (see II), this allowed for different effects of 'rewards' (positive feedback) and 'punishments' (negative feedback) in the model updating process. This version of the HMM is subsequently referred to as HMM_{RP}. An additional version of the model (subsequently referred to as HMM) was fitted where positive and negative feedback were treated equally, with $c = d$.

For both the HMM_{RP} and the HMM, the probability of S_t prior to any outcome observation is calculated for a given trial from the state transition probability (see above) and the posterior probability of S_{t-1}:

$$P(S_t) = \sum_{S_{t-1}} P(S_t|S_{t-1})P(S_{t-1}) \quad (11)$$

After the outcome has been observed, the posterior probability of S_t is updated based on the prior $P(S_t)$ and the outcome O_t:

$$P(S_t) = \frac{P(O_t|S_t)P(S_t)}{\sum_{S_t} P(O_t|S_t)P(S_t)} \tag{12}$$

Softmax action selection

Values were translated into choice probabilities for options *L* and *R* with a softmax action selector for all models (I) – (IV):

$$\mathrm{p(R)} = \frac{1}{1 - \mathrm{e}^{\beta \times (-(\mathrm{V_R} - \mathrm{V_L})}}, \qquad \mathrm{p(L)} = 1 - \mathrm{p(R)} \tag{13}$$

Here, the slope of the sigmoid function and the stochasticity (randomness) of the choices is determined by β, the inverse temperature. For the HMM (see V), state probabilities were used in the softmax function in place of values. Further, in this model the inverse temperature parameter was not included in order to reduce non-identifiable parameter estimation.

Model comparison:

A Hierarchical Bayesian Analysis (HBA; Gelman et al., 2013) was adopted from the hBayesDM package (Ahn et al., 2017) and implemented in the Stan language in R (RStan; Carpenter et al., 2017), to estimate model parameters. Models were fitted separately for both task blocks and both groups (SZ and HC). To compare modes regarding their goodness-of-fit to explain the observed data whilst accounting for model complexity, leave-one-out cross validation was conducted by using the log-likelihood evaluated at the posterior simulations. The results are reported with the leave-one-out information criterion (LOOIC; and corresponding effective number of parameters) in the tables below (S1 – S4), where lower LOOIC values indicate better model fit.

Table S1.
First, volatile task block: Model fit for SZ (n = 30)

Model	**LOOIC**	**no. of parameters**
WSLS	5746 (25)	1
RL_{RW}	5077 (42)	2
RL_{CF}	5024 (47)	2
RL_{RP}	5012 (55)	3
RL_{PH}	5071 (58)	4
HMM	4973 (48)	2
HMM_{RP}	**4813 (71)**	**3**

Table S2.
First, volatile task block: Model fit for HC (n = 30)

Model	LOOIC	no. of parameters
WSLS	5262 (27)	1
RL_{RW}	4332 (53)	2
RL_{CF}	4275 (49)	2
RL_{RP}	4154 (56)	3
RL_{PH}	4332 (62)	4
HMM	4120 (45)	2
HMM_{RP}	**3923 (68)**	**3**

Table S3.
Second, cued task block: Model fit for SZ (n = 30)

Model	LOOIC	no. of parameters
WSLS	5669 (26)	1
RL_{RW}	5146 (42)	2
RL_{RW_reset}	4406 (50)	2
RL_{CF}	4940 (51)	2
RL_{CF_reset}	4369 (48)	2
RL_{RP}	5020 (61)	3

RL_{RP_reset}	3948 (60)	3
RL_{PH}	5152 (51)	4
RL_{PH_reset}	4406 (51)	4
HMM	4902 (44)	2
$HMM_{\_reset}$	4315 (47)	2
HMM_{RP}	4254 (79)	3
HMM_{RP_reset}	**3826 (70)**	**3**

Table S4.
Second, cued task block: Model fit for HC (n = 30)

Model	**LOOIC**	**no. of parameters**
WSLS	5627 (27)	1
RL_{RW}	5237 (40)	2
RL_{RW_reset}	4907 (44)	2
RL_{CF}	4384 (43)	2
RL_{CF_reset}	4299 (42)	2
RL_{RP}	5133 (62)	3
RL_{RP_reset}	4240 (53)	3
RL_{PH}	5241 (61)	4

RL_{PH_reset}	4378 (54)	4
HMM	4839 (44)	2
$HMM_{\_reset}$	4225 (43)	2
HMM_{RP}	4502 (63)	3
HMM_{RP_reset}	**4088 (69)**	**3**

References (suppl.):

Ahn, W.Y., Haines, N., Zhang, L., 2017. Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. Comput Psychiatr 1, 24-57.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1).

den Ouden, Hanneke E.M., Daw, Nathaniel D., Fernandez, G., Elshout, Joris A., Rijpkema, M., Hoogman, M., Franke, B., Cools, R., 2013. Dissociable Effects of Dopamine and Serotonin on Reversal Learning. Neuron 80(4), 1090-1100.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian data analysis. CRC press.

Gläscher, J., Hampton, A.N., O'Doherty, J.P., 2008. Determining a Role for Ventromedial Prefrontal Cortex in Encoding Action-Based Value Signals During Reward-Related Decision Making. Cerebral Cortex 19(2), 483-495.

Pearce, J.M., Hall, G., 1980. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review 87(6), 532-552.

Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement., in: Black, A.H., Prokasy, W.F. (Eds.), Classical Conditioning II: Current Research and Theory. Appleton Century Crofts, New York, pp. 64–99.

Schlagenhauf, F., Huys, Q.J.M., Deserno, L., Rapp, M.A., Beck, A., Heinze, H.-J., Dolan, R., Heinz, A., 2014. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. NeuroImage 89, 171-180.

Worthy, D.A., Todd Maddox, W., 2014. A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W.K. Estes. Journal of Mathematical Psychology 59, 41-49.

Paper III

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 09 July 2020
doi: 10.3389/fpsyg.2020.01469

Objective Versus Subjective Effort in Schizophrenia

Isabel Kreis[1], Steffen Moritz[2] and Gerit Pfuhl[1]*

[1] Department of Psychology, Faculty of Health Sciences, UiT – The Arctic University of Norway, Tromsø, Norway,
[2] Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Background and Objectives: Performance on cognitive tasks is often impaired in individuals with schizophrenia (SCZ), possibly resulting from either cognitive deficits (e.g., limited working memory capacity) or diminished mental effort or both. Investment of mental effort itself can be affected by cognitive resources, task load, and motivational factors and has thus proven difficult to measure. Pupil dilation during task performance has been proposed as an objective measure, but it remains unclear to what extent this converges with self-reports of perceived task demands, motivation, and invested effort. The current study tried to elucidate this question.

