MONASH University

Self in Autism:
A Predictive Perspective

Kelsey Savita Perrykkad

A thesis submitted for the degree of Doctor of Philosophy

at Monash University in 2021.

Philosophy Department

School of Philosophical, Historical and International Studies

Faculty of Arts, Monash University

Copyright Notice

Table of Contents

Copyright Notice i

Table of Contents ii

Abstract iii

Publications During Enrolment iv

Thesis including published works declaration v

Acknowledgements vii

Preface to the Thesis 1

Chapter 1. Modelling Me Modelling You: The Autistic Self 11

Chapter 2. Adaptive Behaviour and Predictive Processing Accounts of Autism 48

Chapter 3. Are differences in self-cognition a characteristic feature of autism? Evidence from psychiatric traits, self-concept and shape-label matching 54

Chapter 4. Fidgeting as Self Evidencing: A predictive processing account of non-goal-directed action 91

Preface to The Squares Task: Chapters 5-7 102

Chapter 5. The Effect of Uncertainty on Prediction Error in the Action-Perception Loop 109

Chapter 6. Judgements of Agency and Block-wise Volatility: A Pilot Study 123

Chapter 7. The Beach Task: Environmental Niche Selection Under Uncertainty 151

Chapter 8. When Big Data Aren't The Answer 198

Discussion and Conclusion 205

References in Thesis Text 211

Addendum 1. Chapter 5 Supplementary Materials 216

Abstract

Historically, accounts of the self were intimately related to the soul and other essential and immovable aspects of our character. In contrast, modern conceptions from cognitive science suggest that representations of the self are subject to many of the same cognitive processes as other representations. The self, however, is one of our most important representations and has a unique, reflexive character. Our self-representation shapes how we act in the world, and the feedback we receive in turn shapes how we represent ourselves. The reciprocal interaction between the environment and ourselves is essential to how our cognitive system actively builds and maintains the self over our lifetime.

Previous theories have suggested that the way individuals on the autism spectrum build and maintain representations of themselves differs from neurotypical individuals. This research has been motivated by social deficit theories of autism. In this thesis, I investigate the self in autism anew from primarily sensory and cognitive perspectives, leveraging concepts from the predictive processing framework, which offers new approaches and insights. Under this framework, autism is characterised by differences in modelling or predicting the world under uncertainty (which impacts both perception and action).

The thesis adopts an interdisciplinary approach to the question of how autistic self-cognition differs from neurotypical self-cognition. The theoretical chapters are based in conceptual, analytic and argumentative methods from philosophy of cognitive science. These include arguments about the interpretation of evidence about autistic self-cognition from many cognitive domains, the best way to define and explain autism, how to provide in-depth predictive processing accounts of perplexing aspects of self related processing, and the limitations on conclusions we can draw from the tools we use to measure autistic traits. The experimental chapters operationalise three different aspects of self-cognition and look at their relation to autistic traits. These are self-concept clarity as measured by two self-report questionnaires, self-prioritisation in early processing of arbitrary shape-label pairs, and in several experiments, judgement of agency in environments with uncertainty in the mapping between actions and their outcomes. These experiments also give us a deeper insight into self-cognition in the neurotypical case, and highlight just how much we still have to learn.

Findings from the thesis show that individuals with more autistic traits are more prone to act early in the face of rising uncertainty. This may result from a self-model with less hierarchical depth in autism – with more resources at lower, sensory parts of the cognitive hierarchy and less capacity at more integrated levels. The thesis also raises questions about the appropriate core features of autism. The body of work demonstrates that experimental psychology and philosophy can work in tandem to teach us about the nature of the self, autism, and the self in autism.

Publications During Enrolment

***Perrykkad, K.**, Lawson, R. P., Jamadar, S., & Hohwy, J. (2021). The effect of uncertainty on prediction error in the action perception loop. *Cognition, 210*, 104598.

***Perrykkad, K.**, Hohwy, J. (2020) Fidgeting as Self-evidencing: a predictive processing account of non-goal-directed action. *New Ideas in Psychology*, 56,100750.

***Perrykkad, K.**, Hohwy, J. (2020) Modelling Me, Modelling you: the Autistic Self. *Review Journal of Autism and Developmental Disorders*, 1-31.

***Perrykkad, K.**, Hohwy, J. (2019) When Big Data Aren't The Answer. *Proceedings of the National Academy of Sciences*, 201902050.

Wilson, W. J., Downing, C., **Perrykkad, K.**, Armstrong, R., Arnott, W. L., Ashburner, J., & Harper-Hill, K. (2019). The 'acoustic health'of primary school classrooms in Brisbane, Australia. *Speech, Language and Hearing*, 1-8.

Carroll, A., Gillies, R.M., Cunnington, R., McCarthy, M., Sherwell, C., **Palghat, K.**, Goh, F., Baffour, B., Bourgeois, A., Rafter, M., & Seary, T. (2019). Multimodal, cooperative interventions: Changes in science attitudes, beliefs, knowledge and physiological arousal. *Information and Learning Sciences*, 120(7/8), 409-425.

van de Cruys, S., **Perrykkad, K.**, Hohwy, J. (2019) Explaining hyper-sensitivity and hypo-responsivity in autism with a common predictive coding-based mechanism. *Cognitive Neuroscience*, 1-2.

***Perrykkad, K.** (2019) Adaptive Behaviour and Predictive Processing Accounts of Autism. *Brain and Behavioural Sciences*, 42.

van der Kruk, Y., Wilson, W. J., **Palghat, K**., Downing, C., Harper-Hill, K., & Ashburner, J. (2017). Improved Signal-to-Noise Ratio and Classroom Performance in Children with Autism Spectrum Disorder: a Systematic Review. *Review Journal of Autism and Developmental Disorders*, 1-11.

Palghat, K., Horvath, J. C. and Lodge, J. M. (2017) The hard problem of 'educational neuroscience'. *Trends in Neuroscience and Education*, 6: 204-210.

***** = Included in the thesis

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes five original papers published in peer reviewed journals (including three original studies and two original commentaries) and one submitted publication. The core theme of the thesis is self-cognition in autism spectrum conditions from the perspective of predictive processing. The ideas, development and writing of all the papers in the thesis were the principal responsibility of myself, Kelsey Perrykkad, working within the School of Philosophical, Historical and International Studies under the primary supervision of Jakob Hohwy, and secondary supervision of Sharna Jamadar of the Turner Institute for Brain and Mental Health.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research. Assistance in data collection for Chapter 6 was provided by research assistants, Disha Sasan and Alex Wulff. Jonathan Robinson translated my experimental code from Chapters 5 and 6 to run in an online environment for Chapter 7. Rebecca Lawson provided guidance on Chapters 5 and 6 and provided example eye-tracking code.

I have renumbered pages but have not renumbered sections of submitted and published papers for consistent presentation within the thesis.

Student name: Kelsey Perrykkad

Student signature: *[signature removed in final version]* **Date:** 29/03/21

I hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author, I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor name: Jakob Hohwy

Main Supervisor signature: *[signature removed in final version]* **Date:** 29/03/21

In the case of published and submitted chapters 1-5 and 8, my contribution to the work involved the following:

Thesis Chapter	Publication Title	Status *(published, in press, accepted or returned for revision, submitted)*	Nature and % of student contribution	Co-author name(s) Nature and % of Co-author's contribution	Co-author(s), Monash student Y/N
1	Modelling Me, Modelling You: the Autistic Self	Published	90%. Concept, literature collection and analysis and writing.	(1) Jakob Hohwy, concept and editing. Input into manuscript: 10%	No
2	Adaptive Behaviour and Predictive Processing Accounts of Autism	Published	100%		
3	Are differences in self-cognition a characteristic feature of autism? Evidence from psychiatric traits, self-concept and shape-label matching	Submitted	90%. Concept, method, data collection, data analysis and writing	(1) Jakob Hohwy, concept and editing. Input into manuscript: 10%	No
4	Fidgeting as self-evidencing: A predictive processing account of non-goal-directed action	Published	90%. Concept, literature collection and analysis and writing.	(1) Jakob Hohwy, concept and editing. Input into manuscript: 10%	No
5	The Effect of Uncertainty on Prediction Error in the Action Perception Loop	Published	70%. Concept, Method, Task Coding, Data Collection, Data Analysis, and writing.	(1) Rebecca P. Lawson, concept, method, analysis assistance and editing. Input into manuscript: 10% (2) Sharna Jamadar, method, data collection support, and editing. Input into manuscript: 10% (3) Jakob Hohwy, concept, method and editing. Input into manuscript: 10%	No No No
8	When big data aren't the answer	Published	90%. Concept and writing.	(1) Jakob Hohwy, concept and editing. Input into manuscript: 10%	No

Acknowledgements

The people I would like to thank are innumerable. I simply cannot name all of the people without whom I could not have produced this thesis, but I will do my best. To everyone who made me feel welcome and valued in the academic community, you played more of a role than you know. The image of the lone philosopher in their ivory tower could not be further from my experience of research life. One thing I have learned in my years of academia so far is that collaboration is what makes all the hard work manageable and fun. I hope to carry this collaborative spirit forward to my future, both in and out of academia.

I would like to thank all my participants, without whom none of the experimental work would be possible. I would also like to thank members of the autistic community who share their experiences publicly, on twitter, and in autobiographies. Reading your writing was invaluable in writing this thesis. I promise to reach out in a more dialogic way for future research.

My supervisors, Jakob Hohwy and Sharna Jamadar, have been amazing mentors to me. I thank both of them for their advice and encouragement over the PhD, even in the midst of a life-altering global pandemic they were both easily accessible and supportive. I look forward to working with both of you in the future. Also to Becky Lawson, I am honoured to have been able to work with you from across the globe. Thanks for making me feel so welcome and valued whenever I visited Cambridge; and thank you for your ongoing long-distance support. Jonno, you stepped up in the middle of the pandemic to dedicate much of your time help me out of a sticky situation – I will forever be grateful for that level of care in an uncertain time. I also appreciate the help of Disha and Alex with data collection for the pilot study.

In my academic journey, I have been very fortunate to have many mentors along the way who have made me the researcher I am today. From Oxy, I thank Professors Morrissey, Brighouse and Levitan – a powerhouse of three fantastic female mentors who stoked the early sparks into fires of curiosity and passion about what I do now. Thanks also to Simmy Poonian, Jeff Bednark, and Ross Cunnington for introducing me to agency research and for believing in my potential.

One of the most enjoyable things about my PhD candidature was the phenomenal community of philosophers, neuroscientists and psychologists who supported me on my journey in Melbourne, Brisbane and around the world. The journey would have been much more difficult without you all by my side. I will always remember conversations in the lunchroom (or on zoom after a lab meeting, or over a beer at the Jazz Club, or celebrating someone’s milestone at the Nott) with the members of the Cognition and Philosophy lab. It is a rare and special group of incredible people, and I hope to stay in touch with all of you for years to come. In particular, I would like to thank my closest academic brothers, Andrew Corcoran and Stephen Gadsby, and my closest academic sister, Manuela Kirberg. I could message you at any time, day or night, and you were there to hear my rantings. Thanks to Noam Gordon, for always being in the office for the first few months of my PhD, and helping

me establish a strong routine. To Julian Matthews, for our shared love of all things aesthetic. Special mention too to Andy McKilliam - my critical thinking doppelganger. And to all the other lab members, Iwan, Niccolo, Ricardo, Simon, Rafik, Mateusz, Roger, Jasmine, James, we have created a special environment to learn about the workings of the mind together. To the many visiting students to the Cognition and Philosophy Lab, who brought new ideas and perspectives and kept things interesting and alive, I thank you too. Debriefs in the park while painting with Manja Engel were irreplaceable.

To the members of the Thesis Consolation group, Chase Sherwell, Megan Campbell and Natalie Rens, thanks for consoling me even after your theses were long completed and for teaching me what it means to share a lab. To my Tuesday dinner group – thanks for talking about things that had nothing to do with academia once a week throughout the whole PhD.

My family is everything to me, and I thank them for being there every step along the way. To Aunt Kelly, for proofreading the final version of my thesis – one day, I will get a dog and name it Bonferroni! To grandma, for your endless intellectual spirit. My dog, Apollo, is a great stress reliever, and he was always up for a good cuddle and some playtime at the end of a long day. Mom and Dad, thanks for listening to all the intricate details about when an experiment went wrong (or right), always looking out for my stress levels (and heart rate), and always telling me you knew I could do it and how proud you were of me. To Kendall, thanks for helping me make decisions when I felt overwhelmed, and doing it with style ("Glass of champagne anyone? She made a decision!"). The core is strong; together, we can do anything.

And finally, endless appreciation to my husband, Andrew. Thank you for being there to listen to all the boring details at the end of every day, on the commute home, or just on the couch with a glass of wine or whiskey. Thanks for sitting in the study, piloting my experiment late into the night after a long day of your own PhD struggles. Thanks for always sharing your birthday with conferences. I am so glad that we decided to embark on this together; the (future) Doctors Perrykkad. I can't wait to see what happens next.

This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

Preface to the Thesis

Aims and scope

This thesis combines three key areas of research: Autism Spectrum Conditions, Self-Cognition, and the Predictive Processing framework. I use tools from both cognitive science and philosophy to answer the following research question: How does self-cognition differ in Autism Spectrum Condition? The thesis contains four experimental chapters (Chapters 3, 5-7), and four theoretical chapters that do not present new empirical data, but refer to or review existing literature to build an argument (Chapters 1-2, 4, 8). The theoretical chapters are not systematic reviews or meta-analyses, but rather develop concepts through analysis, argumentation, and conceptual interpretation of existing empirical data as evidence employing the methodology of philosophy of cognitive science. I will begin by situating these three key areas and outlining the thesis' progression.

Autism Spectrum Condition

Autism spectrum condition (ASC, autism), affects the way an individual perceives and interacts with the world and others. Recent estimates in Australia suggest that 2.4-3.9% of the population has autism (May, Sciberras, Brignell, & Williams, 2017). Being autistic manifests diversely both in terms of severity (or impact on daily needs) and in terms of the domains of life it most impacts. For instance, while some people on the spectrum cannot communicate verbally at all, others merely have difficulties interpreting semantically ambiguous phrases whose meaning is derived from more complex social and contextual cues. While in many cases autism is associated with high levels of distress, many autistic traits do not necessitate help-seeking and therefore many individuals on the spectrum may go without diagnosis. According to the latest Diagnostic and Statistical Manual (DSM-5), the current official symptoms include difficulties with social-emotional reciprocity; non-verbal communication; relationships (initiation, comprehension and maintenance); repetitive movements, use of objects or speech; rigid adherence to routines and insistence on sameness; restricted intense interests; and atypical sensory responses (American Psychiatric

Association, 2013). While autism is a lifelong condition, it is often referred to as developmental due to its early onset and particularly strong effects in childhood (including delayed developmental milestones).

In this thesis, I will prefer the term Autism Spectrum Condition (ASC, or simply, autism). This choice in term is intentional. It emphasises the diversity of symptom presentation between individuals and a broad range of symptom severity through the use of the word spectrum (as opposed to the historical usage of Autistic Disorder, for example). And further, it is normatively neutral with respect to the net value or deficit implied by the cognitive differences it delineates. I also opt for the identity-first label, ‘autistic’, to person-first language, such as ‘person with autism’, in this thesis in response to the preferences of the autistic community themselves (Bury, Jellett, Spoor, & Hedley, 2020).

Given the diversity in the way autism presents across individuals, and the disparate cognitive domains in the diagnostic criteria, Autism Spectrum Condition proves a particularly nebulous target for a unified cognitive theory. For an overview of the current landscape of autism research including the autistic perspective, see Fletcher-Watson and Happé (2019). Previous prominent theories of autism include social deficit accounts such as the theory of mind theory (Baron-Cohen, Leslie, & Frith, 1985), the social orienting and social motivation theories (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012); and information processing theories such as the weak central coherence account (Happé, 1999; Happe & Frith, 2006), executive dysfunction theories (Ozonoff, Pennington, & Rogers, 1991), and the extreme-male brain or systematising theory (Baron-Cohen, 2002). Where the social deficit accounts often fail to account for restricted and repetitive behaviours and interests (RRBI), the information processing theories struggle to explain the clinically salient social difference-makers in autism, which has lead some to suggest that different parts of the diagnostic criteria may be underwritten by independent cognitive mechanisms (Frith & Happé, 1994; Happe & Frith, 2006).

However, if prevailing clinical ontology is correct, and ‘autistic’ is a coherent label that identifies a co-occurring cluster of features, then a unified theory is preferable. Splitting theories of autism into component mechanisms that explain this criterion or that criterion but do not explain why they frequently co-occur instead supports splitting of the current diagnosis into two or more distinct conditions. This would, in effect, destroy or dramatically alter the category all together. While this might ultimately prove the preferable route, a parsimonious theory of autism should address all the criteria and why they regularly co-occur.

Self-cognition

It is no accident that trying to understand the self has been central to practices in both philosophy and cognitive science since the beginning. What kind of individual are we? Is there an unmovable, everlasting part of us? Do we change over time? How much? Why do we act in certain ways or display certain traits that are different from other people? How do we build narratives about our lives? How we represent ourselves, including how stable we see ourselves to be and how effectively we can affect intended changes on the world, can have huge impacts on mental wellbeing.