Methods: A visual version of the digit span task was administered in a sample of 29 individuals with a diagnosis from the SCZ spectrum and 30 individuals without any psychiatric disorder. Pupil size was recorded during the task, whereas self-reported invested effort and task demand were measured afterward.

Results: No group difference was found for working memory capacity, but individuals with SCZ showed diminished trial-by-trial recall accuracy, showed reduced pupil dilation across all task load conditions, and reported higher perceived task demands.

Conclusion: Results indicate reduced effort investment in patients with SCZ, but it remains unclear to what extent this alone could explain the lower recall performance. The lack of a direct link between objective and subjective measures of effort further suggests that both may assess different facets of effort. This has important implications for clinical and research settings that rely on the reliability of neuropsychological test results when assessing cognitive capacity in this patient group.

Keywords: digit span, mental effort, task load, motivation, schizophrenia, pupillometry

OPEN ACCESS

Edited by:
Elisa Di Rosa,
Washington University in St. Louis, United States

Reviewed by:
Adam Culbreth,
Washington University in St. Louis, United States
Wing Chung Chang,
The University of Hong Kong, Hong Kong

****Correspondence:***
Isabel Kreis
isabel.v.kreis@uit.no

Specialty section:
This article was submitted to Psychopathology, a section of the journal Frontiers in Psychology

Received: *06 March 2020*
Accepted: *02 June 2020*
Published: *09 July 2020*

Citation:
Kreis I, Moritz S and Pfuhl G (2020) Objective Versus Subjective Effort in Schizophrenia. Front. Psychol. 11:1469. doi: 10.3389/fpsyg.2020.01469

INTRODUCTION

Working memory deficits are commonly reported in persons with schizophrenia (SCZ; e.g., Horan et al., 2008; Ventura et al., 2009; Freeman et al., 2014) and have been explained by a lack of processing resources (Nuechterlein and Dawson, 1984; Granholm et al., 1997). However, persons with SCZ, particularly when negative symptoms are prevalent, seem to be less willing to engage with physically (Gold et al., 2013; Barch et al., 2014; Bergé et al., 2018) or cognitively effortful tasks (Wolf et al., 2014; Gold et al., 2015; Culbreth et al., 2016; Reddy et al., 2018; Chang et al., 2019) and, when engaged, tend to exert less effort during task performance (Gorissen et al., 2005; Granholm et al., 2006, 2016). Accordingly, diminished performance on cognitive tests of

persons with SCZ might be explained not only by real cognitive impairments or limited resources but also by reduced invested effort (Gorissen et al., 2005). This has important implications for neuropsychological test situations in both clinical and research applications and led some authors to call for a combined assessment of neuropsychological performance and mental effort in persons with SCZ (Gorissen et al., 2005).

Mental effort has been described as the mediating processes between the theoretically achievable level of performance determined by task demands and cognitive capacity, and the actual level of performance achieved (Shenhav et al., 2017). These processes are affected by both cognitive and motivational factors, including personal goals, incentives, personality, and metacognitive knowledge (Fisher and Ford, 1998; Paas et al., 2005). Effort is inherently aversive and costly, as it requires the mobilization of energy (Gaillard, 1993; Fairclough and Houston, 2004; Shenhav et al., 2017). Hence, reduced effort exertion in persons with SCZ may be related to an overestimation of those (internal) costs (Gold et al., 2015; Shenhav et al., 2017) and could be related to a decreased tolerance of strain (van den Bosch and Rombouts, 1997). Measuring mental effort accurately has proven difficult. Studies investigating the willingness to exert effort often quantify this as choosing hard (high task demand) over easy tasks (low task demand) in favor of a larger monetary reward. Results may thus be confounded by subjective evaluation of monetary reward (see, e.g., Culbreth et al., 2016; Chang et al., 2019). In contrast, during standard neuropsychological assessments, no explicit external rewards are available, and patients usually cannot choose between hard and easy tasks. Measuring actual effort exertion in these contexts must therefore rely on different and more task-independent measures, for example, post-assessment self-reports (Moritz et al., 2017a). A more objective marker of mental effort exertion is pupil dilation during task performance (Granholm et al., 2016; van der Wel and van Steenbergen, 2018). The assumption that pupil dilation reflects effort allocation rests on the observation of positive correlations between pupil dilation and performance (Van Der Meer et al., 2010; Rondeel et al., 2015). Accordingly, smaller task-related pupil responses in persons with SCZ have been interpreted as an indication of reduced mental effort in SCZ and were found to be related to the severity of negative symptoms and defeatist attitudes (Granholm et al., 2006, 2016). Surprisingly, only a few studies investigated to what extent this objective measure of mental effort converges with self-reports of invested effort and motivation in these samples. Moreover, the role of subjectively perceived task demands and experienced strain remains unexplored, despite its likely detrimental role in effort investment (van den Bosch and Rombouts, 1997; Gold et al., 2015).

The current study aimed to investigate the relationship between working memory capacity, recall accuracy, pupil dilation, and subjective measurements of perceived task demands and motivated effort in a sample of participants with SCZ as compared to a sample without any psychiatric diagnosis. Participants with SCZ were expected to show smaller working memory capacity, recall accuracy, and pupil dilation as compared to participants without any psychiatric disorder across conditions of differing task demands. Further, patients were hypothesized to report higher strain caused by the task demands overall in combination with lower motivated effort. The self-report measures of strain and motivated effort were expected to correlate with the severity of negative symptoms.

MATERIALS AND METHODS

Inpatients and outpatients with a diagnosis from the SCZ spectrum were contacted directly and through the distribution of leaflets at the Department of Psychiatry and Psychotherapy of the University Medical Center Hamburg-Eppendorf (UKE), Germany. Healthy control participants were recruited through leaflets and posts on social media and student job websites. Participants had to meet the following inclusion criteria: (1) 18–65 years of age, (2) very good command of the German language, (3) IQ above 80, (4) capacity to give informed consent, (5) no substance dependence, (6) no recreational drug consumption within 1 week prior to the assessment (excluding alcohol, nicotine, and caffeine), (7) no history of neurological disorders, (8) normal or corrected-to-normal eyesight, and (9) a primary diagnosis of SCZ or schizoaffective disorder (SCZ group; DSM-V, American Psychiatric Association, 2013) or no psychiatric diagnosis at all (HC group). For all participants, written informed consent was obtained prior to the study. The study was approved by the local ethics committee of psychologists at the UKE.