Philosophers, like David Hume, have carefully analysed the self, many of them questioning its very existence:

> But self or person is not any one impression, but that to which our several impressions and ideas are supposed to have a reference. … But there is no impression constant and invariable. … It cannot, therefore, be from any of these impressions, or from any other, that the idea of self is derived; and consequently there is no such idea. … For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe anything but the perception. … The identity, which we ascribe to the mind of man, is only a fictitious one …
>
> *(Hume, 1741, pp. Book I, Part IV Sect. VI)*

Psychologists too, have been thinking about and researching selves for a very long time. William James developed a multi-faceted understanding of the self:

> The consciousness of Self involves a stream of thought, each part of which as 'I' can 1) remember those which went before, and know the things they knew; and 2) emphasize and care paramountly for certain ones among them as *'me'* and *appropriate to these* the rest. The nucleus of the *'me'* is always the bodily existence felt to be present at the time. Whatever remembered-past-feelings *resemble* this present feeling are deemed to belong to the same *me* with it. Whatever other things are perceived to be associated with this feeling are deemed to form part of that me's *experience;* and of them certain ones (which fluctuate more or less) are reckoned to be themselves *constituents* of the me in a larger sense… This me is an empirical aggregate of things objectively known… The same brain may subserve many conscious selves, either alternate or coexisting …
>
> *(James, 1890, pp. 400-401)*

I take a very cognitive approach to understanding the self. In some ways it is a mental representation unlike others. It develops over the longest timescale (integrating experiences over one's entire life) and is perhaps uniquely reflexive, in that the vehicle of the representation is what is represented. On the other hand, in being a mental representation, it is subject to the same cognitive processes as any other mental representation. Attempts to study the neural basis of the self have yielded disperse and non-specific networks (Gillihan & Farah, 2005). This reflects the fact that self-cognition is integrated with many other cognitive processes and uncovering its nature can be quite complex.

The self is represented in many cognitive domains (see Chapter 1 for more detail). Involvement of self-specific stimuli in very low-level sensory domains can enhance sensory processing. Consider for example the classic cocktail party effect where hearing your own name in a noisy room elicits instant attention (Moray, 1959). Perhaps surprisingly, there is also attentional advantage for self-referring labels when they are assigned arbitrarily to meaningless stimuli such as shapes (Sui, He, & Humphreys, 2012)(see Chapter 3). As James suggests, relevance to the self also has impacts on memory as seen in the self-reference effect, where self-referents such as personality traits that apply to oneself have a recall advantage (Symons & Johnson, 1997). Autobiographical memories are also an important aspect of establishing a sense of self that has temporal stability. The self also involves recognition of one’s own image, bodily representations and recognising the sensory consequences of one’s actions as such (judgement of agency – see Chapters 5-7). Of course, part of our self-representation is the explicit self-concept – how we talk and think about our own self (see Chapter 3). There are then many ways to operationalise the self for experimental research in cognitive science.

Self-cognition in autism

While self-cognition might be inextricably intertwined with some of the other core features of autism (which will become apparent in Chapters 1 and 4), it is a relatively neglected area of autism research. In some ways, turning away from the focus on autism as a social disorder to a focus on what it is like to be a self with autism is a return home. The word “autism” is derived from the greek, *autos*, which means self. While originally named this way for the apparent withdrawal or self-sufficiency of autistic children (as described by Kanner (1943)), there has previously been some focus on differences in self-cognition in autism.

The previous approach to differences in the self in autism (Frith & Happé, 1999; Lombardo & Baron-Cohen, 2010; Uddin, 2011; Williams, 2010), as we will see in Chapter 1, has been motivated primarily by the theory of mind theory of autism. In this theory, autism is explained by an inability to attribute independent mental states to others (Baron-Cohen, 2000; Baron-Cohen et al., 1985). So too then, the story goes, because of failures of inference about minds in general, autistic individuals might have difficulty accurately representing themselves (Frith & Happé, 1999). While on the right track, taking this folk psychological route has some limitations.

At its core, the theory of mind theory of autism (and the autistic self) postulates differences in forming inferences. Making inferences about minds involves selecting between different models of the world based on the available sensory evidence. This is what we do to make inferences about many other things besides minds. What if we could go beyond mere folk psychology and instead understand the autistic self as based in differences in inferential processes cast in a general, computational, cognitive framework? I argue that turning to the predictive processing framework will give us unique insights as we move forward to understand the self in autism.

Predictive Processing

Predictive processing is an increasingly popular framework used to understand the mind. The most devoted proponents of predictive processing (particularly of the associated ‘free energy principle’) will argue that it is not just a theory of the mind, but life and, even broader, any physical process. Others go the opposite way. Those who tend to use the term ‘predictive coding’ restrict their focus to small populations of neurons involved in perception; usually vision. In this thesis, I use the broader term ‘predictive processing’. This captures both perceptual and active inference, but seeks to be agnostic about the involvement of the free energy principle.

Predictive processing, in its far-reaching scope, has budding and competing theories of both autism and selfhood (see Chapter 1). In this thesis, I do not aim to adjudicate between the various theories of autism from predictive processing. Nor do I develop a mathematically rigorous notion of the self under predictive processing. What I do is take some of the core concepts and functions from predictive processing and apply them to the self in autism. We will see that this gives us a promising route forward for understanding the self, autism, and the self in autism.

The predictive processing framework asserts that the primary function of the brain is to minimise the difference between expected and actual input from the world. At each information processing node, incoming information (e.g. from sensory organs, or other nodes) is compared with an expected value, and the difference between these two (i.e. the prediction error) gets passed up to the next level of the neural hierarchy. In effect, then, the brain builds a model of what the outside world is most likely to be, and updates this based on the mismatches in feedback from the actual world.

The data that the brain uses to build these models is noisy. This is due to both the complexity in the confluence of sensory signals from different sources, and internal noise from imperfect detectors and ways of passing signals between nodes of the physical network. As such, to optimise the use of the available data to form quality predictions, the cognitive system must be sensitive to the precision, or inversely, the uncertainty, of both the expectations and the prediction errors. A very small deviation from what is expected in a chaotic and noisy environment should not be taken as seriously as the same deviation in an environment with a high signal to noise ratio. These measures of uncertainty are also modelled at multiple hierarchical levels. Specifically, the brain models both the expected variability in the sensory input, and the volatility of that signal, that is, how often the signal is expected to change (in mean and/or width of the expected distribution).

This cognitive framework is not merely intended to account for perception, but also action. Not only does the brain predict the next most likely state of the world, but also the likely states of the organism. When the likely state of the organism fails to meet the current state of the organism, the neural system then corrects for this error by changing the bodily state, e.g. through movement. This is called active inference. Policy selection thus becomes a very important part of active inference, in that choosing among candidate future states will depend not only on the current inferences about and prior preferences for states of the world and the body, but also on the interaction between possible courses of action (or policies) and the likely effects these will have on future states.

These are many of the core features which will serve as a continuous thread throughout the thesis – the hierarchical nature of cognition, prediction error, active inference, policy selection and uncertainty (both variability and volatility).

For the purposes of the thesis, I therefore adopt a predictive processing explanation of Autism Spectrum Condition in which autism is characterised by differences in modelling or predicting the world under uncertainty (which impacts both perception and action). Since the processes by which an individual models themselves is also based in these same mechanisms, self-cognition should likewise be affected. Forming an inference about the nature of oneself is done under the same constraints and using the same quantities as other inferences under this framework. So instead of motivating a difference in self from interpersonal origins, we can motivate differences in self-cognition from these domain general cognitive mechanisms. In this way, the predictive processing framework may also prove to be a unifying approach to the self in autism.

Thesis roadmap

The aim of this thesis is to better understand self-cognition in Autism Spectrum Condition using tools from the predictive processing framework. In the first few chapters of the thesis, I treat self-cognition as a fairly unified construct involving plastic, reflexive, mental representations of the temporal and spatial continuity of the representing organism itself and its likely states. Chapters 1-3 address the overarching question of whether differences in self-cognition should be considered a central or defining characteristic of autism.

Chapter 1 provides a comprehensive review of experimental work on self-cognition in autism spectrum condition, covering 146 original studies across more than 22 paradigms in the 8 major domains of action, memory, self-prioritisation (attention), self-recognition, body, internal states, language, and explicit self-knowledge. It begins with an in-depth introduction to autism, the self and predictive processing. This chapter concludes with a proposal, based on data from the literature review, that the autistic self-model is “flatter” than a neurotypical self in the sense that its cognitive architecture has more functional components in lower parts of the hierarchy and less rich representation further up the hierarchy. Importantly, this flatter self is not a lesser self, but will have different features and will thrive in different environments than a self with a different or “thicker” structure.

Chapter 2 is based around a published commentary on Jaswal and Akhtar (2019), a Brain and Behavioural Sciences article that criticises the Social Motivation Theory of autism. The chapter as a whole focuses on the core features of autism, both historically and in present day, and how predictive processing can provide a unified account, unlike previous social theories. The linking text preceding the commentary briefly outlines the history of definitions of autism and some of the social deficit accounts of autism. The commentary elaborates on Chapter 1 by further discussing the relative advantages of a predictive processing framework in addressing some of the particular behaviours that Jaswal and Akhtar (2019) claim are not accounted for by the Social Motivation Theory of autism. While this chapter does not discuss particular differences related to the self, it bolsters the argument for focusing on the predictive processing account of autism as a foundation for the rest of the thesis.

Chapter 3 is the first experimental chapter. In this chapter, I aim to answer the question of whether self-cognition is as central to autism as it is to other psychiatric conditions that are defined by symptoms involving self-representation. To do this, I collected data measuring traits from five different psychiatric conditions, and measuring self-representation at two levels of the cognitive hierarchy, from over 300 participants. I compared the strength of the relationship between the self-representation scores (explicit self-concept and implicit self-prioritisation) and autistic traits, to the relationship between the same self scores and traits for self-related psychiatric conditions (borderline personality disorder and schizophrenia) and non-self-related psychiatric conditions (depression and anxiety). Results showed that explicit self-concept was less predictive of autistic traits than it was of any of the other psychiatric traits. The results of this study also emphasise the importance of narrowing focus on a smaller aspect of self-cognition, as there was no correlation between the explicit self-report and the implicit self-prioritisation task.

From this point in the thesis, the focus narrows to investigate one aspect of self-cognition which I found was anomalous in the review presented in Chapter 1 for showing no difference between autistic and non-autistic populations. This aspect is making a judgement of agency. A judgement of agency is the explicit, conscious report that one’s actions were the cause of a sensory event. This process is essential to understanding self-cognition since acting in the world is how we garner the evidence that furnishes inferences about the self. Recognising sensory effects as caused by our own actions is the first step in this process. Focusing on judgement of agency allowed us to leverage more concepts from the predictive processing framework, and the remainder of the thesis focuses on active inference as an explanation of action and action-selection, and the impact of environmental uncertainty on self-inferences.

Chapter 4 acts as a transition and presents an in depth introduction to how actions can be used as a response to uncertainty that threatens the self-model. The basic argument is that fidgeting and secondarily, autistic stimming, can be understood as the agent choosing actions that reliably reduce uncertainty. When this policy selection is successful, it reaffirms the agent’s existing model, and is thus self-evidencing. The paper introduces many of the conceptual tools used in the following experimental chapters, and demonstrates the utility of the predictive processing framework for understanding autistic behaviours in a novel way.

The majority of the experimental work in the thesis is contained in Chapters 5-7. These three chapters describe three versions of a judgment of agency task, which I will call *The Squares Task*. These chapters are preceded by linking preface text introducing the experimental paradigm and the significant aspects that speak to other parts of the thesis. In this task, participants are asked to move the mouse to determine which one of many squares on the screen they control. This task was chosen because it is similar in its basic principles to the task used in the majority of the past literature on judgement of agency in autism (Grainger, Williams, & Lind, 2014; Russell & Hill, 2001; Williams & Happé, 2009). Crucially, in contrast to these previous studies, the experiments presented here manipulate variability and volatility in the movement of the stimuli. In Chapter 5, both are manipulated within a trial, in Chapter 6 (which reports pilot data only, as the study progress was halted by COVID-19) volatility is manipulated across whole blocks, and finally in Chapter 7, the structure of variability in the sensory environment is again manipulated within a trial, but experience of it is under the control of the participant. In all three experiments, I measure the relationship between numerous dependent variables and autism traits.

Throughout the thesis, I use autism traits in a general population as a first step to understand the self in autism. Chapter 8 takes a critical stance on the primary tool we use to measure autism traits. It is focused on a published commentary discussing particular methodological issues of circularity in results from a well-popularised big data study claiming to confirm the Extreme Male Brain theory of autism (Greenberg, Warrier, Allison, & Baron-Cohen, 2018).

The thesis finishes with an overarching discussion about the findings of the thesis and future directions for interdisciplinary work on the self.

A note on thesis requirements

According to the thesis requirements for theses “including published works” in the Faculty of Arts at Monash University, theses including published works must include any publications in exactly the form the work was published including all formatting, the final published pdfs are therefore inserted into the thesis. The thesis requirements also stipulate that each chapter be accompanied by framing material or linking text, usually introducing and following the published work, which situates the chapter in the thesis as a whole. This means that the chapters that are published or submitted for publication (Chapters 1-5,8) could not be edited for their inclusion in the thesis. For consistency in style across the thesis, the unpublished chapters 6 and 7 are written in a similar style with all of the sections included in a traditional scientific manuscript prepared for submission. As a consequence, some of the introduction and methods material inevitably are somewhat repetitive. This especially holds for methods of the Squares Task Chapters 5-7, and introductory material around key concepts of predictive processing and autism throughout; the overall length and breadth of the thesis is designed to compensate for any repetition.

Chapter 1. Modelling Me Modelling You: The Autistic Self

The following publication offers a comprehensive review and analysis of self-cognition research in Autism Spectrum Conditions. This review addresses definitional questions of both selfhood and autism, provides a systematic canvas of the available literature (up until it was accepted for publication in April 2019) and offers a new way of understanding these results under the predictive processing framework. It acts as a detailed introduction to the themes and aims of the thesis. In the text following the published paper, I provide a brief review of some literature not covered by the published review.

Review Journal of Autism and Developmental Disorders
https://doi.org/10.1007/s40489-019-00173-y

REVIEW PAPER

Modelling Me, Modelling You: the Autistic Self

Kelsey Perrykkad[1] · Jakob Hohwy[1]

Received: 27 August 2018 / Accepted: 19 April 2019

Abstract

The stereotype of autism spectrum conditions (ASC or 'autism') focuses on the social and communicative elements of the diagnostic criteria. In this review, we step back from autism as a social and communicative disorder and focus on the autistic *self*. The autistic self is a key component of the condition which has nevertheless received comparatively little attention. We provide a taxonomy for experimental paradigms in the cognitive sciences that aim to address questions related to the self. We articulate reasons based on domain-general cognitive mechanisms, autobiography and historical conceptions for why the self might differ in ASC. We conclude with elucidating the implications of a predictive processing account of autism on conceptualising the autistic self and how this fits with existing literature, with a focus on context sensitivity, model complexity, learning, integration, active inference and precision. This opens up large scope for future research on unique differences in the autistic self, which could be extended as a framework for understanding the condition as a whole in a new and unified way.

Keywords Autism spectrum conditions · Autistic self · Predictive processing · Self-model · Self-cognition

The self is what unifies our experiences over time and space. As individual human beings, each of our selves is different—we see the world from a unique spatio-temporal perspective; we have both a sense of ownership over our experiences and a sense of authorship over our actions; we think that certain traits apply or do not apply to ourselves; we remember our unique personal histories; and we behave in ways that other people may not. Much of the self is captured by the subjective element of momentary experience and is thus pre-reflective. It shapes and is shaped by both perception and action. Importantly, we are one of many other selves, with whom we have regular interactions.

While most research on autism spectrum condition (ASC, 'autism') tends to focus on its social components, there is increasing evidence that whatever causes the autistic social peculiarities also affects how autistic people perceive and interact with the non-social world. If this is true, then this domain-general mechanism that causes the diverse autistic traits should also affect the autistic self. A focus on the self would indeed align with the first reports of autism from Kanner (1943) where the condition is characterised by a withdrawal into the self; it is from this that the condition gets its name—"autism" comes from the Greek *autos* for self.

Others have come to the hypothesis of altered self-cognition in autism by treating the self as a target of mentalising, which they argue is deficient in autism (Frith and Happé 1999; Williams 2010; Lombardo and Baron-Cohen 2010; Uddin 2011). We will outline some reasons to prefer a Bayesian account to the theory of mind deficit theory of autism, but likewise arrive at the hypothesis that self-cognition will be affected. This approach provides a more parsimonious and encompassing theory of autism while still encouraging better understanding of autistic self-cognition.

Understanding the experience and, thereby, the needs of other people is central to flourishing societies. It encourages reciprocal positive social interactions. However, when people do not immediately fit with our expectations for the mental lives of others, we are struck with a feeling of social discomfort. If I am

Figures and narrative interpretations contained in this paper were presented as posters at the Australasian Cognitive Neuroscience Association 2017 Conference and the Science of Self: Agency and Body Representation 2017 Conference.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s40489-019-00173-y) contains supplementary material, which is available to authorized users.

✉ Kelsey Perrykkad
kelsey.perrykkad@monash.edu

[1] Cognition and Philosophy Lab, Philosophy Department, School of Philosophical, Historical, and International Studies, Monash University, 20 Chancellors Walk, Clayton, VIC 3800, Australia

Published online: 16 May 2019

communicating with you, I can use my understanding of myself to model how you represent yourself. If my model is wrong, then my expectations for your behaviour will be inaccurate, and your behaviours will seem odd. However, if each of us correctly models the other, we can better share our experiences and knowledge. When a neurotypical person talks to an autistic person, often both have an incorrect model of the other. There is now empirical evidence to support the bidirectionality of miscommunication between autistic people and their family members, who are likely the most familiar with their behaviours in a practical sense (Heasman and Gillespie 2017). The neurotypical person predicts the reactions of another neurotypical person (in line with the majority of their past experience), and the autistic person may have difficulty modelling the other's mind in accordance with their condition. The tendency for neurotypical models of empathy and mind reading to be applied to autistic people has been described by autistic academic, Damian Milton, as 'the double empathy problem'. He writes, "the 'double empathy problem': a disjuncture in reciprocity between differently disposed social actors which becomes more marked the wider the disjuncture in dispositional perceptions of the lifeworld – perceived as a breach in the 'natural attitude' of what constitutes 'social reality' for 'non-autistic spectrum' people and yet an everyday and often traumatic experience for 'autistic people'" (Milton 2012, p. 884). Understanding the way autistic people construct a representation of themselves would go a step towards aligning this social mismatch and improving societal function. Many autistic people work tirelessly every day to develop strategies that help them understand the behaviour of others. Instead of berating them to work harder at this to fit in with 'normal' society, the neurotypical community could learn to expect variance and, particularly, work to understand common forms of mental life differing from their own. Through mutual effort, we could relieve some of the burden on the autistic community to conform to neurotypical expectations.