This study was part of a larger project, and the total sample contained 61 participants. Only 59 of those completed the version of the digit span task and the corresponding motivation questionnaire as described here. Analyses of overall performance and questionnaires therefore rely on the data of 59 participants. For trial-wise analyses of pupil dilation and performance, another three participants were excluded due to large amounts of missing pupil data and technical difficulties during pupil recording.

Measures

Visual Digit Span Task

A visual, computerized version of the digit span task was administered. All stimuli were white on gray background. A trial started with the presentation of a fixation cross (4 s). A number of digits between one and nine were then shown one after another (1 s each), with a 1-s interstimulus interval. At the end of each trial, participants had to recall the digits in the order they were presented in and manually type in their responses on a standard keyboard. To keep the task as similar as possible to the standard forward digit span subtest of the Wechsler adult intelligence scale (WAIS-IV; Wechsler, 2008), the amount of digits presented in one trial increased over time: starting off with two digits, an additional digit was added after every second trial until the maximum amount of nine digits. Thus, for each load condition between two and nine, two trials were completed. During digit presentation, pupil size was recorded at a rate of 500 Hz with a desktop-mounted infrared video-based eye tracker (Eyelink 1000, SR Research).

Post-assessment Questionnaire

Self-reported motivation, invested effort, and subjective task demand were assessed after completion of the digit span task. The scales were newly compiled from items of the NASA Task Load Index (N-TLX; Hart and Staveland, 1988) and an authorized adaptation of items from the Momentary Influences, Attitudes and Motivation Impact on Cognitive Performance Scale (MIAMI; Moritz et al., 2017b) to cover topics such as motivation, invested effort, perceived task difficulty, and strain. In total, 17 items were posed on a Likert scale from 1 (completely disagree) to 4 (completely agree) (example items: "The task was very easy."; "I was very motivated.").

Clinical Assessments

Clinical diagnoses (SCZ group) or the absence thereof (HC group) was confirmed with the Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998). Positive and negative symptoms were assessed with the Positive and Negative Symptoms Scale (PANSS; Kay et al., 1987) within the SCZ group. Since the validity of the original PANSS dimensions has been criticized, particularly with regard to the negative symptoms scale (van der Gaag et al., 2006; Khan et al., 2013), negative symptom scores were calculated both according to the original publication (subsequently PANSS-N) and according to the scoring suggestions by van der Gaag et al. (2006; subsequently PANSS-N_{vdGaag}). As a proxy for premorbid intelligence, the German multiple choice vocabulary test (Lehrl et al., 1995) was administered.

Analysis

For overall analyses of working memory capacity, questionnaire responses, and clinical assessments, Spearman correlations and Mann–Whitney *U*-tests were used due to violated normality assumptions. Non-parametric effect sizes are reported as Cliff's delta d_C. For trial-wise analyses of recall accuracy, load condition, group membership, and pupil dilation, linear mixed regression models were built hierarchically and compared with the likelihood-ratio chi-squared test. For detailed model comparison and model parameters at each step, see **Supplementary Tables S1–S3**. All confirmatory testing was conducted with a significance level of 0.05, using the R programming language (R version 3.5.1, R Core Team, 2018).

Pupil Dilation Preprocessing

Eye blinks and artifacts were detected with a custom-built filter based on the pupil signal's velocity and removed through cubic-spline interpolation (Mathôt et al., 2018). The signal was then smoothed with a 3-Hz low-pass Butterworth filter, and periods of missing and aberrant data spanning more than 1000 consecutive milliseconds were treated as NA. Baseline pupil size for every trial was calculated as the mean pupil size of the 200 ms prior to the first digit. Percentage change in pupil size from baseline was then calculated for each sample of the trial. Baseline-corrected pupil dilation at each digit was then calculated by averaging the signal across the 1-s period after digit onset. Consistent with Granholm et al. (2016), the average pupillary response to the last digit presented in each trial was the main variable of interest. Only trials with less than 25% of missing data and where less than 50% of the signal used to calculate this main variable had been interpolated were submitted to subsequent analyses.

RESULTS

There were no significant group differences in any of the demographic variables or premorbid intelligence (see **Table 1**).

The SCZ group consisted of five inpatients and 24 outpatients. Thereof, 24 participants reported taking antipsychotic medications (83%; first generation: 1; second generation: 19; both first and second generations: 4). The mean percentage of the clinically recommended maximum dosage (Kane et al., 2003) was 60.94 (SD = 78.84). One participant took additional anticholinergic and 11 (38%) took other psychotropic drugs.

An exploratory factor analysis with varimax rotation revealed two subscales of the post-assessment questionnaire. The first one reflected perceived task demands and to what extent participants felt challenged and stressed (including items such as "In my opinion, the task was very difficult." and "I felt very stressed."). This scale included seven items and was labeled "ease" due to its reverse coding (i.e., lower values reflect higher experienced task demands). The possible score range was 7–28, and Cronbach's alpha was 0.82. The second scale reflected self-reported motivation and invested effort (including items such as "I was very motivated." and "I put in a lot of effort and gave it my best shot."). This scale encompassed eight items and was labeled "motivated effort" to distinguish it from effort driven by task demands (for full scales, see **Supplementary Material**). The possible score range was 8 to 32, and Cronbach's alpha was 0.81.

Overall Analyses: Maximum Digit Span and Correlation With Questionnaire Scales

General working memory capacity was assessed as the maximum number of correctly recalled digits in a row in the task overall, independent of load condition. The SCZ and the HC group only differed at a statistical trend (Md_{SCZ} = 6, Md_{HC} = 7; W = 551.1, p = 0.07, d_C = 0.27). Both groups reported similar motivated effort (Md_{SCZ} = 25, Md_{HC} = 28; W = 541.5, p = 0.11, d_C = 0.24). However, participants with SCZ reported smaller values for ease, i.e., they felt more challenged and strained by the task (Md_{SCZ} = 16, Md_{HC} = 19; W = 617.5, p = 0.01, d_C = 0.42).