Towards this aim, in this article, we provide a comprehensive overview of research on the autistic self. We assert that the predictive processing framework will prove fruitful for understanding both the self and autism. We begin by describing predictive processing and provide predictive processing accounts of autism and the self. We then motivate the hypothesis that the self differs in autism. The argument culminates with a review of existing empirical evidence from the cognitive sciences, synthesised to shed light on how predictive processing may elucidate a positive account of the autistic self.

Predictive Processing

Predictive processing is an ever more popular approach in computational neuroscience, theoretical neurobiology, computational psychiatry and philosophy. For relevant reviews and introductions, see Friston (2010, 2017b)), Hohwy (2013), Clark (2015) and Friston et al. (2014). At its most basic, predictive processing asserts that the brain works to minimise prediction error. In this way, it approximates Bayesian inference over the long-term. The brain predicts sensory input, computes the discrepancy between the predicted and actual input and then changes either the prediction (by adjusting the internal model, called 'perceptual inference', see Fig. 1—solutions 1a and 2a) or the environment (through action, also called 'active inference', see Fig. 1—solutions 1b and 2b) such as to minimise this discrepancy. Through this cycle, we construct generative models of the causes of our sensory input in the outside world and try to fit these models with the evidence we have for them. This approach is simple enough to have parsimonious explanatory power in brain function, but has enough moving parts to accommodate a range of complex empirical findings including those of neuropsychiatry (Hohwy 2013; Clark 2015; Friston et al. 2014).

Crucial to predictive processing is the hierarchical structure of the resulting models. Elements at the most shallow layers of the structure are temporally and spatially small. Elements deeper in the hierarchical structure bring together these more specific cell populations, to abstract features over longer temporal and spatial scales and allow for more stable perceptions such as object recognition. Predictions feed from the deeper layers to the shallow layers, and evidence feeds back through the system.

In the predictive processing explanations of perception and behaviour, inference of underlying (or hidden) causes is performed based on probability distributions created from accumulated prior evidence. This structure also comprises second-order expectations about the shape (variance, or inversely, *precision*) of these probability distributions. These second-order expectations influence how we choose to reduce prediction error by adjusting the model's learning rate (Mathys et al. 2011; Feldman and Friston 2010). This determines how highly new information is weighted against prior experience in updating the model. Only two numbers are needed to represent each prediction, the mean and the variance. Figure 1 is a representation of the two kinds of prediction error based on these statistics: (1) how precise the signal is expected to be and (2) *what* we expect to see.

Note several key features of predictive processing, which will be relevant for autism and the self. Models are malleable and constantly changing in response to the current sensory evidence together with higher-order properties of predictions, such as the expected uncertainty in a given context. A given hierarchical model can be more or less *complex* and is thus subject to considerations of model fit and simplicity. Getting the balance between active inference (acting to change the evidence the model receives) and perceptual inference (adjusting the model to optimise perception) right depends on prior expectations for optimal model complexity. Lastly, prediction error minimisation can happen more or less

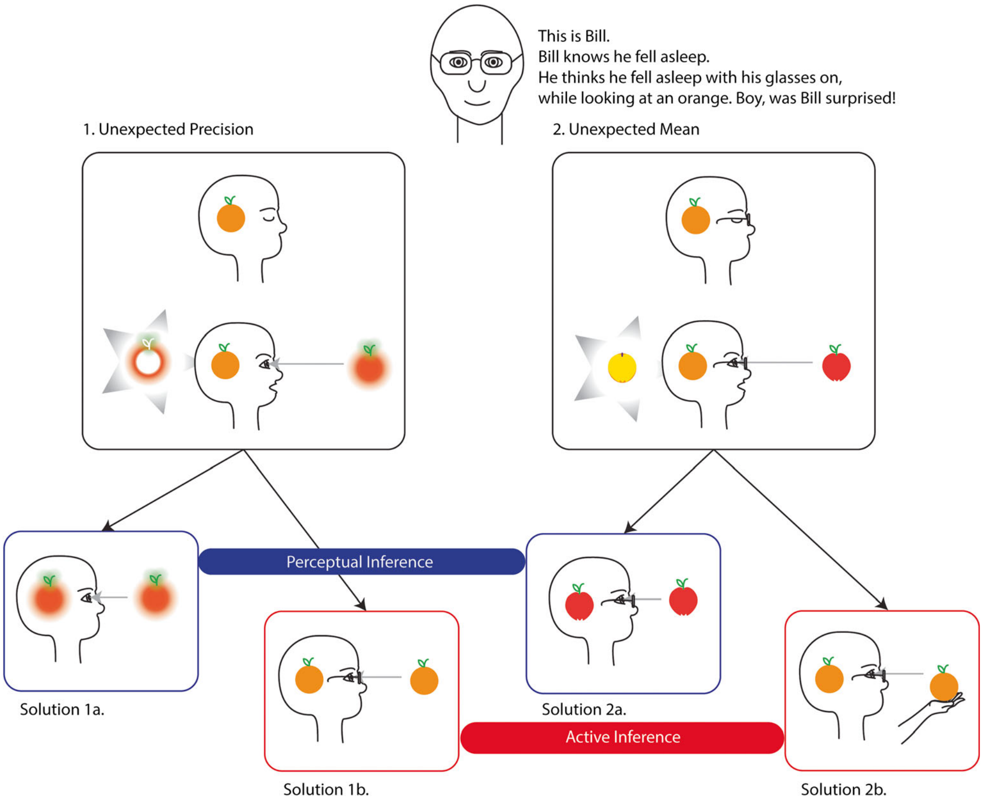


Fig. 1 Types of prediction error and ways to minimise it. Bill knowingly falls asleep with his glasses on, looking at an orange. As such, his expectations on awaking are to see a clear image of an orange in front of him. The ways in which this can go wrong and the solutions to it demonstrate different kinds of prediction error and prediction error minimisation strategies—all of which we experience in one form or another throughout our daily lives. 1. Unexpected precision—when the signal is more or less clear than the expected signal. In this example, this could be caused by Bill not actually having his glasses on when he fell asleep. 2. Unexpected mean—when the content of the signal is different than expected. In this example, when Bill actually wakes up to an apple in front of him. These prediction errors can be reduced in two ways: a. perceptual inference—This is when the internal model is adjusted to match the received input. In this case, Bill would be changing his expectations. b. Active inference—This is when the external world is acted upon so that it matches the original expectation (which is held fixed). So for solution 1b, Bill has put his glasses on, to correct the unexpected (im)precision. In solution 2b, Bill has acted to replace the apple in his visual field with an orange. In predictive processing, active inference explains *all* actions, that is, all actions are efforts to change the external world to reduce prediction error. Figure created in Adobe Illustrator

concurrently at multiple levels of the hierarchy and, therefore, at multiple temporal scales. For example, exploratory behaviour may cause short-term prediction error but one can reduce prediction error more in the long run by more broadly sampling the world.

What Is Autism Spectrum Condition?

ASC is defined by the latest *Diagnostic and Statistical Manual*, DSM-5 (American Psychiatric Association 2013), as a combination of deficits in social communication and repetitive patterns of behaviour with an onset in early childhood. Autism is a lifelong condition, though is often referred to as 'developmental' due to its early onset and particularly strong effects in childhood (including delayed developmental milestones). The current official symptoms include difficulties with social–emotional reciprocity; verbal and non-verbal communication; relationships (initiation, comprehension and maintenance); repetitive movements, use of objects or speech; rigid adherence to routines and insistence on sameness; restricted intense interests; and atypical sensory responses (American Psychiatric Association 2013). The *spectrum* terminology captures the diverse presentations and broad range of severity across individuals (Baird et al. 2003). Some autistic people cannot communicate (verbally or non-verbally), obsessively engage in violent self-harm, have frequent dramatic reactions to external sensory stimuli (such as textures, light or sound) and may seem generally overwhelmed by the world. On the other hand, many autistic people have ostensibly typical behaviour, which, in many cases, does not necessitate help-seeking and therefore often goes undiagnosed. The condition's prevalence has been estimated to be approximately 0.62% of people worldwide (Elsabbagh et al. 2012), though recent estimates in individual countries range from 1 to 3.9% (Australia (May et al. 2017); USA (Christensen et al. 2016); UK (Baron-Cohen et al. 2009)). The apparent and well-popularised increase in the prevalence of ASC is likely due to increasing awareness and changing definitions (Baird et al. 2003; Fisch 2012).

Differences in self-cognition have never been included in the official descriptions of autism. However, as this review will articulate, there is robust evidence that the self manifests differently in ASC. The keen focus on social aspects of the condition has been an obstacle for autism research, leading to a dearth of holistic accounts that include sensory and self-related symptoms as core features of ASC. While social aspects may be clinically salient, it is likely, given the multitude of other domains that are affected, that these are downstream consequences of a difference in computational structure or low-level cognitive processing phenotype.

Bayesian and Predictive Processing Accounts of Autism

In the last 5 years, Bayesian or predictive processing framework has increasingly been brought to bear on ASC (Van de Cruys et al. 2014; Pellicano and Burr 2012; Lawson et al. 2014), for review see Palmer et al. (2017). Within this perspective, there are different appeals to aspects of Bayesian inference and predictive processing, suggesting various types of perceptual and active inference differences in autism, leading to different overall generative models in autistic individuals. The high inflexible precision of prediction errors in autism (HIPPEA) account proposed by Van de Cruys et al. (2014) suggests that autism stems from overly high and inflexible estimations of the precision of sensory samples, resulting in a chronically high learning rate. Pellicano and Burr (2012) argue that priors in autism are unusually weak and nonspecific (see also Mitchell and Ropar (2004)); for discussion, see (Brock 2012).

On these accounts, social difficulties in ASC would be understood as a specific situation in which more complex inferences must be made, and therefore, a pronounced behavioural difference is identified. In order to understand social contexts, the individual must predict the behaviour of other people which is even more complicated still, because we have to predict how other people are predicting (and deciding scope of) relevant information and then predict how our own behaviour influences these external causes. They thus involve more higher order expectations about context shifts. Restricted and repetitive patterns of behaviour, interests or activities, can be interpreted under this framework as both using action to make sensory input more predictable and a solution where you put yourself into a niche in which prediction is stable and prediction error is easily minimised (akin to occupying a still, darkened room) (Constant et al. 2018). Applying a general predictive processing difference to social contexts thus quickly generates phenomena much like those seen in the autism spectrum and the variety of difficulties faced by the ASC population.

Predictive processing theories appear to have a particular advantage in scaffolding explanations which unite the strengths and weaknesses of autistic performance across modalities and domains. Like some other cognitive explanations of autism, these theories predict that autistic people are less susceptible to mistakes made by typically developing participants due to reduced interference of the global context in multiple sensory domains including audition and vision (Foxton et al. 2003; Happe and Frith 2006). They can also explain low-level discrimination talents such as the finding that autistic people have superior pitch perception (O'Connor 2012).

The unifying nature of predictive processing accounts of the brain, bringing together explanations of both perception and action, is similarly unifying for the symptoms of ASC. Where previous prominent explanations of autism have emphasised either social deficits (theory of mind theory) or sensory differences (weak central coherence account), the predictive processing perspective asserts a particular global neural structure and neatly accommodates differences in social interaction, motor systems and sensory domains. Happé (1999) notes that deficit accounts of autism, including the theory of mind and executive dysfunction theories, are poor at accounting for the particular skills typical of ASC in addition to their difficulties. The theory of mind theory of autism fails to account for the sensory hypo- and hyper-sensitivities, restricted interests, insistence on sameness and savant skills of many autistic individuals, focusing heavily on what could be considered one half of autism's diagnostic criteria (Frith and Happé 1994). Additionally, it often fails to attend to the autistic voice, and has been criticised for being based on misunderstandings of the autistic social inference processes (Milton 2012, 2014a) and sensory differences (Lawson and Dombroski 2015). Weak central coherence theory also does not account for all findings in sensory processing. For example, findings also show that the ASC population has difficulty processing speech sounds which are perceptually complex (O'Connor 2012). This theory especially fails to explain the social difficulties of ASC. Suggestions have been made in the direction of disjointed face processing and integration of broad contextual information (Happé 1999). In other instances, Happé and Frith emphasise the independence of the social deficits and other features of autism (Happe and Frith 2006; Frith and Happé 1994). In many ways, the weak central coherence theory is the closest recent relative of the predictive processing account of autism, but the predictive processing account is more easily extended across these disparate domains (Moutoussis et al. 2014).

Among the advantages of following this route in further research is the generation of new avenues for supportive interventions. If predictive mechanisms are plastic and adaptable, and these are the fundamental source of difficulties for people with ASC, perhaps we may develop methods to cultivate optimal Bayesian prediction generation where it is wanted by people in the autistic community. Precision psychiatry developed under this framework would allow for individually customised intervention (Friston et al. 2014). Targeting the functioning of the general mechanism under this theory has the potential for widespread cognitive changes. As the neurobiological basis of predictive processing is further understood, we may be able to target these behaviourally or through pharmacological intervention. For example, long-term perceptual volatility training might be predicted to improve social communication and language skills. This might be a supported hypothesis under the predictive processing schema but would be a surprising finding under many other prominent theories.

Table 1 Self-conceptions or properties of the self

Self as…		
Agent	Integrative glue	Physical
Authentic	Interpersonal	Plastic
Bodily	Intrinsic	Prediction
Bounded	Invariable	Psychological
Conductor	Invariable	Reflexive
Conscious	Material	Social
Disposition	Minimal	Spiritual
Embodied	Model	Subject
Experiential	Multiple	Transparent
Extrinsic	Narrative	Unborrowed
Historical	Object	Unbroken
Identity	Person	Unconstructed
Illusory	Personal	Unified

How to Characterise the Self

Conceptualisations of Self

The word ‘self’ is philosophically loaded and scientifically fragmented. Barresi and Martin (2011) review its historical usage and assert that “by the end of the twentieth century the unified self had died the death if not of a thousand qualifications, then of a thousand hyphenations” (p. 51). This refers to concepts like self-recognition, self-awareness, self-control, self-esteem, etc. In Table 1, we have listed a subset of the assorted features which are considered constitutive, necessary or sufficient of the self in various texts. It is interesting to note some of the contradictory properties listed here, such as *bodily* and *spiritual*; *unified* and *multiple*; and *invariable* and *plastic* (Table 1). This may give some understanding of why the subject of the self is so fraught in science and philosophy.

Here, we adopt a broad conception of the self, inspired by the predictive processing perspective. Briefly, the self is used to explain temporal and spatial continuity of an organism, which is reflexive and actively shaped by that organism’s actions and internal processes at each moment of experience. Thus, the self involves both sense of ownership over conscious experiences as they happen and memory for one’s personal history (Gallagher 2000). However, interpreted through a Bayesian lens, it is also a model which is malleable and constructed from a complex network of integrated sensory evidence.

Predictive Processing Accounts of Self

A relatively deflationary view of the self from predictive processing is self-consciousness as the process of active inference in systems with sufficient temporal depth. This view is presented recently by Friston (2018). In this particular account, the self falls out of the process of active prediction error minimisation and is thus not associated with an explicit representation within the system. See also Kiverstein (2018).

A less deflationary conception of the self under predictive processing is the self as one of the inferred and modelled hidden causes of sensory experience (Moutoussis et al. 2014; Apps and Tsakiris 2014; Letheby and Gerrans 2017). As put by Apps and Tsakiris (2014), “the notion that there is a ‘self’ is the most parsimonious and accurate explanation for sensory inputs”. By minimising prediction error (what it means to exist as an organism under predictive processing), the organism provides constant evidence for its own existence (Friston 2017a). Since, under this interpretation, the self is just a hierarchically structured bundle of inferred endogenous causes, it is subject to the same properties as our models of other hidden causes, as outlined above. It is plastic and probabilistically defined based on past experience. It is a deep part of the structure (due to its perceived consistency over time and space) meaning that the self is to some extent abstracted away from basic sensory evidence, likely involving complex multisensory integration. It emerges from and is continually shaped by an attempt to minimise surprise in both long-term characteristics and short-term behaviours through both perception and action. Because the self here is a mere inferred network of causes, this approach is compatible with being agnostic, or even sceptical, as to the existence of any real, underlying self as the referent of the internal model (e.g. Letheby and Gerrans 2017). On the other hand, it is also compatible with realism about the self, identifying the very self-model as the self, since it is the actual cause of the relevant changes in sensory input (Hohwy and Michael 2017; Limanowski and Blankenburg 2013). As a meta-model, the self under this interpretation estimates the accuracy of the larger world model and, thus, can forecast its own efficacy qua hidden cause.

From the First Person

A significant contribution to our understanding about the experiential self in autism comes from the writings of autistic people themselves. There is a wealth of books written by known autistic authors; and a comprehensive list of those published before 2006 is provided by Baggs (2013). The publications have not slowed; more recent examples include Higashida (2013) and Hammond (2010). What is clear from these autobiographies is the ability for some autistic people to recount and describe their life stories in great detail.

From two of the most widely read autistic autobiographies of Williams (2009) and Willey (2014), we can see the suggestion that the autistic self might be multiple. It should be noted that Donna Williams' case may be particularly special in this sense as she has been diagnosed with dissociative identity disorder in addition to her autism. However, the majority of the book is spent describing her embodiment as three different individuals—her namesake, Donna; the outgoing manifestation of a young girl she met at the park once as a young child, Carol; and the strong willed monster under the bed, Willie. The following is an example of how she describes this relationship between her multiple selves, "I eventually lost Donna and became trapped in a new way. Carol strove for the unacceptable: social acceptance. In doing so, Carol took the stage. Willie, my other face to the world and the embodiment of total self control, sat immobilized in the audience. Donna was still in the cupboard" (Williams 2009, p. 25). While the strength of this multiplicity is likely not typical of autistic people generally, the theme of multiple selves is present in the autobiographies of other autistic individuals who do not have comorbid dissociative identity disorder. Willey (2014) also implies a multiplicity of self, though in a less prolonged way. For example, she says, "All I had to do was fragment myself. One of me could nod, interject and produce monologues of creativity. The other me heard only my inner thoughts, felt only my irritation at the situation, understood only the need to escape. Neither of me was very good at listening to entire dialogues, but both were very good at hearing the first parts of sentences or even words, and then disregarding the other halves." (Willey 2014, p. 73). This *fragmentation* is also evident in the autethnography of Damian Milton; he writes, "Indeed, my experience of identity has had much more in common with postmodernist notions of 'fragmentation' and incoherence, although not an experience of fluidity or of an easily changeable or disposable identity. Some aspects of what it felt like to be me have seemed like immovable 'clumps'" (Milton 2014b, p. 185).

Another way in which autistic autobiography suggests a different way of constructing a sense of self is the recurring theme of what we call *echolalia of other selves*. Echolalia is a common behaviour in autism, which consists of meaningless and persistent repetition of spoken words or phrases (often in response to hearing the word or phrase from an external source, though a particular individual may develop a habit of repeating a favourite expression). It is a form of mimicry. This behaviour alone has been associated with "limitations in self-other differentiation and/or self-conception"(Hobson 2011, p. 578). However, further evidence for the impact of echolalia on the sense of self comes from the theme of mimicking another person's self, as seen anecdotally in multiple autistic autobiographies (Hammond 2010; Willey 2014; Williams 2009). Donna embodies the little girl she met at the park. Willey (2014), p. 75) says that her autistic daughter recognises her interpersonal echolalia most, "She recognizes the moment I bend my voice or my motions to match someone else's and it drives her to distraction. In no uncertain terms she will demand I stop acting like whomever, that I quit walking this way or that, that I stop pretending to be someone I am not. … I have come to the conclusion that even though I need to stop doing it, it is simply easier to echo, more comfortable and typically more successful superficially to pretend to be someone I am not." Hammond (2010), p. 25) describes the following instance, "Another kid used to chew his fingernails all the time. I remember watching him doing it and being absolutely fascinated by it. So I started biting my nails too, just to be like him in some sense." Lawson and Dombroski (2015), p. 47) describe a similar observation of a young autistic girl, "There was a character on a children's television program who wore her hair in braided pigtails that the ASC child really seemed to like. Since first seeing this character, the child became almost obsessed with wearing her hair in braided pigtails."

While each describing just one individual's experiences, the themes that emerge across multiple autobiographies should be taken seriously for further research. It would be ill-informed to disregard self-report completely in a discussion about autistic selfhood. When neurotypical researchers interpret neurotypical self-related behaviours in an experimental setting, they can ensure that their description accurately accounts for the qualia of their own experience, which may be reasonably generalised to their own population. However, in the case of autism research, where it is done by non-autistic researchers, there is an asymmetry of experience. As such, it is vital to attend to the voices of autistic individuals themselves.

Though it is undeniably important to consider the first-person accounts of autistic experience when making claims about the autistic self, it is important to note here that even in the neurotypical case, the self is an elusive entity, the cognitive components of which are not apparent to introspection; a concern also expressed by McGeer (2004) in a critique of the treatment of autobiography in Frith and Happé (1999). It is not clear that anyone has direct access to a comprehensive and accurate description of their selfhood.

Despite these considerations, the arguments that follow include some small excerpts from autobiography in an effort to include the experiences of those on the spectrum. The best way to ensure that there is no asymmetry of experience in this kind of research is to directly involve autistic individuals as part of the research team (for related discussion, see Milton 2014a).

Experimental Paradigms That Operationalise the Self

Table 2 provides a complete table of definitions for the identified paradigms, with exemplar or seminal references from studies of non-autistic populations. While it is aimed at being exhaustive, almost any experimental paradigm could be adapted to use self-relevant stimuli and could then be counted in this list. Some of these were specifically added for the relevance to the current context of ASC (e.g. alexithymia, or the inability to report, recognise and differentiate emotional states). Additionally, the categories and paradigms listed here are not mutually exclusive, and there are many existing experiments that may fall in multiple categories or compare phenomena across categories. Moreover, categories are based on conceptual groupings and are not necessarily meant to reflect distinct cognitive or neurological processes.

Each of the listed paradigms can be conceived of as testing the function or robustness of the self-model in the face of different manipulations, consistent with the overall predictive processing account of the self. *Action* studies relate to how the brain represents our causal influence on the world, and how we recognise our own self-efficacy. Agency studies have been touted as some of the best examples that test philosophical concepts of both the embodied and the minimal self (Hohwy 2007; Gallagher 2000).

Self-prioritisation studies demonstrate improved attentional allocation to self-relevant stimuli, which also prime the individual to act more quickly in situations that cue their future involvement. This can also be used as a dependent variable for studies aiming to temporarily manipulate the malleable self-representation. For example, temporary self-association paradigms involve manipulating the self-model to include the stimuli to be temporarily included in the model in order to evoke the same preferential processing for this stimulus.

Self-recognition studies involve measuring the identification of the physical self with the internal self-representation, and similarly, *body*-related paradigms demonstrate the malleability of the representation of the physical bounds of the self as predicted by a predictive processing account. *Visual perspective taking* might strike some readers as an unusual addition to this category. However, significant philosophical accounts of embodied selves rely heavily on the fact that our view of the world is from a particular bodily orientation, and in these paradigms, we can manipulate bodily positioning, if only through mentalising or self-simulation.

Studies of *internal state* representations are relevant here because they are the only stimuli that the individual has unique access to, and thus, help define the boundaries of which sensory signals one appropriates as one's own. Seth (2018), for example, argues that the understanding of one's own body as an object is driven by internal signals such as interoception and proprioception; further, he says the phenomenology of being a body is reliant upon active inference on these internal signals allowing us to predictively maintain homeostatic regularity. As such, understanding internal state representation is central to investigating self-representation.

Memory paradigms aim to demonstrate differences in representing one's personal history, and thus, through conscious representations of past experience, tap into conscious constructions of prior expectations for self, which are one of the key parts of a predictive processing story (priors), and set the baseline for expected future states of the organism. The function of episodic memory (here captured under autobiographical memory) has previously been tied to the philosophical concept of a narrative self (Gallagher 2000), or the self constituted by the stories we tell about ourselves (see Schechtman (2011) for a review of these philosophical views). Moreover, the *self-concept* paradigm, which does not involve memory, but does involve abstracting over a past experience to develop a sense of truth about temporally stable and context-independent attributions to self, may be highly related to how we account for inconsistencies (prediction errors) in individual behaviours and incorporate them into a larger story about our lives.

Understanding the accuracy of these kinds of representations and, thus, estimating how well we understand ourselves is captured in *self-knowledge* studies. From a more embodied, interactive perspective, *language* studies are used to indicate self-representation in our communication, so measuring pronoun use such as 'I' is primary in understanding how we publicly represent ourselves to others.

Empirical Evidence That the Self Is Different in Autism

There are existing theoretical reviews of self-cognition in autism of varying levels of exhaustiveness (Hobson 2011; Frith

Table 2 Paradigm taxonomy

Parent category	Paradigm	Description (exemplar citations)
Action	Sense of agency	An implicit sense of control over actions, as measured by (1) intentional binding—the measurable reduction in perceived time between voluntary actions and their sensory consequences *(David et al. 2008b; Haggard et al. 2002; Hughes et al. 2013; Moore et al. 2009), (2)* by changing the synchrony between actions and effects and often measured by perceived window of synchrony (Balslev et al. 2007; Repp and Knoblich 2007) and (3) sensory attenuation following self-performed actions (Bednark et al. 2015; Blakemore et al. 1999).
	Judgement of agency	The dependent variable is an explicit, conscious, conceptual report of whether or not participants did something (Knoblich and Prinz 2001; Repp and Knoblich 2007; Saito et al. 2015; Wegner and Wheatley 1999).
	Monitoring intentions	Meta-cognitive awareness and memory for one's own intentions in acting (Lang and Perner 2002; Russell et al. 2001; Shultz et al. 1980).
Memory	Autobiographical memory	Usually in an interview or questionnaire, participants are asked to describe specific episodic memories (personal lives or experimental content) in response to a cue (question or word) (Fivush 2011; Harris et al. 2014; Prebble et al. 2013).
	Self-reference effect (SRE/SRM)	A memory advantage for self-referents and self-traits (Gillihan and Farah 2005; Philippi et al. 2012; Rogers et al. 1977; Symons and Johnson 1997).
	Semantic	Memory for facts about oneself/one's life; remembering generalisable aspects of oneself (Klein et al. 1996).
	Memory for own actions	Refers to both the 'self-enactment effect': a memory advantage for self-performed actions as opposed to other's actions, or memory for source of certain actions (Baker-Ward et al. 1990; Berberian and Cleeremans 2010).
	Own false beliefs	Testing the ability to attribute false beliefs to the self, either in light of new information (memory—what did I think) or in appearance/reality contrast conditions (it looks like an x, but it is a y) (Atance and O'Neill 2004; Freeman and Lacohée 1995; Grèzes et al. 2004; Hogrefe et al. 1986; Wimmer and Hard 1991).
Self-prioritisation	Self-cuing	At least one cue stimulus is followed by a target stimulus in which self-related stimuli will usually provide a target processing advantage when presented as a cue (Alexopoulos et al. 2012; Higgins et al. 1988; Platek et al. 2004; Woźniak et al. 2018).
	Temporary self-association	A self-irrelevant stimuli is associated with the self through associative learning or artificial labelling within the experimental context (Sui et al. 2012; Woźniak et al. 2018).
	Orienting to own name	Also known as the cocktail party effect, this enhanced processing of one's own name, where it becomes particularly salient amongst other stimuli (Alexopoulos et al. 2012; Imafuku et al. 2014; Moray 1959; Tacikowski and Nowicka 2010; Yang et al. 2013).
	Not otherwise categorised	An advantage to presented self-stimuli which were neither artificially associated with the self nor did the task directly involve self-identification (Baess and Prinz 2017; Tacikowski and Nowicka 2010).
Self-recognition	Mirror self-recognition: sticker/red dot	A behavioural measure of the ability for the visual self to become the object of its own attention qua self (Amsterdam 1972; Bertenthal and Fischer 1978; Chang et al. 2017; Gallup Jr. 1968; Gallup et al. 2011).
	Own face recognition	Face recognition here is distinguished from mirror recognition by the use of a static own-face image instead of responsive, dynamic stimuli (Cahrel et al. 2002; Keenan et al. 1999, 2001; Sugiura et al. 2000; Uddin et al. 2005).
	Not otherwise categorised	Other non-standard tests of recognition of the self in stimuli, including the use of video with delayed playback, recognition of own name, object ownership, body recognition or own biological motion (Jokisch et al. 2006; Miyakoshi et al. 2010; Perrin et al. 2005).
Body	Theory of mind: visual perspective taking	Examples of experimental first person perspective (1PP) by forcing a temporary reallocation of spatio-temporal orientation to give a response from an alternative spatial location including the director task and avatar perspective taking (Dumontheil et al. 2010; Epley et al. 2004; Mattan et al. 2015, 2017; Samson et al. 2005; Surtees et al. 2013; Surtees and Apperly 2012).
	Body representation	The neural or psychological representation of one's own body and its boundaries (De Preester and Tsakiris 2009; Holmes and Spence 2006; Iriki et al. 1996; Longo et al. 2010; Maravita and Iriki 2004; Sedda 2011).
	Body ownership: tactile illusions	In the classic rubber hand illusion, a participant's hand is visually hidden, and a rubber hand placed on the table in front of them in a similar position to their own hand. When both hands are simultaneously stroked with a paintbrush, many people get the illusion that the rubber hand is their own (Botvinick and Cohen 1998; Rohde et al. 2011; Suzuki et al. 2013; Tsakiris and Haggard

Table 2 (continued)

Parent category	Paradigm	Description (exemplar citations)
		2005). Variations not including a rubber hand are subsumed under this broader heading.
Internal states	Introspection	Ability to self-reflect on mental processes and successfully communicate their phenomenology. Instances of introspection can be analysed for quality and content (Carruthers 2009).
	Interoceptive awareness	Awareness of homeostatic signals internal to one's body (Craig 2003; Crucianelli et al. 2016; Ondobaka et al. 2015).
	Alexithymia	A particular form of introspective deficit manifesting in an inability to identify and describe one's own emotions (Bagby et al. 1994; Lane et al. 1998; Lumley et al. 1996; Taylor 1984; Taylor et al. 1991).
Language	Pronoun use	The correct use of pronouns (I, you, he/she) as measured by either parental report or sampling language and analysing for frequency and accuracy of personal pronoun use (Charney 2008; Goodenough 1938; Lewis and Ramsay 2004).
Self-knowledge	Self-concept	Participants are asked to describe themselves, or participants judging whether or not certain words describe them without a memory component (Higgins et al. 1988; Hart and Damon 1986; Damon and Hart 1986; Saxe et al. 2006).
	Meta-knowledge	Studies investigating knowledge or ignorance about one's own knowledge or ignorance about oneself (Burton and Mitchell 2003; Raviv et al. 1990).

and Happé 1999; Williams 2010; Lombardo and Baron-Cohen 2010; Uddin 2011; Lyons and Fitzgerald 2013; Molnar-Szakacs and Uddin 2016; Huang et al. 2017). Here, we diverge from these previous approaches by hypothesising that self-differences in autism are rooted in differences in predictive processing. The advantage of this is that it provides access to the interpretive and explanatory tools provided by the predictive processing framework and the generation of new hypotheses based on these novel interpretations of data from self-cognition research in ASC. We aim to be comprehensive in a cross-disciplinary analysis of self-cognition in autism.

Evidence from Self-Cognition Paradigms

Table 3 lists all the individual studies reporting original empirical investigation using one of the paradigms listed in Table 2 in an autistic population (or use an AQ measurement to divide their results as a primary aim) (Table 3). All but four of these studies use a clinically diagnosed autistic population. As can be seen in Fig. 2b, autism conceived of as a developmental disorder influences the distribution of studied ages in these paradigms. Roughly half of the paradigms are tested in children; there is a bimodal distribution with clusters around the preteen years and late 20s/early 30s. An important avenue for future research will be to study how the self-model develops and changes over the lifespan, including in autistic populations. The details of which studies have child populations are noted in Table 3.

Action

> Everything I did, from holding two fingers together to scrunching up my toes, had a meaning, usually to do with reassuring myself that I was in control and no-one could reach me, wherever the hell I was.
> (Williams 2009)

Based on existing data, performance on judgement of agency tasks is comparable between ASC groups and a typically developing (TD) population (David et al. 2008a; Williams and Happé 2009a; Russell and Hill 2001; Grainger et al. 2014). However, there is some evidence that sense of agency, or at least as measured by action-effect temporal binding (Sperduti et al. 2014) and sensory attenuation following own action (van Laarhoven et al. 2019), is reduced in autism. Some researchers account for these conflicting findings by appealing to a distinction between mechanisms involved in psychological and physical selves (Williams 2010; Uddin 2011; Molnar-Szakacs and Uddin 2016). However, along with Zahavi (2010), we question the conceptual clarity and robustness of this kind of distinction, though commonly used in theoretical discussions of self-cognition. This confusion is particularly apparent in studies of agency, which are usually considered under physical self (Williams 2010; Uddin 2011; Molnar-Szakacs and Uddin 2016), but could easily be understood as having a large component in psychological aspects of the self. Additionally, as we will see later, interoceptive awareness is largely reduced in autism, which would presumably fall under the physical self.