There was a positive relationship between reported ease and maximum digit span across the whole sample (ρ = 0.26, p = 0.04) but no relationship between motivated effort and maximum digit span (ρ = 0.21, p = 0.12). Within the SCZ group, negative symptoms correlated neither with maximum digit span (PANSS-N: ρ = 0.03, p = 0.90; PANSS-N_{vdGaag}: ρ = 0.30, p = 0.13), ease (PANSS-N: ρ = 0.11, p = 0.57; PANSS-N_{vdGaag}: ρ = −0.03, p = 0.87), nor motivated effort (PANSS-N: ρ = 0.03, p = 0.89; PANSS-N_{vdGaag}: ρ = 0.05, p = 0.80). Ease and motivated effort were moderately correlated (ρ = 0.34, $p < 0.01$).

As anticholinergic agents can have detrimental effects on cognitive functions like working memory (Spohn and Strauss,

TABLE 1 | Sample demographics per group (total sample size = 59).

	SCZ (*n* = 29)			HC (*n* = 30)			*P*
	n	***M* (*SD*)**	***Md* (*IQR*)**	***n***	***M* (*SD*)**	**Md (*IQR*)**	
Gender (m/f)	14/15			13/17			0.90
Education ("1"/"2"/"3")	1/2/26			1/5/24			0.51
Age		47.55 (11.66)	51 (15)		45.80 (11.64)	47 (16.75)	0.57
WST		33.52 (3.54)	34 (4)		32.37 (4.55)	34 (6.25)	0.28
PANSS							
Positive Scale		12.07 (4.17)	11 (6)				
Negative Scale		10.41 (3.12)	10 (4)				
Negative Scale$_{vdGaag}$		12.59 (4.21)	12 (4)				
Total score		49.79 (14.24)	45 (15)				
Time since onset		19.38 (12.14)	18 (14)				
Inpatients/outpatients	5/24						

Sample sizes (n), counts, means (M; with standard deviations SD), and medians (Md; with inter-quartile ranges IQR) are displayed. Education was recorded in German school system categories corresponding to completion of 1 = secondary school I (up to age 15), 2 = secondary school II (up to age 16), 3 = sixth form college (up to age 19). WST, German vocabulary test. Negative Scale$_{vdGaag}$, negative symptom scoring according to van der Gaag et al. (2006). P-values for group comparisons are provided for the demographical variables gender and education (chi-squared tests) as well as age and the WST scores (T-test).

1989; Minzenberg et al., 2004) and affect pupil size (Naicker et al., 2016), benztropine mesylate equivalents, where available, were used to assess the anticholinergic load induced by each participant's daily dosage of the prescribed antipsychotics (Minzenberg et al., 2004). There was no difference in maximum digit span (W = 103, p = 0.98) or pupil dilation at the four-digit load condition, i.e., the load condition equivalent to the minimum digit span achieved in this sample (W = 69, p = 0.69), between participants who received an antipsychotic with a known anticholinergic effect ($Md_{digitspan}$ = 6, Md_{pupil} = 2.54, n = 16) and those who did not ($Md_{digitspan}$ = 6, Md_{pupil} = 1.89, n = 13). Anticholinergic load was correlated neither with maximum digit span (ρ = 0.26, p = 0.27, n = 20) nor with average pupil dilation at the four-digit load condition (ρ = 0.15, p = 0.59, n = 16). Similarly, the percentage of maximum dosage of all antipsychotics was not related to the maximum digit span (ρ = 0.11, p = 0.63, n = 23) or average pupil dilation at the four-digit load condition ($\rho = -0.10$, p = 0.67, n = 19).

Trial-Wise Analyses: Recall Accuracy

Trial-wise recall accuracy was measured as the percentage of digits recalled in the correct order on a given trial until the first error was made. To illustrate, within a load condition of eight digits, recall accuracy would be 50% if the first four digits were remembered correctly, but digits from the fifth digit onward were reported in an incorrect order. As seen in **Table 2**, average recall accuracy per load condition expectedly decreased with increasing load. This was confirmed by linear mixed regressions, which revealed main effects of load, $\chi^2(1) = 313.32$, $p < 0.001$, and group, $\chi^2(1) = 4.94$, $p = 0.03$, on recall accuracy, while the interaction between load and group was not significant, $\chi^2(1) = 2.23$, $p = 0.14$. In the winning model with only the two main effects, recall decreased as load increased, $b = -9.89$, $t = -22.11$, $p < 0.001$, and was lower in the SCZ group as compared to the HC group, $b = -6.56$, $t = -2.26$, $p = 0.03$.

TABLE 2 | Average percentage of items recalled in correct order per load condition for each group (N = 56).

Load	SCZ (*n* = 27)		HC (*n* = 29)	
	***M* (*SD*)**	***Md* (*IQR*)**	***M* (*SD*)**	**Md (*IQR*)**
2	100 (0)	100 (0)	100 (0)	100 (0)
3	97.9 (14.6)	100 (0)	100 (0)	100 (0)
4	94.3 (22.1)	100 (0)	98.7 (10.0)	100 (0)
5	90.9 (23.0)	100 (0)	91.6 (22.7)	100 (0)
6	67.1 (36.6)	83.3 (66.7)	81.2 (29.0)	100 (33.3)
7	49.3 (38.0)	35.7 (85.7)	62.5 (37.6)	71.4 (85.7)
8	42.3 (35.0)	25 (62.5)	47.5 (35.2)	37.5 (50)
9	35.1 (36.1)	22.2 (55.6)	43.5 (35.1)	38.9 (58.3)

Means (M; with standard deviations SD) and medians (Md; with inter-quartile ranges IQR) are displayed. Trials with NA entries for pupil dilation excluded per subject for comparability with regression models.