Table 3 Self-cognition paradigms: findings in autism. For descriptions of paradigms, see Table 2

Paradigm	Original empirical studies using an autistic population
Sense of agency	Sperduti et al. (2014); van Laarhoven et al. (2019)
Judgement of agency	David et al. (2008a), Grainger et al. (2014), Russell and Hill (2001)[a], Williams and Happé (2009a)[a], Zalla et al. (2015)
Monitoring intentions	Phillips et al. (1998)[a], Williams and Happé (2010b)[a]
Actions: not otherwise categorised	Grynszpan et al. (2012)
Autobiographical memory	Bowler et al. (2000), Bruck et al. (2007)[a], Crane et al. (2009), Cygan et al. (2018), Goddard et al. (2007, 2017)[a], Klein et al. (1999), Kristen et al. (2014), Williams and Happé (2010a)[a]
Self-reference effect (SRE/SRM)	Gillespie-Smith et al. (2017)[a], Henderson et al. (2009)[a], Lombardo et al. (2007), Toichi et al. (2002), Williams et al. (2017), Yoshimura and Toichi (2014)
Semantic	Crane and Goddard (2008), Goddard et al. (2017)[a], Klein et al. (1999), Toichi et al. (2002)
Memory for own actions	Dunphy-Lelii and Wellman (2012)[a], Farrant et al. (1998)[a], Grainger et al. (2014), Hala et al. (2005)[a], Hare et al. (2007), Hill and Russell (2002)[a], Lind and Bowler (2009a)[a], Maras et al. (2013), Millward et al. (2000)[a], Russell and Jarrold (1999)[a], Summers and Craik (1994)[a], Williams and Happé (2009a)[a], Wojcik et al. (2011)[a], Yamamoto and Masumoto (2018), Zalla et al. (2010)
Own false beliefs	Baron-Cohen (1991, 1992)[a], Bradford et al. (2018), Fisher et al. (2005)[a], Leslie and Thaiss (1992)[a], Perner et al. (1989)[a], Russell and Hill (2001)[a], Williams and Happé (2009b)[a]
Self-cuing	Zhao et al. (2018)
Temporary self-association	Skorich et al. (2017)[b], Williams et al. (2017), Zhao et al. (2018)
Orienting to own name	Cygan et al. (2014), Leekam and Ramsden (2006)[a], Mars et al. (1998)[a], Nadig et al. (2007)[a], Nijhof et al. (2017), Osterling and Dawson (1994)[a], Zwaigenbaum et al. (2005)[a]
Self-prioritisation: not otherwise categorised	Morita et al. (2012), Zamagni et al. (2011)[a]
Mirror self-recognition: sticker/red dot	Dawson and McKissick (1984)[a], Ferrari and Matthews (1983)[a], Hobson and Meyer (2005)[a], Reddy et al. (2010) [a], Spiker and Ricks (1984)[a]
Own face recognition	Chakraborty and Chakrabarti (2015)[b], Uddin et al. (2008)[a]
Self-recognition: not otherwise categorised	Burling et al. (2019)[b], Dissanayake et al. (2010)[a], Dunphy-Lelii and Wellman (2012)[a], Lind and Bowler (2009b) [a], Neuman and Hill (1978)[a], Root et al. (2015)[a]
Theory of mind: visual perspective taking	Baron-Cohen (1989b)[a], Begeer et al. (2010)[a], Dawson and Fernald (1987)[a], Hamilton et al. (2009)[a], Reed (2002), Reed and Peterson (1990)[a], Russo et al. (2018)[a], Santiesteban et al. (2015), Warreyn et al. (2005)[a], Zwickel et al. (2011)
Body representation	Asada et al. (2017)[a], Bertilsson et al. (2018), Russo et al. (2018)[a], Vasudeva and Hollander (2017)
Body ownership: tactile illusions	Cascio et al. (2012)[a], Blakemore et al. (2006), Greenfield et al. (2017)[a], Greenfield et al. (2015)[a], Guerra et al. (2017), Mul et al. (2019), Palmer et al. (2013)[b], Palmer et al. (2015)
Interoceptive awareness	Barttfeld et al. (2012), Elwin et al. (2012), Fiene and Brownlow (2015), Garfinkel et al. (2016), Mul et al. (2018), Schauder et al. (2015)[a], Shah et al. (2016), Thaler et al. (2017)
Introspection	Baron-Cohen (1989a)[a], Hurlburt et al. (1994)
Alexithymia	Allen et al. (2013), Berthoz and Hill (2005), Bird et al. (2010, 2011), Cook et al. (2013), Griffin et al. (2015)[a], Heaton et al. (2012), Hill et al. (2004), Hobson et al. (2018)[a], Mul et al. (2018),

Table 3 (continued)

Paradigm	Original empirical studies using an autistic population
	Shah et al. (2016), Silani et al. (2008), Szatmari et al. (2008)[a]
Pronoun use	Baltaxe (1977)[a], Bartak and Rutter (1974)[a], Dascalu (2018)[a], Dunphy-Lelii and Wellman (2012)[a], Jordan (1989)[a], Lee et al. (1994)[a], Lombardo et al. (2007), Mizuno et al. (2011), Prévost et al. (2018)[a], Silberg (1978)[a]
Self-concept	Farley et al. (2010)[a], Jackson et al. (2012), Lai et al. (2018), Lee and Hobson (1998), Lombardo et al. (2010), Scheeren et al. (2010)[a]
Meta-knowledge	Capps et al. (1995)[a], Dritschel et al. (2010)[a], Elmose and Happé (2014)[a], Furlano & Kelley (2019)[a], Mitchell and O'Keefe (2008), Sasson et al. (2018), Vickerstaff et al. (2007)[a]

[a] Mean participant age was under 18 years (or middle value in given age range where mean is unavailable)

[b] Study *only* completed on autism quotient (AQ) measurements, no formal diagnosis of participants

A review by Zalla and Sperduti (2015) suggests that the potentially conflicting findings in these studies may be understood by differentiating prospective and retrospective aspects of agency, and assert that retrospective agency may be impaired in ASC. It has also been found that despite similar performance, autistic people rely more heavily on external cues to agency, such as visual input, rather than internal cues, such as proprioception (Zalla et al. 2015). However, while a different weighting between sensory inputs is a common finding in autistic self-cognition research, this conclusion is often reversed, such as in bodily ownership and force adaptation domains (Gowen and Hamilton 2013; Greenfield et al. 2015) in which the findings seem to more consistently show over-reliance on proprioception rather than visual cues (see also "Body" section below). On the whole, autistic people have poorer retrospective recognition of their original intentions when the outcome is accidental or reflex-driven (Phillips 1993; Phillips et al. 1998; Williams and Happé 2010b), which may lend support to this idea. Though contrary to this, Russell and Hill (2001) found intact abilities to monitor own intentions in the face of an incongruent outcome.

Memory

> I mark my life by moments in time, captured like morning glories at dawn, small and simple, yet fine and real. Moments define me, they make me complete. I envision the times that come together to form who I am.
> (Willey 2014)

Memory for agent, as measured by accuracy of source identification following a card game, is generally found to be

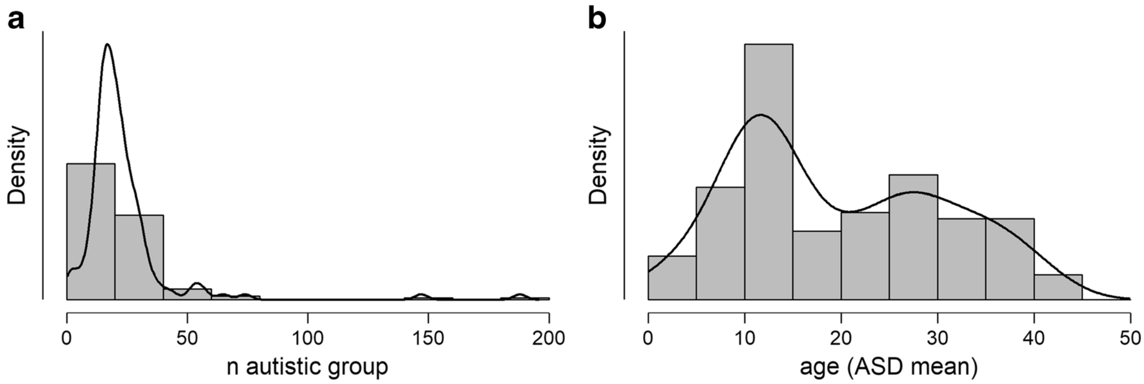


Fig. 2 Population sampled for studies of autism and self. **a** Density plot showing the number of participants in the autistic group for $n = 135$ empirical studies of the self in autism (six studies had two samples which were separately included and six studies or substudies did not use a diagnosed autistic population and so were excluded from this plot). **b** Density plot showing mean reported age of the autistic population (or general sample for $n = 6$ studies without a diagnosed autistic population) included in $n = 141$ studies of the self in autism. Where mean reported age is not reported or inferable from data provided, the midrange is included for this plot. Figure is a JASP output

reduced in ASC (Dunphy-Lelii and Wellman 2012; Lind and Bowler 2009a; Russell and Jarrold 1999; Yamamoto and Masumoto 2018), though there is one exception (Farrant et al. 1998). In other variations of this kind of task, Grainger et al. (2014), Hill and Russell (2002) and Zalla et al. (2010) found action source memory is intact in an autistic population.

Additionally, in typical populations, one usual finding is an advantage for the memory of one's own actions over the actions of others, also called the self-enactment effect. Research in the self-enactment effect in autism is mixed, but is one of the most studied of any self in autism paradigm summarised here. A few studies (Lind and Bowler 2009a, b; Grainger et al. 2014; Summers and Craik 1994; Yamamoto and Masumoto 2018) show the expected advantage of memory for self-performed over observed actions in an autistic population. Similarly, Hill and Russell (2002) and Hala et al. (2005) found no difference between groups in distinguishing previously performed actions from lures. Other research shows that there is no self-enactment effect in autism (Hare et al. 2007; Millward et al. 2000; Russell and Jarrold 1999; Wojcik et al. 2011). Williams and Happé (2009a) found no difference between their tested populations on memory for own actions, but additionally found no self-enactment effect for either group. One study even show a reversed effect with an advantage for memory of others' actions over their own in ASC (Millward et al. 2000). In a tabular summary, Grainger et al. (2014) show that overall there is little evidence for differences in self-enactment effect between autistic cohorts and matched typical samples, though this particular representation fails to include whether the typical group's enactment effect was significant in the first instance (which can be difficult to infer from the summary statistics reported in a within-group analysis without the original authors having tested for this specifically).

Experiments that test memory for one's own previous beliefs show that like typically developing children, children with autism more easily identify their own false prior beliefs than others' past false beliefs (Baron-Cohen 1991, 1992; Fisher et al. 2005; Perner et al. 1989; Russell and Hill 2001; Leslie and Thaiss 1992). Fisher et al. (2005) show poorer performance for autistic participants on the own false belief task as compared to a control group with moderate learning difficulties, but similar performance across self and other conditions. This may be clarified by the design of Williams and Happé (2009b) in which children did not say what they thought aloud until after the reality was revealed, and own false belief performance was worse than false belief attribution to others. This suggests that memory for what one has said is not impaired in autism, but mindreading one's past self is more difficult for this population.

Episodic memory also seems to be affected in ASC. Autobiographical episodic memories are generally found to be less specific and/or fewer in number in ASC populations (Bruck et al. 2007; Crane and Goddard 2008; Goddard et al. 2007; Klein et al. 1999). However, Williams and Happé (2010a) found no deficit in reporting own emotional past events. Semantic memory for self, such as yes/no responses to questions about life events or facts about one's life (e.g. name two of your elementary school teachers), seems to be more intact in autism (Crane and Goddard 2008; Kristen et al. 2014), though results are mixed (Bruck et al. 2007) and have been suggested to be age related (Lind 2010). Goddard et al. (2017) found that autistic participants felt a weak connection to their memories, though they relied on episodic memories more than trait memory to reveal aspects of their personalities. In contrast, Bowler et al. (2000) found that memories for words in autistic participants were reliant on semantic rather than episodic processes.

In the typical population, there is a memory advantage for self-referents and self-traits. This is called the self-reference effect. Some evidence shows that this memory advantage is absent or decreased in ASC compared to TD (Henderson et al. 2009; Lombardo et al. 2007; Toichi et al. 2002; Yoshimura and Toichi 2014). However, in a recent study of perceptual self-reference effect in which participants associated themselves temporarily with a shape, ASD participants showed no impairment in memory for self-referents compared to controls (Williams et al. 2017).

Self-Prioritisation

There is relatively little research amongst the paradigms identified which involve increased attention or processing for self-related stimuli. The lack of orienting to one's own name is an exception, and it is a common warning sign for children who may be later identified as having ASC. Empirical evidence in children and infants confirms this red flag (Leekam and Ramsden 2006; Mars et al. 1998; Nadig et al. 2007; Osterling and Dawson 1994; Zwaigenbaum et al. 2005). EEG data from adults with ASC demonstrated equal increases in P300 amplitude to both own name and others' name which were overall larger than TD participants. This activation was related to diminished activation of the rTPJ and attenuated lateralisation of the neural response to own name and demonstrated a lack of self-specific prioritisation (Cygan et al. 2014; Nijhof et al. 2017). Recent work by Zhao et al. (2018) found no significant difference in self-prioritisation in a temporary association task, suggesting usual self-prioritisation effects in both groups. However, there were differences between groups in how signals were integrated from different modalities (gaze direction and auditory signals) when self was used as a cue, and this was significantly correlated with autistic symptom severity.

Self-Recognition

> Carol came in through the mirror. Carol looked just like me, but the look in her eyes betrayed her identity. It was Carol all right. I began to talk to her, and she copied me. I was angry. I didn't expect her to do that. My expression asked her why, and hers asked me. I figured that the answer was a secret.
> (Williams 2009)

One study by Uddin et al. (2008) shows a typical pattern of neural activation for own face stimuli, but reduced premotor/prefrontal activity for other people's faces. In the same vein, there are many studies investigating mirror self-recognition and delayed self-recognition (on video) in autism. The classic paradigm in this area involves surreptitiously placing a red dot on the participant's face and comparing dot-directed behaviour with baseline face-touching behaviour. It is thought that this demonstrates the ability to recognise that the person in the mirror is related to their own body and that they can investigate changes in the mirror by investigating parts of their own body. Early studies of developmental self-recognition find a proportion (50–86% of participants, 4–8 years old) of their autistic subjects successfully show self-recognition (Dawson and McKissick 1984; Ferrari and Matthews 1983; Spiker and Ricks 1984). There seems to be similar performance between TD and ASC children on delayed self-recognition tasks (Lind and Bowler 2009a, b; Dissanayake et al. 2010; Dunphy-Lelii and Wellman 2012; Neuman and Hill 1978; Root et al. 2015). However, this ability may be related to a global developmental delay in autism, as most typically developing children can perform mirror self-recognition before 2 years old (Ferrari and Matthews 1983; Amsterdam 1972). A case study by Root et al. (2015) suggests that despite intact recognition at short lags in video playback, autistic participants may not be as sensitive to prolonged and unexpected delays in feedback.

Body

> When someone facing me moved their left arm, I moved my right arm. When they moved their right arm, I moved my left arm and so on and so forth. I knew all along that I was making a mistake, but no matter what I did and no matter how many times I told myself things like 'her right arm equals my left arm,' I could not transfer the knowledge to the movement. After a few weeks of bilateral torture, I figured out I might find some success if I practiced our dance steps from the back row; a vantage point that allowed me to carbon copy the people who were facing the same direction I was.
> (Willey 2014)

In an extension of the red-dot mirror self-recognition paradigm, autistic children have been shown to use their bodies less in communicating the location of a sticker to other children (Hobson and Meyer 2005).

Due to the intense focus of researchers on theory of mind and social processes in autism over the last two decades, there is abundant research on visual perspective taking. We include it here for its unique relationship to bodily theories of the self and the possibility of shedding light on aspects of the self relating to first-person perspective. Despite the wealth of related studies, there is little consensus for ASC in this area. Some studies show good performance and no difference between ASC and TD populations (Baron-Cohen 1989b; Begeer et al. 2010; Reed and Peterson 1990; Zwickel et al. 2011; Santiesteban et al. 2015). Others show difficulties which are revealed by slower responses or lower accuracy in ASC participants (Dawson and Fernald 1987; Hamilton et al. 2009; Reed 2002; Schwarzkopf et al. 2014; Warreyn et al. 2005; Russo et al. 2018). For a detailed review on visual perspective taking in autism, see Pearson et al. (2013), who claim that different strategy use by participants across different paradigms may lead to the overall mixed results observed.

The rubber hand illusion allows experimenters to empirically manipulate the sense of bodily ownership. Findings from Cascio et al. (2012) and Greenfield et al. (2017) suggest that an ASC population is less susceptible to the illusion, and specifically found that autistic participants show delayed onset of the illusion and reduced embodiment with incongruent stimulus at shorter delays. Other findings in ASC indicate that even when autistic participants report subjective experience of the illusion, they demonstrate weaker influence of this perception on subsequent movements (Palmer et al. 2013, 2015). This is consistent with the idea that autistic subjects rely more on proprioceptive information to bodily ownership (Greenfield et al. 2015). These findings were recently extended by Mul et al. (2019) to the full body illusion, who found significantly reduced experience of this illusion in the ASC group.

Internal States

> For me I've always had trouble understanding these emotions and how to express them in an appropriate way as an adult.
> (Hammond 2010)

> Perhaps I did not lack the feeling of hunger, or needing to go to the toilet, needing to sleep. Perhaps my preoccupation with remaining a step away from fully conscious made it necessary for my mind to deny the awareness of these needs; certainly I would ignore the signs, feeling faint, anxious or grumpy, yet always too busy to stop for such things.
> (Williams 2009)

There has been a plethora of research in the area of interoception and alexithymia in autism, and there is some evidence that the two are related (Shah et al. 2016). On the whole, there seems to be a reduced sensitivity to both internal bodily signals to hunger, thirst, tiredness, toileting need, etc. (DuBois et al. 2016; Elwin et al. 2012; Fiene and Brownlow 2015; Mul et al. 2018) and an increased prevalence of alexithymia in autism (Kinnaird et al. 2019; Szatmari et al. 2008; Hill et al. 2004; Griffin et al. 2015; Berthoz and Hill 2005). The presence of alexithymia has been correlated both with autism severity (Griffin et al. 2015; Hill et al. 2004; Bird and Cook 2013) and other symptoms of autism, especially in the realm of social cognition, including face perception (Cook et al. 2013; Bird et al. 2011) and theory of mind and empathy-related processes (Bird et al. 2010; Silani et al. 2008). Recently, it has been shown to affect the likelihood of receiving an autism diagnosis (Hobson et al. 2018). See also Hatfield et al. (2017) for a review of interoception in autism interpreted through a weak central coherence account.

Language

> Strangely, it took me four more years [at age six] to realize that normal children refer to themselves as 'I'. (Williams 2009)

While there are many differences in language abilities related to ASC, here, we focus on the only capacity specifically related to self-cognition—pronoun use. The correct use of pronouns (I, you, he/she) in verbal communication is demonstrative of discriminability and recognition of various agentive concepts, indicative of the recognition of different selves. There is a consistent reduction in the frequency of use of the 'I' or 'me' pronouns across studies including participants of all ages (Baltaxe 1977; Dunphy-Lelii and Wellman 2012; Jordan 1989; Lee et al. 1994; Lombardo et al. 2007; Dascalu 2018). Additionally, differences in pronoun recognition were associated with neural differences in ASC (Mizuno et al. 2011). The source of this difficulty is likely mixed and may rely on social, cognitive or grammatical origins (Charney 1980; Fay 1979).

Self-Knowledge

> I didn't start to warm to my dad again until I was about 22 and it remains a huge mystery to me as to why it ever happened. Asperger's is like that. For whatever reason, it creates mysterious behaviour, confusing even the Aspie! (Hammond 2010)

There is relatively little research in the self-knowledge category, but this may be due to the difficulty in capturing the entirety of a person's knowledge about themselves or successfully measuring the differences in self-concept (which here is broader than self-esteem and self-efficacy, including all the beliefs about oneself) between populations. Studies are conflicting about which aspects of the self-concept might be different in autism, with Farley et al. (2010) finding less reference to self as agent in ASC, but Jackson et al. (2012) finding comparable descriptions of self as physical and active, but rather reduced in the areas of self as object, social and psychological. Contrary again, Lee and Hobson (1998) only found group differences in descriptions of self in the social domain. Evidence shows that meta-cognition about social skills is intact in autism (Capps et al. 1995; Vickerstaff et al. 2007), as well as for memory performance (Elmose and Happé 2014). However, a recent study by Sasson et al. (2018) shows that adults with ASD have poor meta-cognition about the personality traits that they portray to others, despite showing no difference in self-attribution of personality traits. Autistic children have been shown to overestimate their own academic competency compared to neurotypical children, but are significantly more accurate when provided with feedback (Furlano and Kelley 2019). There also might be neural differences in self-trait attribution (Lombardo et al. 2010). In two studies, ASC participants claimed that others have an equal or better understanding of themselves than they do (Dritschel et al. 2010), whereas typically developing people tend to claim privileged self-knowledge outstripping others' knowledge of oneself.

Introspection in ASC has been measured using self-reported mental states at randomly sampled intervals and suggests autistic mental life largely consists of imagery (Hurlburt et al. 1994). This is also consistent with some autobiographic reports (Grandin 1996). Relatedly, Baron-Cohen (1989a) found that autistic participants fail to discriminate between appearance and reality, which is interpreted as being unaware of their own mental states. Ultimately, it is difficult to robustly capture self-knowledge and introspection empirically, so there is limited evidence upon which to base claims in this area.

Summary and Identification of Gaps in Existing Research

This body of research suggests the following about the autistic self. There is much evidence that memory for self is reduced in autism, particularly episodic memory and memory for own false beliefs. Attentional orientation to own name is attenuated in autism from both behavioural and EEG studies. There seems to be a delay in mirror self-recognition, though this is likely eventually acquired for the majority of children with autism. There is evidence for a difference in cue integration for body ownership evidenced by implicit reactions to the rubber hand illusion and full body illusion. There is evidence that ASC is associated with difficulties in interoception and identifying and reporting own emotions. Additionally, pronoun use is either reduced in frequency or more interchangeable than in typical populations. On the other hand, a few self-cognition paradigms show reliably intact abilities in autism. These areas include judgement of agency, delayed self-recognition from video recordings and possibly semantic memory for self-facts.

For some areas of self-cognition research, there are very few studies on a particular aspect of autistic self-cognition. These include sense of agency (as opposed to judgement of agency), self-prioritisation studies such as self-cuing and temporary self-association, own face recognition (from a still image), body representation and introspection. For a graph of study distribution across paradigms see the Supplementary Materials.

A Predictive Processing Account of the Self in Autism

> The human saga is just not reliable enough for me to predict. Social situations are not the only things I find unreliable, and hence, untrustworthy and uncomfortable. My sense of visual perception often plays tricks on me, making it difficult for me to do ordinary tasks... Generally speaking, I know I should not rely on my own visual perception, but practically speaking, it is sometimes impossible to rely on anything else.
> (Willey 2014)

For the purposes of the following theoretical explanations and hypotheses, we adopt the self as meta-model conception of self. This asserts that a subset of the predictive process, specifically, the reflexive part which regulates and evaluates the performance of the system itself, *is* the self (Hohwy and Michael 2017). This will also imply the self as hidden cause conception of the self as one of the many hidden causes of sensory input as it will infer this representation as the source of actions and their sensory consequences. Similarly, we broadly subscribe to the predictive processing accounts of autism (Van de Cruys et al. 2014; Pellicano and Burr 2012; Lawson et al. 2014), using evidence presented above from autistic self-cognition to feedback on current theories of predictive mechanisms as they manifest in ASC without straightforwardly adopting any existing theory for the source of predictive differences in autism. We suggest themes under which a predictive model of self may be operationalised and reveal itself as having a more ‘flat’ hierarchical structure in autism than in the neurotypical case.

Overall, we have seen that the autistic self seems to involve fewer robust and stable features than the neurotypical self. The notion that the self is ‘thin’ in autism is not new (Hobson 2011). Relatedly, recent work by Constant et al. (2018) has used a similar model of the self in autism derived from HIPPEA to explain autistic environmental niche construction as a way of minimising prediction error for the self-model more effectively. Additionally, many aspects of the self as studied by the empirical paradigms reviewed above seem to indicate distinctive differences in the function and structure of the autistic self. However, within the Bayesian brain framework, we can reconceptualise the metaphorical notion of ‘thinness’ of self in a broader, more theoretically substantive framework. A thin or flat self is not a lesser self, but rather a self with different properties and different optimal environments.

Figure 3 illustrates the autistic self-model we propose based on the reviewed evidence and its relationship to the predictive processing framework as outlined below. The assertion is that the autistic brain is using a differently structured generative model to deal with the problem of perception (which would also change the response space (cf. policy selection) and, therefore, the behaviours of the individual). We suggest that the structure of the internal generative model (or recurrent neural network) in autistic people is flatter than in neurotypical individuals (which may be compensated for by richness of nodes at lower levels, providing the foundations for the typical central coherence findings of improved attention to and identification of details). In this model of autistic information processing, there are fewer, or less precise and informative, nodes in the upper, more long-term end of the temporal–spatial hierarchy, and in contrast, more nodes towards the variable, more sensory end of the hierarchy. If this kind of structure explains information processing in general in autism, and the self-model accurately represents this, then the self-model will also have this flatter structure. We suspect, based on the range of self-related stimuli, that nodes indexing the self-model are likely distributed across the hierarchy, as represented by the darker grey nodes in Fig. 3. Consequences of this self-model and explanations of how the existing evidence articulated above might lead us to this structure are further elaborated in the following sections.

Context Sensitivity

> *Remember I can't apply the rules from one party to another exactly because each party, and the people there, are never the same. They are always different in really subtle ways. (This may also apply to friends and other people.)*
> Hammond (2010) on how to cope, section titled “going to parties”

> *The significance of what people said to me, when it sank in as more than just words, was always taken to apply only to that particular moment or situation. Thus, when I once received a serious lecture about writing graffiti on Parliament House during an excursion, I agreed that I'd never do this again, and then ten minutes later, was caught outside writing different graffiti on the school wall. To me, I was not ignoring what they said, nor was I being funny: I had not done exactly the same thing as I had done before. My behaviour puzzled them; but theirs puzzled me, too. It was not so much that I had no regard for their rules as*

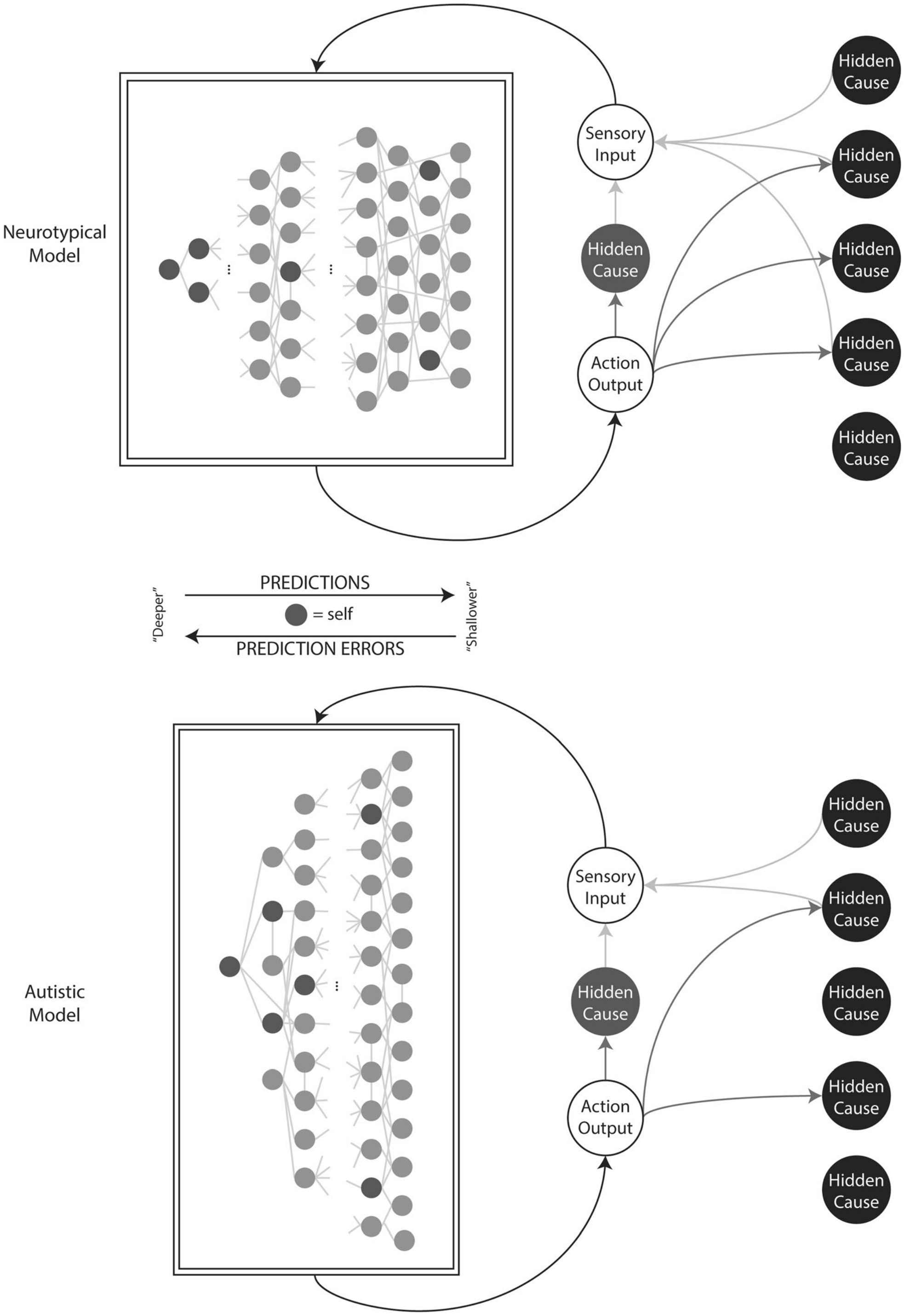


Fig. 3 Autistic self model. The image on the top represents predictive processing in the neurotypical case. The model inside the box depicts the model of the world, which represents the hidden causes on the right-hand side of the image. The self, if explicitly represented, may be represented as one of the hidden causes of the world (self as hidden cause theory, as pictured by the dark grey hidden cause between the action output and the sensory input, which may be a *mis*-representation), or by a subset of the internal nodes which monitor the performance of the system as a whole (self as meta model theory, pictured by the dark grey nodes inside the model of the world). The relationship between the modelled hidden causes and the world is mediated by action output and sensory input and, thus, may not capture all the actual hidden causes. Our proposed autistic model is pictured on the bottom half of the image. The internal model has the same number of nodes, but is differently distributed, such that more nodes sit at the 'shallow' parts of the hierarchy, which represent fine-grained hidden causes (i.e. at small spatio-temporal scale). The self-representation would change as a result of this general model shape, as it influences which hidden causes the individual perceives, and how they expect to be able to interact with them causally. Figure created in Adobe Illustrator

> *that I couldn't keep up with the many rules for each specific situation. I could put things into categories but this type of generalizing was very hard to grasp.* (Williams 2009)

One of the ways in which this predictive processing model of autistic information processing might influence the self-model is through its effects on context sensitivity. Being context sensitive means recognising when the statistical regularities and underlying hidden causes may have changed. It is an important part of generating socially appropriate behaviour. Determining the scope of rules and generalising these allows for efficient processing of complex environments. However, as many researchers have noted outside of the Bayesian framework (Plaisted 2001), the ASC population seems to struggle with context appropriate behaviour. In certain cases, where strong priors prove unhelpful, avoiding the influence of obvious context cues can lead to better performance. Instead of context sensitivity, the autistic cognitive style may lend itself to context specificity.

Consider the structure proposed in Fig. 3. While it sub-optimally approximates Bayes in the long run, it may more accurately reduce prediction error in the medium to short term. This reflects a fundamental aspect of the overall imperative of an organism to minimise prediction error, namely the trade-off between the current rate of prediction error minimisation on the one hand, and on the other, the overall time scale on which prediction error is expected to be minimised. We propose, in effect, that autistic people are more preoccupied with the current rate of prediction error minimisation, and sub-optimally represent transient and volatile increases in prediction error in the pursuit of long-term prediction error minimisation. However, while this structure leads to improved performance on certain perceptual tasks in ASC, it also leads to a less comprehensive and less substantive long-term self-model.

On the one hand, this could manifest in an overly generalised self-model, for which small pieces of information that would normally be considered short-term inconsistencies may be overextended. This is supported by anecdotal evidence from autistic autobiography (Willey 2014; Williams 2009; Hammond 2010). For example, Williams creates an entire alternate identity based on a single interaction with a young girl at a park, which influences her self-concept for the rest of her life. She says, "This stranger, who I had only met once, was to change my life… Later, I became Carol." (Williams 2009). For Willey, she describes her process of self-realisation as follows, "I mark my life by moments in time, captured like morning glories at dawn, small and simple, yet fine and real. Moments define me, they make me complete. I envision the times that come together to form who I am." (Willey 2014). She also says that she is conscious of emphasising small moments in her memory, "I … found ways to elaborate a few isolated examples into what would then pass for a myriad of good times. … Looking far over my shoulder, I can call to mind people who must have been interested in my friendship. I can see a boy I knew as if it was yesterday. I can hear conversations we had and interests we shared. But more important, I can remember his face and the expressions he made as we talked. Today if he looked at me like he did then, I believe I would have seen the kindness and gentleness that was his. I never did much with this boy when I had the chance." (Willey 2014). Hammond offers us an insight to this happening in another domain, that of emotional associations with certain contexts, "Once Rebecca, who was an older girl at the school, looked over the toilet stall when I was going to the toilet. She was saying some not very nice things and it really freaked me out about using the toilets there. Even today I still hate public toilets because of this incident." (Hammond 2010).

Conversely understood, in individual cases, the flat self-model could mean a more accurate self-understanding within very specific contexts than in typically developing individuals. Rather than one temporally consistent and context general self, autistic people may have a less unified, coherent self-model with different primary traits in different contexts. Each of these ways of being might be a more accurate representation of the model's performance within that particular context. This may be reflected in the weaker responsiveness to impacts of behaviour on social reputation seen in autism, in which long-term social standing is disregarded for short-term prediction error minimisation (Izuma et al. 2011). It is also consistent with the multiplicity and fragmentation described in the autistic autobiography above.

On the other hand, if the context general and precise self-representation is retained, it is likely to be poorly fit to most environments. This is comparable to omitting representations of long-term regularities in explaining shifting short-term properties in other domains. For example, if you wanted to explain the regularity of a sense of comfort and warmth as you walk into work in the morning, you should not only account for the first-order variance of this feeling coinciding with the presence of a particular co-worker or a certain amount of sunshine entering the window on that day, but also for the possible variance of those causes such as a chronic illness for that co-worker or particular seasons. These deeper hidden causes, which can be understood as responsible for changing contexts, contribute to our expectations for volatility (changing variance) in the short-term expectations. In the case of the autistic self, this would mean that long-term invariances (as captured in the idea of personal identity over time and space) are more poorly represented than short-term, context-specific self-inferences. This is akin to an overfitted self-model, consistent with the volatility overestimating hypothesis presented in Lawson et al. (2017). The thickness of the self-representation and its behavioural consequences for each individual accounts for the observed variation in the presentation of ASC across individuals. Individual variation in the shape of the hierarchy would manifest itself in qualitative and clinically relevant differences in behaviour.

While our proposal is at the level of information processing and not physical neuronal instantiation, there are neural findings consistent with our model. Specifically, structural magnetic resonance imaging (MRI) of white matter connectivity and functional network connectivity (using fMRI) in autistic populations show enhanced short-range connectivity and reduced long-range connections involved in more integrated processing (Belmonte et al. 2004; Courchesne and Pierce 2005; O'Connor and Kirk 2008). These findings have been used to explain hyperacuity in low-level auditory tasks, while performing worse in speech processing tasks (O'Connor 2012). Neural findings consistent with the weak central coherence theory of autism will generally also be consistent with the predictive processing accounts. Existing neurobiological accounts of autism, such as abnormal neuronal weighting of excitation and inhibition (Rubenstein and Merzenich 2003) and the suggested GABAergic deficit (Robertson et al. 2016), may also prove to be interestingly related to the physical instantiation of this proposal, especially with the potential involvement of the oxytocin system (Quattrocki and Friston 2014). However, further research is required to elucidate this relationship.