Trial-Wise Analyses: Pupil Dilation

As seen in **Figure 1**, in the HC group, trial-wise pupil dilation to the last digit increased with increasing processing load before it reached asymptote and decreased in higher load conditions. In contrast, this inverse U-shaped relationship was less prevalent in the SCZ group, and pupil dilation was smaller across almost all load conditions. These observations were confirmed by linear mixed regressions. Given the observed inverse U-shaped relationship between load and pupil dilation, both linear and quadratic load terms were tested as predictors. There was no significant effect for the linear load term, $\chi^2(1) = 0.95$, $p = 0.33$; the reverse was true for the quadratic one, $\chi^2(1) = 18.50$, $p < 0.001$. There was a significant main effect of group, $\chi^2(1) = 4.07$, $p = 0.04$. The interaction between load and group was not significant, $\chi^2(1) = 1.05$, $p = 0.31$, but the interaction between quadratic load and group indicated a trend, $\chi^2(1) = 2.89$, $p = 0.09$. In the winning model, which included the main effects only, both the linear and quadratic load terms were significantly

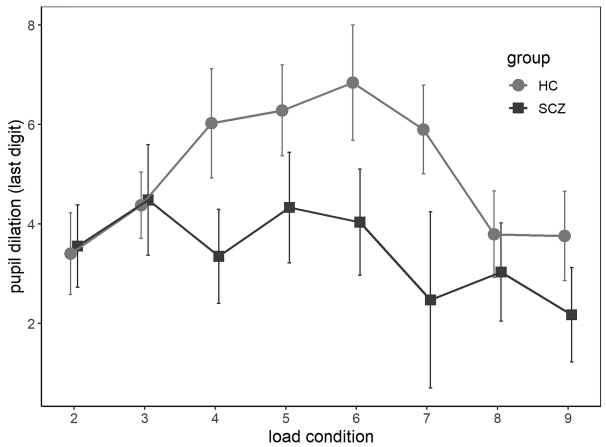


FIGURE 1 | Average baseline-corrected pupil dilation at the last digit of each load condition (2–9) for participants with diagnosis from the schizophrenia spectrum (SCZ group) and without (HC group). Error bars reflect standard errors of the mean.

related to pupil dilation, linear: b = 2.08, t = 4.06, p = < 0.001; quadratic: b = -0.20, t = -4.31, p = < 0.001. Further, participants with SCZ showed generally smaller pupil dilation across load conditions, b = -1.77, t = -2.04, p = 0.046. Notably, there was no group difference in baseline pupil size across all trials, $\chi^2(1) = 2.37$, $p = 0.12$.

Trial-Wise Analysis: Can Pupil Dilation at Last Digit Predict Recall Accuracy?

In another linear mixed regression analysis, the final model from Section "Trial-Wise Analyses: Recall Accuracy" was extended to establish if pupil dilation could predict variance in performance above and beyond the amount explained by load condition and group membership. Adding pupil dilation to the model indeed improved it significantly, $\chi^2(1) = 4.58$, $p = 0.03$. In this model, coefficients for load and group were consistent with the results of Section "Trial-Wise Analyses: Recall Accuracy," with performance decreasing as load increased, b = -9.86, t = -22.22, $p < 0.001$, and being lower in the SCZ as opposed to the HC group, b = -6.00, t = -2.04, p = 0.046. In line with an interpretation of pupil size as a measure of invested mental effort, larger pupil dilation predicted better performance, b = 0.32, t = 2.15, p = 0.03.

To test if this relationship was similar for all load and group conditions, interaction effects were added. The interaction term of load and group was not significant, $\chi^2(1) = 1.63$, $p = 0.20$, and therefore excluded from further models. However, the interactions between load and pupil dilation, $\chi^2(1) = 5.14$, $p = 0.02$, and between group and pupil dilation, $\chi^2(1) = 4.59$, $p = 0.03$, improved the model significantly. The final model therefore included load, group, pupil dilation, and the interactions between load and pupil, as well as group and pupil. Here, recall accuracy decreased with increasing load, b = -10.34, t = -20.62, $p < 0.001$, but in the presence of the interaction terms, there was no significant main effect for group, b = -3.05, t = -0.99, p = 0.33, or pupil dilation, b = -0.05, t = -0.11, p = 0.91. There was a meaningful trend for the interaction between load and pupil dilation, b = 0.12, t = 1.86, p = 0.06, indicating that the detrimental effect of load on performance was smaller on trials with larger pupil responses. Further, the interaction between group and pupil dilation was significant, b = -0.65, t = -2.16, p = 0.03, suggesting that pupil dilation was less predictive of performance in the SCZ as compared to the HC group.

Overall Analysis: Pupil Dilation and Subjective Effort in Max Span Condition

Linear mixed regression analyses for pupil dilation in the four-digit trials were conducted to explore the relationship between pupil dilation and the self-report questionnaire. This load condition was chosen because four was the minimum working memory capacity within the whole sample. Thus, a negative relationship between pupil dilation and maximum digit span within this condition would be expected as participants with more available cognitive resources would need to invest less effort (relative to their cognitive capacity) than persons with fewer resources. Adding self-reported motivated effort and perceived ease to the model while controlling for capacity and group would then give an indication to what extent pupil dilation is affected by motivational factors in addition. Since motivated effort and ease were correlated, two separate models were built. In the motivated effort model, only the group effect that had already being observed across all load conditions achieved marginal significance (b = -3.02, t = -1.97, p = 0.05, n = 54), but no effect of maximum digit span (b = -0.48, t = -0.90, p = 0.37, n = 54) or motivated effort (b = 0.04, t = 0.20, p = 0.84, n = 54) was found. Results from the ease model were similar, with no effects for maximum digit span (b = -0.36, t = -0.68, p = 0.50, n = 54) or ease (b = -0.26, t = -1.19, p = 0.24, n = 54), but smaller pupil dilation in the SCZ group (b = -3.49, t = -2.26, p = 0.03, n = 54). Within the SCZ group, the average pupil dilation in the four-digit trials was not related to negative symptoms (PANSS-N: ρ = 0.01, p = 0.95, n = 25; PANSS-N_{vdGaag}: ρ = -0.09, p = 0.68).

DISCUSSION

This study investigated the relationship between performance in a working memory task, self-reported motivated effort and ease, and objective effort allocation as indexed by pupil dilation in individuals with a clinical diagnosis from the SCZ spectrum (SCZ group) and individuals with no psychiatric disorder (HC group).