The notion of context-specific self-models in ASC may explain findings related to memory for self in autism. In this framework, we see trait-based memories or semantic memory for self as consisting of long-term generalisation of self across different temporal and physical contexts. Episodic memory is infused with a first-person perspective which is temporally and physically specified. Under this model, the less robust self (over time) would rely more heavily on the first-person experience and, therefore, on episodic memory rather than models of the self that generalise heavily across time and space (as in semantic memory for self-traits). This would be supported by findings showing that autistic people show diminished self-reference effect for traits and with the findings by Goddard et al. (2017), which showed that autistic people relied more on episodic memories more than trait memory to reveal aspects of their personalities. When the first-person perspective is intentionally removed from control stimuli, as in Bowler et al. (2000), memories for words in autistic participants were reliant on semantic rather than episodic processes. This might emphasise the specific importance for first-person experience in the construction of a self-model in autism.

In the literature investigating visual self-recognition in autism, there were no consistent deficits reported. Visual self-recognition, as in mirrors and film, is likely very context dependent. There are certain contexts where one might expect to see one's own face, as in reflective surfaces. Further research might investigate the flexibility to extend visual self-recognition to unexpected contexts using virtual reality (for example, encountering and interacting with an avatar of oneself which does not mirror behaviour) to test the extent to which context might more quickly erode the less robust autistic self-recognition in various contexts.

Self as Hidden Cause

Our proposal that the self in autism is manifested by a flat model has implications for the self as an inferred hidden cause. As we expressed above, representations higher in the hierarchy, of which we propose there are fewer in autism, are more temporally invariant. Representations lower in the hierarchy are more transient and, thus, depend more on the first-person perspective and momentary lived experience. For self-representation, the flatter structure we propose in autism would lead to a more variable self-representation, which is more dependent on the first-person perspective. In some ways, this is counterintuitive; a more robust and deep self-structure leads to less reliance on specific experiences.

This would mean that the representation of self as cause is less stable in autism. The prior for an ever-present, invariant self who affects change in the world and effectively minimises prediction error over long durations is weaker under this model structure. For an autistic person, the answer to the question "Who am I?" may be more heavily weighted on the current sensory input.

This line of reasoning is highly related to paradigms that investigate sense and judgement of agency. In these paradigms, participants are measured on their ability to determine when they have causal control over the world. Although it has been consistently reported that people with autism show similar performance to neurotypical participants on wholly predictable consequences in tests of judgement of agency, there is evidence of reduced performance on sense of agency tasks relating causal agency to expected temporal durations (shorter for intentional actions) (Sperduti et al. 2014). This might indicate a reduced robustness in identifying the sensory consequences of one's own actions in autism. We have also discussed evidence that identifying one's own intentions following unintended consequences is reduced in autism (Phillips 1993; Phillips et al. 1998; Williams and Happé 2010b). There is also suggestion that autistic people rely more heavily on external cues to agency rather than proprioceptive information (Zalla et al. 2015). This evidence suggests an imprecise representation of the sense of self as a cause. The internal model of self is less robust and is more susceptible to deviations from expected input.

There is also ample evidence that there are high rates of alexithymia in autism. The increased prediction error in relation to self over the long run that would be expected in this kind of model (due to overfitting) could be directly contributing to differences in autistic emotional processing and the high rates of anxiety in autism. Emotion has been cast in the language of prediction error as reflecting the rate of prediction error minimisation (and is thus directly relevant to the self as meta-model) (Van de Cruys 2017). That is, high rates of anxiety may be related to lower than expected rates of prediction error minimisation due to a failure to adequately account for volatility in sensory input. Further, an imprecise self-model would not be able to accurately identify the internal hidden causes of own internal states and emotional responses, leading to alexithymia. The results above also show that interoception is impaired in autism. Being able to account for one's own internal states involves a strong sense that things within oneself can cause sensory feedback (without a conscious intent to self-stimulate). Without a robust model that one's own body is a source of a lot of sensory input, it would require much more prediction error to activate the model of self as cause of these sensations.

Personal pronoun use can also be understood as indicative of robust and consistent identification of selves in the environment. By using the word 'I', we can easily communicate when the agent of a verb (or action) is the unified model which we accept can cause environmental change. Avoiding identification of selves in language may indicate some confusion or lack of specificity in identifying long-term sources of environmental change. By saying "the dog was walked" rather than "I walked the dog", we can avoid linguistically identifying the causal source of the act. This is also consistent with a less robust sense of temporally-persistent self. It acknowledges changes in the environment but does not identify these changes with a particular unified causal actor, persisting over longer time periods.

Learning, Model Adjustment and Integration of Priors and Evidence

> As long as things followed a set of rules, I could play along. Rules were - and are - great friends of mine. I like rules. They set the record straight and keep it that way. You know where you stand with rules and you know how to act with rules. Trouble is, rules change and if they do not, people break them.
> (Willey 2014)

Assuming a flatter information processing structure in autism would also have consequences for the ways in which the model is changed by surprising evidence and how it accounts for prediction error across the hierarchy of hidden causes. The information that propagates through the system will also shape the self-model in a loop.

There is evidence from body ownership paradigms that different sources of sensory information with relevance to the self-model are differently integrated in autism. Specifically, there is evidence that autistic subjects rely more on proprioceptive information to bodily ownership (Greenfield et al. 2015). However, we have also seen evidence in agency studies that there is an overreliance on external cues as compared to proprioceptive cues in attributions of agency in autism (Zalla et al. 2015). These together suggest that the balance between different cues to self-involvement or presence is different in autism and suggests poor context sensitivity for this integration.

Note that other predictive processing theories propose differences at this level of detail—changing the weighting or values of different aspects of an individual calculation between two nodes as in Fig. 3 (and then generalised across the system). Specifically, here we refer to the HIPPEA, overlearning of volatility and weak priors accounts (Pellicano and Burr 2012; Van de Cruys et al. 2014; Lawson et al. 2017). As the system develops with these detailed discrepancies, the model of the world would be shaped by the weighted prediction errors. This would impact how the self-model (and the computational structure overall) is built and maintained by the system. Thus, our proposal may be compatible with these other observations and hypotheses.

Attention and Accumulation of Model Evidence

The self gains evidence for itself through the effectiveness of the model in reducing prediction error. If the overall model is poorer at reducing long-term prediction error, there would be less evidence for the self-model over long time scales. Conversely, if there is better prediction error minimisation in the short term, then there would be stronger evidence for the self-model's existence at shorter time scales.

There are various ways that the self can increase evidence for its own existence. One way to do this is through active inference. If one can selectively sample sensory stimuli that is perfectly (or at least more) predictable, it would improve the evidence we have for our own existence (which is likely a very strong evolutionary prior). This may explain the self-stimulation behaviours of people with autism (aka. stimming). While the individual can perform a repetitive and highly predictable action upon themselves (such as arm flapping, flicking their fingers in front of their eyes or even spinning to control visual stimuli), the world becomes more predictable than it was before, and the individual has greater evidence for the self-model as it explains away that incoming error, though predominantly in the short term.

Perseveration on highly regular stimuli and environments may also be the autistic person's attempt to combat unexpectedly low rates of prediction error minimisation. Over-attention to one stimulus in the environment in predictive coding accounts is demonstrative of an attempt to increase precision. This is also called the monotropic tendency in autistic attention (Murray et al. 2005). This may happen at higher levels of the hierarchy, such as in savant knowledge areas (which are often of highly regular domains) or in lower levels of the hierarchy, such as in stimming. This is consistent with the model of autism proposed by Van de Cruys et al. (2014) which asserts that inflexibly high weight is given to prediction errors, which would encourage the active inference solutions to stabilise the environment by creating predictable stimulation of sensory apparatus, or situating the self in highly predictable environments. This is especially true where all dominant sensations are self-controlled, and thus, there is no conflicting prediction error coming from uncontrollable aspects of the world. In non-stimming situations, poorly processing prediction error from other sources may overwhelm the system and reduce the evidence for one's own existence. Attention that can be paid to sources such as when another person calls your name may be reduced, and such stimuli might not become salient. This is consistent with the observation that children with autism often fail to orient to their name when called. This could also explain somewhat why internal states are not attended to—without strong evidence for one's own existence, autistic people may attribute sensations coming from within to external sources, which may further increase prediction error as it fails to capture the truth and further reduces evidence for self.

Developing a Self by Active Inference

Relatedly, one of the ways in which the self is shaped and developed is through exploration of the world through active inference. It constitutes one of the ways we learn about ourselves as causal powers and model ourselves as somewhat consistent over time and space. If the self is less robust to perturbations in expected sensory outcomes, the system may not use active inference in the same way as a more stable system.

Recent work by Constant et al. (2018) emphasises the particular role active inference under an autistic self-model that is 'collapsed' might play in explaining autistic behaviours. Autistic individuals create a very distinct environmental niche through their self-imposed strict rules and repetitive behaviours. The variability of this created niche is directly related to the temporal thickness of a model (i.e. the depth of the hierarchy) (Friston 2018). Prima facie, the overly restricted behaviours of autistic individuals do not seem compatible with the hypothesis that the self-model is constantly changing and in some ways overly flexible. However, because action and perception are part of a highly integrated causal loop, the overspecification of the self-model (at the lower levels of the hierarchy) can be the direct cause of significantly more restricted environmental niches. The model tries to minimise prediction error from a very flexible self-model by acting on the world to enforce the restricted environments in which that self-model effectively reduces prediction error. This would explain the particular and focused energy with which some autistic individuals engage in their chosen areas of expertise, and also the insistence on sameness and learned routines, which are highly predictable (Lawson and Dombroski 2015). As such, in the world that an autistic individual creates for themselves, if it could be perfectly controlled and isolated, this model (Fig. 3) would predict that they would be better at self-modelling than the neurotypical model in its own created environment. Though a new area of research, this proposed flexibility in self-representation might be related to recent findings about the correlation between ASC and gender variance and dysphoria (Heylens et al. 2018; Strang et al. 2014; Vermaat et al. 2018; Janssen 2018; Øien et al. 2018).

This line of reasoning may be investigated using agency paradigms in which self-identification can only be made through active inference in a noisy environment. Such a paradigm would help to highlight the ways different solutions to the problems of perception (Fig. 1) are each utilised depending on the environmental conditions (variability and volatility).

Motor dysfunction so frequently occurs in autism (as confirmed by a formal meta-analysis by Fournier et al. (2010)) that some have suggested it should be included as a core clinical feature. Motor differences include generally slower and less accurate movements that have been candidate signs for an autism diagnosis include slower limb movements, poor and slow manual dexterity, poor ball skills (throwing and catching), instability, impaired gait (such as heel or toe walking), reduced coordination in locomotor skills and stiffness (Gowen and Hamilton 2013). It has been suggested that these differences originate in higher, integrative parts of the cognitive hierarchy (Gowen and Hamilton 2013). Notably, dysfunction in self-representation including of body positioning, possible states of being, long-term policies about optimal methods for goal attainment are intimately interrelated with our motor system, as the self is expressed in active inference through the bodily effectors. The known impairments in the motor system in autism would be consistent with both sensory dysfunction (as signals get sub-optimally integrated) and from a weak self-concept (with less top-down shaping of action representation).

More conclusive evidence in autistic populations for particular peculiarities or similarities in the area for memory for own actions may also speak to this line of enquiry.

Implications for Future Research

It will be challenging to generate and test hypotheses from predictive coding model. There are many moving parts to the information processing model as a whole, and capturing or testing how the model of the model (that is, the self) is affected by differences in its structure can require controlling for many variables. However, there is encouraging research outside of the literature on the self, which successfully engages with predictive coding hypotheses.

The predictive processing account emphasises environmental regularities at various time scales and how an internal model driven by prediction errors might dynamically capture these regularities. As such, experimental designs that manipulate probabilities in stimuli, specifically their variability and volatility, will prove most useful.

Generally, in cognitive science, progress in the area of predictive processing has been made through simulation studies (Friston and Frith 2015; Kanai et al. 2015) as well as using methods from the neural (Fardo et al. 2017; Iglesias et al. 2013; Starkweather et al. 2017) and psychological domains (Vossel et al. 2014). In autism, some work has already been done in testing some of the theories which are generated by the predictive processing view by manipulating these variables (Skewes and Gebauer 2016; Skewes et al. 2015; Lawson et al. 2017; Manning et al. 2016; Smith et al. 2017; von der Lühe et al. 2016; Gonzalez-Gadea et al. 2015).

As confidence builds that predictive processing theories prove fruitful in autism research, the crucial next step is to apply these methods to the paradigms listed above to test how the informational processing structure in autism might also affect the self-model. Based on the theoretical predictions made here, we would expect that autistic individuals would show decreased performance on tasks that depend on accurate modelling of volatility and cues to changing contexts. For example, we would expect that the autistic self-concept might be relatively specific to time and place, and involve fewer accurate long-term generalisations. Investigating metacognition with regard to long-term traits as compared to family member or close friend evaluations may reveal this. In paradigms requiring integrating information about long-term priors and immediate sensory information, such as in tactile illusions or intentional binding in ecologically valid settings, we might expect more accurate (and less illusory) perception in autistic subjects.

When the autistic person has access to effective active inference, we expect perseveration in specific action outcome contingencies, rather than exploratory actions for new rewards. That is, where the individual can reliably control their environment such as to occupy a low prediction error-producing state, we expect autistic subjects to be more inclined to occupy that state, even if they could learn something new about themselves as a hidden cause by trying novel actions; in other words, we suggest as a useful construct a self-related exploration–exploitation distinction balance, which might be maintained differently in autism. This may also impact how autistic people recognise and remember their own control in action. If actions were not externally dictated in an agency-based task, different results may be obtained than when actions are pre-specified. Further, tasks which are optimally performed through exploratory risk taking, such as volatile probability reward tasks, may be sub-optimally completed by autistic participants, revealing this discrepancy in their use of active inference. In tasks where there is one, stable optimal action, autistic participants may perform better, by assuming no change in the underlying probabilities. Additionally, we might expect prediction effects for memory for own action. If the outcome of an action was strongly predicted, we might expect a better memory for that action due to its attribution to self in autism. In neurotypical participants, however, we might expect that actions which have unexpected outcomes, but that still maintained a sense of control, would be the strongest remembered.

Visual perspective taking is one area that involves counterfactual reasoning and using priors to reconstruct potential current sensory input. Under our proposal, we would expect poorer performance from autistic subjects on this kind of task as it would be hard to suppress the current sensory input to reason about an alternative state in space without deeper nodes mediating this discrepancy in perceived position based on prior expectations.

Based on our proposed differences in autistic self-representation, we would also expect autistic participants to be better at temporary self-association tasks as they have a flexible and more short-term self-model, which might be more easily shifted by suggestions like this—however, they would likely need more than just a verbal instruction to induce such an incorporation into the self-concept. Despite being less influenced by strong priors, autistic subjects will expect less volatility in underlying probabilities and, thus, may be unable to stably eliminate prediction error caused by the unexpected mismatches across sensory domains. Providing more consistent and lasting evidence for the new self-association may ensure its adoption. Self-prioritisation in general may rely on associations between signals activating the self qua cause and the increased likelihood that one will need to act. With a less robust sense of self as cause, we might expect that these paradigms should show a deficit for autistic subjects.

Mistaken inferences about the self commonly occur in typically developing adults, and these we expect would differ in the autistic self. This need not always be interpreted as an autistic deficit, as psychologically beneficial misperceptions of self might enhance mental health if perceived more keenly by autistic subjects, or, if alleviated, a more veridical perception of self for the particular bias may equally confer benefits. For example, one of these areas that seem particularly promising is the superiority illusion. This effect reflects the widespread perception that oneself is superior to average. It has been shown that this illusion is dependent on the dopaminergic system in the brain (Yamada et al. 2013), which has also been implicated in mediation of precision expectations and balancing of top-down and bottom-up information under predictive processing (Friston et al. 2012). It can be expected based on our proposal that this illusion should differ in autism.

Further, expected differences in autistic self-processing should be experimentally contrasted with more classical disorders of the self. These include dissociative identity disorder, Cotard’s delusion, disorders of bodily awareness such as phantom limbs and asomatognosia, ego dissolution under psychoactive substances and schizophrenia. This would improve understanding of the particular form the self takes in autism, and how its processing is unique (or not) from other disorders of the self. Some of these have already been described using a predictive processing framework (Letheby and Gerrans 2017; Adams 2018; Corlett 2017; Fletcher and Frith 2009; Seth et al. 2012), more of which would allow for hypothesis-driven, well-controlled contrasts between the self-processing in these conditions.

A good theory should equally have some predictions which would disconfirm the proposal. Here, if we found typical or improved ability to form accurate generalisations about oneself across long time periods (and across domains), this might provide evidence against the flatter structure we propose here. Stronger still, if all past research was found inaccurate and we saw no deficit or improved performance in autistic populations across the self-cognition paradigms, we would have no reason to expect an underlying difference in the autistic self. Intact memory for self especially would be inconsistent with our proposal given its time-sensitive nature. Similarly, we would expect this to extend to future projections of the self.

The structure of the autistic self may lead autistic agents to ignore long-term patterns of behaviour stemming from deep internal agentive causes (the deep, causally effective self-concept) and attribute own behaviour to noise or external cues. This would impact both how the person perceives their role in the world and how they choose to interact with the world. Long-term goals for changing context might be both absent and undervalued in choosing day-to-day actions under a flat self-model.