While there was no significant group difference in working memory capacity measured as maximum digit span, the SCZ group showed decreased recall accuracy on a trial-by-trial basis. The absence of a significant difference in maximum digit span may seem surprising, as working memory deficits in SCZ are well established. However, not all studies using the digit span task have replicated this finding (Park and Holzman, 1992; Franke et al., 1993). In the current study, participants had multiple opportunities to demonstrate their general working memory capacity, as performance in all trials were considered when assessing maximum digit span. In contrast, trial-by-trial assessment of recall accuracy may have been more sensitive to

momentary fluctuations in attention, which in turn might be affected by motivation (Engelmann et al., 2009). Given similar general capacity in both groups, at first glance, the differences in trial-wise performance seem more likely to have been caused by reduced effort rather than by a general lack of cognitive resources. In line with this, pupil dilation was reduced in the SCZ group across all load conditions, suggesting that participants with SCZ indeed invested less effort while doing the task. The inverse U-shaped relationship between load and pupil dilation was present across groups, though more prominent in the HC group, and can be interpreted as a detachment from the task at hand as task demands exceed available cognitive resources and thus decreasing expectations of success (Granholm et al., 2016). While some studies found group differences in pupil dilation only for high task demands (Granholm et al., 1997, 2006), others have reported differences across all demands, similar to the findings of this study (Granholm et al., 2016). Such discrepancies are likely the result of methodological differences and categorization of high and low demands. While the interaction effect between load and group on pupil dilation did not reach significance, the descriptive results suggest that pupil dilation was actually similar in trials where task load was below four digits (see **Figure 1**).

The interpretation of trial-wise pupil dilation as a measure of effort was supported by its positive relationship with trial-wise recall accuracy in a basic linear mixed regression model. In the regression model with interaction terms, recall accuracy of participants with larger pupil responses declined less as task load increased. Thus, increased task load can be compensated with an increase in invested effort. Nevertheless, the significant interaction between pupil dilation and group suggested that the positive relationship between pupil dilation and performance was smaller, if not absent, in the SCZ group. This makes it difficult to conclude if decreased trial-by-trial performance in this group can truly be attributed to less effort and proposes the role of additional explanatory factors. Interestingly, participants with SCZ reported feeling more challenged and stressed by the task, and this feeling of strain was correlated with maximum digit span and with motivated effort across the entire sample. On the one hand, it is likely that limited cognitive capacity leads to higher perceived task demands and strain. On the other hand, the cognitive resources available might not be exploited fully in situations where the task environment induces stress, which in turn may lead to an increase in perceived strain (Fairclough and Houston, 2004). Momentary sensitivity to stress has, in fact, been found to negatively affect cognitive functioning in SCZ (Morrens et al., 2007). Similarly, a generally reduced tolerance of strain in persons with SCZ could potentially explain the pattern of findings including heightened self-reported strain, smaller pupil dilation, and impaired recall accuracy across all load conditions (van den Bosch and Rombouts, 1997). This interpretation fits also well with the idea that persons with SCZ may invest less effort as a consequence of an overestimation of the costs associated with it (Gold et al., 2015; Shenhav et al., 2017). However, self-reported ease (i.e., reversed strain) did not predict pupil dilation in the four-digit trials and neither did self-reported motivated effort. Further, self-reported effort did not differ between groups, conflicting with the finding of smaller pupil dilation in SCZ across the task. This indicates little convergence between subjective and objective measures of effort, which may be linked in part to the way both constructs were measured (trial wise vs. after task completion) and to the fact that self-reports can be biased by lack of retrospective insight as well as social desirability.

None of our variables of interest correlated with negative symptom severity. This may seem unexpected, as previous studies have demonstrated a negative relationship between negative symptom severity and effort investment (e.g., Gorissen et al., 2005; Wolf et al., 2014) or that effort investment was predominantly impaired in subgroups scoring high on negative symptoms (Granholm et al., 2006; Bergé et al., 2018). However, other findings indicate that the relationship between effort investment and negative symptoms may, in fact, be non-linear and moderated by other factors, such as defeatist attitudes (Granholm et al., 2016; Reddy et al., 2018). Given the small sample size and the rather low average negative symptom score of the patient sample, no subgroups of high- and low-scoring patients were compared in the current study. The low scores were likely related to the large percentage of outpatients who tend to express fewer negative and other symptoms (e.g., Kasckow et al., 2001). Note that inconsistencies in findings regarding negative symptoms can further be related to the fact that measurement instruments differ across studies. The PANSS, which was chosen here, has received criticism for not reflecting the latest research results on negative symptoms (Kumari et al., 2017), which poses a limitation on the interpretability of the findings.

Further limitations of the study include the rather small sample sizes (particularly for the analyses including medication variables), the fact that medication was self-reported, the heterogeneity of the sample in terms of mixing in- and outpatients and including participants with schizoaffective disorders, as well as the possibility that matching groups by level of education may have contributed to the selection of an atypical, high-achieving group of persons with SCZ (Resnick, 1992). All of these factors may explain why some results from previous studies could not be replicated. The sample may have also been biased by the large proportion of chronically ill patients who, in turn, have been exposed to antipsychotic medication for long periods of their lives.

One potential limitation of the design is the fact that task load conditions were not randomized to ensure comparability with the standard version of the digit span subtest from the WAIS-IV (Wechsler, 2008). However, depletion or fatigue effects (Hagger et al., 2010) cannot account for the consistently smaller pupil dilation in SCZ across all load conditions. Another limitation is that subjective effort was only assessed after task completion with scales that have not been externally validated, although they were derived from well-validated measures.

Taken together, the findings of this study demonstrate once again the complex relationships between performance, effort, cognitive resources, and task demands. The results involving pupil dilation suggest that, in cognitive tasks, participants with SCZ might indeed exert less mental effort. However, it remains unclear to what degree this accounts for impaired momentary

performance in this sample and to what extent this is linked to the higher perceived strain imposed by task demands. To accurately judge the outcome of clinical or research-related neuropsychological assessments, these and other motivational factors have to be taken into account. Importantly, the lack of convergence between subjective and objective measures of effort might indicate that both objective and subjective measures can complement each other in unique ways and should thus be both considered for applications in this context.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**. Data as well as task and questionnaire material are available in an Open Science Framework repository: DOI 10.17605/OSF.IO/GCH97.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of psychologists at the University Medical Center Hamburg-Eppendorf, Hamburg, Germany. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version. IK, SM, and GP designed the study and edited the manuscript. IK collected and analyzed the data and wrote the manuscript.

FUNDING

This work was supported by a grant from The Research Council of Norway under Grant FRIMEDBIO 262338, including a personal overseas research grant to IK (Project No. 297347). The publication charges for this article have been funded by the publication fund of UiT The Arctic University of Norway.

ACKNOWLEDGMENTS

We thank all participants that dedicated their time and effort. We also thank the supportive staff of both the outpatient and the inpatient clinic of the Department of Psychiatry and Psychotherapy of the UKE and the members of the Clinical Neuropsychology Working Group of the Department of Psychiatry and Psychotherapy of the UKE for providing all necessary facilities and their indispensable guidance during clinical assessments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01469/full#supplementary-material

REFERENCES

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, 5 Edn. Washington DC: American Psychiatric Association.

Barch, D. M., Treadway, M. T., and Schoen, N. (2014). Effort, anhedonia, and function in schizophrenia: reduced effort allocation predicts amotivation and functional impairment. *J. Abnorm. Psychol.* 123, 387–397. doi: 10.1037/a0036299

Bergé, D., Pretus, C., Guell, X., Pous, A., Arcos, A., Pérez, V., et al. (2018). Reduced willingness to invest effort in schizophrenia with high negative symptoms regardless of reward stimulus presentation and reward value. *Compr. Psychiatry* 87, 153–160. doi: 10.1016/j.comppsych.2018.10.010

Chang, W. C., Westbrook, A., Strauss, G. P., Chu, A. O. K., Chong, C. S. Y., Siu, C. M. W., et al. (2019). Abnormal cognitive effort allocation and its association with amotivation in first-episode psychosis. *Psychol. Med.* 2, 1–11. doi: 10.1017/s0033291719002769

R Core Team (2018). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Vienna: R Core Team.

Culbreth, A., Westbrook, A., and Barch, D. (2016). Negative symptoms are associated with an increased subjective cost of cognitive effort. *J. Abnorm. Psychol.* 125, 528–536. doi: 10.1037/abn0000153

Engelmann, J., Damaraju, E., Padmala, S., and Pessoa, L. (2009). Combined effects of attention and motivation on visual task performance: transient and sustained motivational effects. *Front. Hum. Neurosci.* 3:4 doi: 10.3389/neuro.09.004.2009

Fairclough, S. H., and Houston, K. (2004). A metabolic measure of mental effort. *Biol. Psychol.* 66, 177–190. doi: 10.1016/j.biopsycho.2003.10.001

Fisher, S. L., and Ford, J. K. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Pers. Psychol.* 51, 397–420. doi: 10.1111/j.1744-6570.1998.tb00731.x

Franke, P., Maier, W., Hardt, J., Frieboes, R., Lichtermann, D., and Hain, C. (1993). Assessment of frontal lobe functioning in schizophrenia and unipolar major depression. *Psychopathology* 26, 76–84. doi: 10.1159/000284803

Freeman, D., Startup, H., Dunn, G., Černis, E., Wingham, G., Pugh, K., et al. (2014). Understanding jumping to conclusions in patients with persecutory delusions: working memory and intolerance of uncertainty. *Psychol. Med.* 44, 3017–3024. doi: 10.1017/S0033291714000592

Gaillard, A. (1993). Comparing the concepts of mental load and stress. *Ergonomics* 36, 991–1005. doi: 10.1080/00140139308967972

Gold, J. M., Kool, W., Botvinick, M. M., Hubzin, L., August, S., and Waltz, J. A. (2015). Cognitive effort avoidance and detection in people with schizophrenia. *Cogn. Affect. Behav. Neurosci.* 15, 145–154. doi: 10.3758/s13415-014-0308-5

Gold, J. M., Strauss, G. P., Waltz, J. A., Robinson, B. M., Brown, J. K., and Frank, M. J. (2013). Negative symptoms of schizophrenia are associated with abnormal effort-cost computations. *Biol. Psychiatry* 74, 130–136. doi: 10.1016/j.biopsych.2012.12.022

Gorissen, M., Sanz, J. C., and Schmand, B. (2005). Effort and cognition in schizophrenia patients. *Schizophr. Res.* 78, 199–208. doi: 10.1016/j.schres.2005.02.016

Granholm, E., Morris, S. K., Sarkin, A. J., Asarnow, R. F., and Jeste, D. V. (1997). Pupillary responses index overload of working memory resources in schizophrenia. *J. Abnorm. Psychol.* 106, 458–467. doi: 10.1037/0021-843x.106.3.458

Granholm, E., Ruiz, I., Gallegos-Rodriguez, Y., Holden, J., Link, P. C., and Gallegos-Rodriguez, Y. (2016). Pupillary responses as a biomarker of diminished effort associated with defeatist attitudes and negative symptoms in

schizophrenia. *Biol. Psychiatry* 80, 581–588. doi: 10.1016/j.biopsych.2015.08.037

Granholm, E., Verney, S. P., Perivoliotis, D., and Miura, T. (2006). Effortful cognitive resource allocation and negative symptom severity in chronic schizophrenia. *Schizophr. Bull.* 33, 831–842. doi: 10.1093/schbul/sbl040

Hagger, M. S., Wood, C., Stiff, C., and Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: a meta-analysis. *Psychol. Bull.* 136, 495–525. doi: 10.1037/a0019486

Hart, S. G., and Staveland, L. E. (1988). "Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.*, 52, 139–183. doi: 10.1016/s0166-4115(08)62386-9

Horan, W. P., Braff, D. L., Nuechterlein, K. H., Sugar, C. A., Cadenhead, K. S., Calkins, M. E., et al. (2008). Verbal working memory impairments in individuals with schizophrenia and their first-degree relatives: findings from the Consortium on the Genetics of Schizophrenia. *Schizophr. Res.* 103, 218–228. doi: 10.1016/j.schres.2008.02.014

Kane, J., Leucht, S., Carpenter, D., and Docherty, J. (2003). Expert consensus guidelines for optimizing pharmacologic treatment of psychotic disorders. *J. Clin. Psychiatry* 64(Suppl. 12), 5–19.

Kasckow, J. W., Twamley, E., Mulchahey, J. J., Carroll, B., Sabai, M., Strakowski, S. M., et al. (2001). Health-related quality of well-being in chronically hospitalized patients with schizophrenia: comparison with matched outpatients. *Psychiatry Res.* 103, 69–78. doi: 10.1016/s0165-1781(01)00260-8

Kay, S. R., Fiszbein, A., and Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* 13, 261–276. doi: 10.1093/schbul/13.2.261

Khan, A., Lindenmayer, J. P., Opler, M., Yavorsky, C., Rothman, B., and Lucic, L. (2013). A new Integrated Negative Symptom structure of the Positive and Negative Syndrome Scale (PANSS) in schizophrenia using item response analysis. *Schizophr. Res.* 150, 185–196. doi: 10.1016/j.schres.2013.07.007

Kumari, S., Malik, M., Florival, C., Manalai, P., and Sonje, S. (2017). An assessment of five (PANSS, SAPS, SANS, NSA-16, CGI-SCH) commonly used symptoms rating scales in schizophrenia and comparison to newer scales (CAINS, BNSS). *J. Addic. Res. Therapy* 8:324.

Lehrl, S., Triebig, G., and Fischer, B. (1995). Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *ACTA Neurol. Scand.* 91, 335–345. doi: 10.1111/j.1600-0404.1995.tb07018.x

Mathôt, S., Fabius, J., Van Heusden, E., and Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behav. Res. Methods* 50, 94–106. doi: 10.3758/s13428-017-1007-2

Minzenberg, M. J., Poole, J. H., Benton, C., and Vinogradov, S. (2004). Association of Anticholinergic Load With Impairment of Complex Attention and Memory in Schizophrenia. *Am. J. Psychiatry* 161, 116–124. doi: 10.1176/appi.ajp.161.1.116

Moritz, S., Klein, J. P., Desler, T., Lill, H., Gallinat, J., and Schneider, B. C. (2017a). Neurocognitive deficits in schizophrenia. Are we making mountains out of molehills? *Psychol. Med.* 47, 2602–2612. doi: 10.1017/s0033291717000939

Moritz, S., Stöckert, K., Hauschildt, M., Lill, H., Jelinek, L., Beblo, T., et al. (2017b). Are we exaggerating neuropsychological impairment in depression? Reopening a closed chapter. *Exp. Rev. Neurother.* 17, 839–846. doi: 10.1080/14737175.2017.1347040

Morrens, M., Krabbendam, L., Bak, M., Delespaul, P., Mengelers, R., Sabbe, B., et al. (2007). The relationship between cognitive dysfunction and stress sensitivity in schizophrenia. *Soc. Psychiatry Psychiatr. Epidemiol.* 42, 284–287. doi: 10.1007/s00127-007-0170-7

Naicker, P., Anoopkumar-Dukie, S., Grant, G. D., Neumann, D. L., and Kavanagh, J. J. (2016). Central cholinergic pathway involvement in the regulation of pupil diameter, blink rate and cognitive function. *Neuroscience* 334, 180–190. doi: 10.1016/j.neuroscience.2016.08.009

Nuechterlein, K. H., and Dawson, M. E. (1984). Information processing and attentional functioning in the developmental course of schizophrenic disorders. *Schizophr. Bull.* 10, 160–203. doi: 10.1093/schbul/10.2.160

Paas, F., Tuovinen, J. E., Van Merrienboer, J. J., and Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educ. Technology Research and Development* 53, 25–34. doi: 10.1007/bf02504795

Park, S., and Holzman, P. S. (1992). Schizophrenics show spatial working memory deficits. *Arch. Gen. Psychiatry* 49, 975–982. doi: 10.1001/archpsyc.1992.01820120063009

Reddy, L. F., Horan, W. P., Barch, D. M., Buchanan, R. W., Gold, J. M., Marder, S. R., et al. (2018). Understanding the association between negative symptoms and performance on effort-based decision-making tasks: the importance of defeatist performance beliefs. *Schizophr. Bull.* 44, 1217–1226. doi: 10.1093/schbul/sbx156

Resnick, S. M. (1992). Matching for education in studies of schizophrenia. *Arch. Gen. Psychiatry* 49, 246–246.

Rondeel, E., Van Steenbergen, H., Holland, R., and van Knippenberg, A. (2015). A closer look at cognitive control: differences in resource allocation during updating, inhibition and switching as revealed by pupillometry. *Front. Hum. Neurosci.* 9:494.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 59 (Suppl 20), 22–57.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., et al. (2017). Toward a Rational and Mechanistic Account of Mental Effort. *Annu. Rev. Neurosci.* 40, 99–124. doi: 10.1146/annurev-neuro-072116-031526

Spohn, H. E., and Strauss, M. E. (1989). Relation of neuroleptic and anticholinergic medication to cognitive functions in schizophrenia. *J. Abnorm.l Psychol.* 98, 367–380. doi: 10.1037/0021-843X.98.4.367

van den Bosch, R. J., and Rombouts, R. (1997). Coping and cognition in schizophrenia and depression. *Compr. Psychiatry* 38, 341–344. doi: 10.1016/S0010-440X(97)90930-5

van der Gaag, M., Hoffman, T., Remijsen, M., Hijman, R., de Haan, L., van Meijel, B., et al. (2006). The five-factor model of the Positive and Negative Syndrome Scale II: a ten-fold cross-validation of a revised model. *Schizophr. Res.* 85, 280–287. doi: 10.1016/j.schres.2006.03.021

Van Der Meer, E., Beyer, R., Horn, J., Foth, M., Bornemann, B., Ries, J., et al. (2010). Resource allocation and fluid intelligence: insights from pupillometry. *Psychophysiology* 47, 158–169. doi: 10.1111/j.1469-8986.2009.00884.x

van der Wel, P., and van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: a review. *Psycho. Bull. Rev.* 25, 2005–2015. doi: 10.3758/s13423-018-1432-y

Ventura, J., Hellemann, G. S., Thames, A. D., Koellner, V., and Nuechterlein, K. H. (2009). Symptoms as mediators of the relationship between neurocognition and functional outcome in schizophrenia: a meta-analysis. *Schizophr. Res.* 113, 189–199. doi: 10.1016/j.schres.2009.03.035

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV)*, Vol. 22. San Antonio, TX: NCS Pearson, 816–827.

Wolf, D. H., Satterthwaite, T. D., Kantrowitz, J. J., Katchmar, N., Vandekar, L., Elliott, M. A., et al. (2014). Amotivation in schizophrenia: integrated assessment with behavioral, clinical, and imaging measures. *Schizophr. Bull.* 40, 1328–1337. doi: 10.1093/schbul/sbu026

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.