Conclusion

In contrast to other theories of the autistic self, the predictive processing account suggests that the autistic self is an authentic self in just the same sense as the neurotypical self. It is an inference that prediction error minimisation is being performed by a deeply hidden cause—the self qua efficacious causal process. The shape of this self-model simply differs as a result of the time scales over which prediction error is predominantly minimised. The autistic self is thus not any more secluded from the world, as was suggested by early observers. The autistic self-model marks out a particular way of existing in the world: of being a self. Despite the complexities and controversies in explanations of both the self and autism, there seems to be a viable path forward for research coming from the predictive processing perspective in both areas.

Acknowledgements The authors would like to thank Uta Frith and Peter Enticott for their detailed comments on an earlier draft of this paper.

Funding Information This research was supported by an Australian Government Research Training Program (RTP) Scholarship and Australian Research Council Discovery Grant DP160102770.

References

Adams, R. A. (2018). Chapter 7—Bayesian inference, predictive coding, and computational models of psychosis A2—Anticevic, Alan. In J. D. Murray (Ed.), *Computational psychiatry* (pp. 175–195): Academic.

Alexopoulos, T., Muller, D., Ric, F., & Marendaz, C. (2012). I, me, mine: automatic attentional capture by self-related stimuli. *European Journal of Social Psychology, 42*(6), 770–779. https://doi.org/10.1002/ejsp.1882.

Allen, R., Davis, R., & Hill, E. (2013). The effects of autism and alexithymia on physiological and verbal responsiveness to music. *Journal of Autism and Developmental Disorders, 43*(2), 432–444. https://doi.org/10.1007/s10803-012-1587-8.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychobiology, 5*(4), 297–305.

Apps, M. A., & Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neuroscience and Biobehavioral Reviews, 41*, 85–97. https://doi.org/10.1016/j.neubiorev.2013.01.029.

Asada, K., Tojo, Y., Hakarino, K., Saito, A., Hasegawa, T., & Kumagaya, S. (2017). Brief report: Body image in autism: evidence from body size estimation. *Journal of Autism and Developmental Disorders*, 1–8.

Atance, C. M., & O'Neill, D. K. (2004). Acting and planning on the basis of a false belief: its effects on 3-year-old children's reasoning about their own false beliefs. *Developmental Psychology, 40*(6), 953–964. https://doi.org/10.1037/0012-1649.40.6.953.

Baess, P., & Prinz, W. (2017). Face/agent interference in individual and social context. *Social Cognition, 35*(2), 146–162.

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto alexithymia scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*(1), 23–32. https://doi.org/10.1016/0022-3999(94)90005-1.

Baggs, A. M. (2013). Autistic authors booklist and facts. *Autonomy, the Critical Journal of Interdisciplinary Autism Studies,* 1(2).

Baird, G., Cass, H., & Slonims, V. (2003). Diagnosis of autism. *BMJ, 327*, 488–493.

Baker-Ward, L., Hess, T. M., & Flannagan, D. A. (1990). The effects of involvement on children's memory for events. *Cognitive Development, 5*(1), 55–69.

Balslev, D., Cole, J., & Miall, R. C. (2007). Proprioception contributes to the sense of agency during visual observation of hand movements: evidence from temporal judgments of action. *Journal of Cognitive Neuroscience, 19*(9), 1535–1541. https://doi.org/10.1162/jocn.2007.19.9.1535.

Baltaxe, C. A. M. (1977). Pragmatic deficits in the language of autistic adolescents. *Journal of Pediatric Psychology, 2*(4), 176–180. https://doi.org/10.1093/jpepsy/2.4.176.

Baron-Cohen, S. (1989a). Are autistic children "behaviorists"? An examination of their mental-physical and appearance-reality distinctions. *Journal of Autism and Developmental Disorders, 19*(4), 579–600.

Baron-Cohen, S. (1989b). Perceptual role taking and protodeclarative pointing in autism. *British Journal of Developmental Psychology, 7*(2), 113–127. https://doi.org/10.1111/j.2044-835X.1989.tb00793.x.

Baron-Cohen, S. (1991). The development of a theory of mind in autism: deviance and delay? *Psychiatric Clinics of North America, 14*(1), 33–51.

Baron-Cohen, S. (1992). Out of sight or out of mind? Another look at deception in autism. *Journal of Child Psychology and Psychiatry, 33*(7), 1141–1155.

Baron-Cohen, S., Scott, F. J., Allison, C., Williams, J., Bolton, P., Matthews, F. E., et al. (2009). Prevalence of autism-spectrum conditions: UK school-based population study. *The British Journal of Psychiatry, 194*(6), 500–509. https://doi.org/10.1192/bjp.bp.108.059345.

Barresi, J., & Martin, R. (2011). History as prologue: western theories of the self. In S. Gallagher (Ed.), *The Oxford handbook of the self* (pp. 33–57). Oxford: Oxford University Press.

Bartak, L., & Rutter, M. (1974). The use of personal pronouns by autistic children. *Journal of Autism and Childhood Schizophrenia, 4*(3), 217–222. https://doi.org/10.1007/bf02115227.

Barttfeld, P., Wicker, B., Cukier, S., Navarta, S., Lew, S., Leiguarda, R., et al. (2012). State-dependent changes of connectivity patterns and functional brain network topology in autism spectrum disorder. *Neuropsychologia, 50*(14), 3653–3662. https://doi.org/10.1016/j.neuropsychologia.2012.09.047.

Bednark, J. G., Poonian, S., Palghat, K., McFadyen, J., & Cunnington, R. (2015). Identity-specific predictions and implicit measures of agency. *Psychology of Consciousness: Theory, Research, and Practice, 2*(3), 253.

Begeer, S., Malle, B. F., Nieuwland, M. S., & Keysar, B. (2010). Using theory of mind to represent and take part in social interactions: comparing individuals with high-functioning autism and typically developing controls. *European Journal of Developmental Psychology, 7*(1), 104–122.

Belmonte, M. K., Allen, G., Beckel-Mitchener, A., Boulanger, L. M., Carper, R. A., & Webb, S. J. (2004). Autism and abnormal development of brain connectivity. *Journal of Neuroscience, 24*(42), 9228–9231.

Berberian, B., & Cleeremans, A. (2010). Endogenous versus exogenous change: change detection, self and agency. *Consciousness and Cognition, 19*(1), 198–214.

Bertenthal, B. I., & Fischer, K. W. (1978). Development of self-recognition in the infant. *Developmental Psychology, 14*(1), 44–50. https://doi.org/10.1037/0012-1649.14.1.44.

Berthoz, S., & Hill, E. L. (2005). The validity of using self-reports to assess emotion regulation abilities in adults with autism spectrum disorder. *European Psychiatry, 20*(3), 291–298. https://doi.org/10.1016/j.eurpsy.2004.06.013.

Bertilsson, I., Gyllensten, A. L., Opheim, A., Gard, G., & Sjödahl Hammarlund, C. (2018). Understanding one's body and movements from the perspective of young adults with autism: a mixed-methods study. *Research in Developmental Disabilities, 78*, 44–54. https://doi.org/10.1016/j.ridd.2018.05.002.

Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry, 3*, e285. https://doi.org/10.1038/tp.2013.61.

Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain, 133*(5), 1515–1525. https://doi.org/10.1093/brain/awq060.

Bird, G., Press, C., & Richardson, D. C. (2011). The role of alexithymia in reduced eye-fixation in autism spectrum conditions. *Journal of Autism and Developmental Disorders, 41*(11), 1556–1564. https://doi.org/10.1007/s10803-011-1183-3.

Blakemore, S.-J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience, 11*(5), 551–559. https://doi.org/10.1162/089892999563607.

Blakemore, S.-J., Tavassoli, T., Calò, S., Thomas, R. M., Catmur, C., Frith, U., et al. (2006). Tactile sensitivity in Asperger syndrome. *Brain and Cognition, 61*(1), 5–13. https://doi.org/10.1016/j.bandc.2005.12.013.

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature, 391*(6669), 756.

Bowler, D. M., Gardiner, J. M., & Grice, S. J. (2000). Episodic memory and remembering in adults with Asperger syndrome. *Journal of Autism and Developmental Disorders, 30*(4), 295–304.

Bradford, E. E. F., Hukker, V., Smith, L., & Ferguson, H. J. (2018). Belief-attribution in adults with and without autistic spectrum disorders. *Autism Research*, 0(0). https://doi.org/10.1002/aur.2032.

Brock, J. (2012). Alternative Bayesian accounts of autistic perception: comment on Pellicano and Burr. *Autism, 14*, 209–224.

Bruck, M., London, K., Landa, R., & Goodman, J. (2007). Autobiographical memory and suggestibility in children with autism spectrum disorder. *Development and Psychopathology, 19*(01), 73–95.

Burling, J. M., Kadambi, A., Safari, T., & Lu, H. (2019). The impact of autistic traits on self-recognition of body movements. *Frontiers in Psychology*, 9(2687). https://doi.org/10.3389/fpsyg.2018.02687.

Burton, S., & Mitchell, P. (2003). Judging who knows best about yourself: developmental change in citing the self across middle childhood. *Child Development, 74*(2), 426–443. https://doi.org/10.1111/1467-8624.7402007.

Cahrel, S., Poiroux, S., Bernard, C., Thibaut, F., Lalonde, R., & Rebai, M. (2002). ERPs associated with familiarity and degree of familiarity during face recognition. *International Journal of Neuroscience, 112*(12), 1499–1512. https://doi.org/10.1080/00207450290158368.

Capps, L., Sigman, M., & Yirmiya, N. (1995). Self-competence and emotional understanding in high-functioning children with autism. *Development and Psychopathology, 7*(1), 137–149.

Carruthers, P. (2009). How we know our own minds: the relationship between mindreading and metacognition. *The Behavioral and Brain Sciences, 32*(2), 121–138; discussion 138-182. https://doi.org/10.1017/S0140525X09000545.

Cascio, C. J., Foss-Feig, J. H., Burnette, C. P., Heacock, J. L., & Cosby, A. A. (2012). The rubber hand illusion in children with autism spectrum disorders: delayed influence of combined tactile and visual input on proprioception. *Autism, 16*(4), 406–419. https://doi.org/10.1177/1362361311430404.

Chakraborty, A., & Chakrabarti, B. (2015). Is it me? Self-recognition bias across sensory modalities and its relationship to autistic traits. *Molecular Autism, 6*(1), 20. https://doi.org/10.1186/s13229-015-0016-1.

Chang, L., Zhang, S., Poo, M.-m., & Gong, N. (2017). Spontaneous expression of mirror self-recognition in monkeys after learning precise visual-proprioceptive association for mirror images. *Proceedings of the National Academy of Sciences, 114*(12), 3258–3263.

Charney, R. (1980). Pronoun errors in autistic children: support for a social explanation. *International Journal of Language & Communication Disorders, 15*(1), 39–43. https://doi.org/10.3109/13682828009011369.

Charney, R. (2008). Speech roles and the development of personal pronouns. *Journal of Child Language, 7*(3), 509–528. https://doi.org/10.1017/S0305000900002816.

Christensen, D. L., Baio, J., Braun, K. V. N., Bilder, D., Charles, J., Constantino, J. N., et al. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2012. MMWR. Surveillance Summaries*,* 65.

Clark, A. (2015). *Surfing uncertainty: prediction, action, and the embodied mind*: Oxford University Press.

Constant, A., Bervoets, J., Hens, K., & Cruys, S. V. d. (2018). Precise worlds for certain minds: an ecological perspective on the relational self in autism. *TOPOI* (The Relational Self - Basic Forms of Self-Awareness), https://doi.org/10.1007/s11245-018-9546-4.

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science, 24*(5), 723–732. https://doi.org/10.1177/0956797612463582.

Corlett, P. R. (2017). I predict, therefore I am: perturbed predictive coding under ketamine and in schizophrenia. *Biological Psychiatry, 81*(6), 465–466. https://doi.org/10.1016/j.biopsych.2016.12.007.

Courchesne, E., & Pierce, K. (2005). Brain overgrowth in autism during a critical time in development: implications for frontal pyramidal neuron and interneuron development and connectivity. *International Journal of Developmental Neuroscience, 23*(2), 153–170. https://doi.org/10.1016/j.ijdevneu.2005.01.003.

Craig, A. D. (2003). Interoception: the sense of the physiological condition of the body. *Current Opinion in Neurobiology, 13*(4), 500–505. https://doi.org/10.1016/S0959-4388(03)00090-4.

Crane, L., & Goddard, L. (2008). Episodic and semantic autobiographical memory in adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(3), 498–506.

Crane, L., Goddard, L., & Pring, L. (2009). Specific and general autobiographical knowledge in adults with autism spectrum disorders: the role of personal goals. *Memory, 17*(5), 557–576.

Crucianelli, L., Krahé, C., Jenkinson, P. M., & Fotopoulou, A. (2016). Interoceptive ingredients of body ownership: affective touch and cardiac awareness in the rubber hand illusion. *Cortex*, https://doi.org/10.1016/j.cortex.2017.04.018.

Cygan, H. B., Tacikowski, P., Ostaszewski, P., Chojnicka, I., & Nowicka, A. (2014). Neural correlates of own name and own face detection in autism spectrum disorder. *PLoS One, 9*(1), e86020. https://doi.org/10.1371/journal.pone.0086020.

Cygan, H. B., Marchewka, A., Kotlewska, I., & Nowicka, A. (2018). Neural correlates of reflection on present and past selves in autism spectrum disorder. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-018-3621-y.

Damon, W., & Hart, D. (1986). Stability and change in children's self-understanding. *Social Cognition, 4*(2), 102–118.

Dascalu, C.-M. (2018). Is the self-reference of autistic children atypical? The case of two french autstic children. *Language Processing and Disorders*, 236.

David, N., Gawronski, A., Santos, N. S., Huff, W., Lehnhardt, F.-G., Newen, A., et al. (2008a). Dissociation between key processes of social cognition in autism: impaired mentalizing but intact sense of agency. *Journal of Autism and Developmental Disorders, 38*(4), 593–605. https://doi.org/10.1007/s10803-007-0425-x.

David, N., Newen, A., & Vogeley, K. (2008b). The "sense of agency" and its underlying cognitive and neural mechanisms. *Consciousness and Cognition, 17*(2), 523–534. https://doi.org/10.1016/j.concog.2008.03.004.

Dawson, G., & Fernald, M. (1987). Perspective-taking ability and its relationship to the social behavior of autistic children. *Journal of Autism and Developmental Disorders, 17*(4), 487–498. https://doi.org/10.1007/bf01486965.

Dawson, G., & McKissick, F. C. (1984). Self-recognition in autistic children. *Journal of Autism and Developmental Disorders, 14*(4), 383–394.

De Preester, H., & Tsakiris, M. (2009). Body-extension versus body-incorporation: is there a need for a body-model? *Phenomenology and the Cognitive Sciences, 8*(3), 307–319. https://doi.org/10.1007/s11097-009-9121-y.

Dissanayake, C., Shembrey, J., & Suddendorf, T. (2010). Delayed video self-recognition in children with high functioning autism and Asperger's disorder. *Autism, 14*(5), 495–508. https://doi.org/10.1177/1362361310366569.

Dritschel, B. M., Wisely, M., Goddard, L., Robinson, S., & Howlin, P. (2010). Judgements of self-understanding in adolescents with Asperger syndrome. *Autism, 14*(5), 509–518. https://doi.org/10.1177/1362361310368407.

DuBois, D., Ameis, S. H., Lai, M.-C., Casanova, M. F., & Desarkar, P. (2016). Interoception in autism spectrum disorder: a review. *International Journal of Developmental Neuroscience, 52*, 104–111. https://doi.org/10.1016/j.ijdevneu.2016.05.001.

Dumontheil, I., Küster, O., Apperly, I. A., & Blakemore, S.-J. (2010). Taking perspective into account in a communicative task. *Neuroimage, 52*(4), 1574–1583. https://doi.org/10.1016/j.neuroimage.2010.05.056.

Dunphy-Lelii, S., & Wellman, H. M. (2012). Delayed self recognition in autism: a unique difficulty? *Research in Autism Spectrum Disorders, 6*(1), 212–223. https://doi.org/10.1016/j.rasd.2011.05.002.

Elmose, M., & Happé, F. (2014). Being aware of own performance: how accurately do children with autism spectrum disorder judge own memory performance? *Autism Research, 7*(6), 712–719.

Elsabbagh, M., Divan, G., Koh, Y. J., Kim, Y. S., Kauchali, S., Marcin, C., et al. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research, 5*(3), 160–179. https://doi.org/10.1002/aur.239.

Elwin, M., Ek, L., Schröder, A., & Kjellin, L. (2012). Autobiographical accounts of sensing in Asperger syndrome and high-functioning autism. *Archives of Psychiatric Nursing, 26*(5), 420–429. https://doi.org/10.1016/j.apnu.2011.10.003.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*(3), 327.

Fardo, F., Auksztulewicz, R., Allen, M., Dietz, M. J., Roepstorff, A., & Friston, K. J. (2017). Expectation violation and attention to pain jointly modulate neural gain in somatosensory cortex. *Neuroimage*.

Farley, A., Lopez, B., & Saunders, G. (2010). Self-conceptualisation in autism: Knowing oneself versus knowing self-through-other. *Autism, 14*(5), 519–530. https://doi.org/10.1177/1362361310368536.

Farrant, A., Blades, M., & Boucher, J. (1998). Source monitoring by children with autism. *Journal of Autism and Developmental Disorders, 28*(1), 43–50.

Fay, W. H. (1979). Personal pronouns and the autistic child. *Journal of Autism and Developmental Disorders, 9*(3), 247–260.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience, 4*.

Ferrari, M., & Matthews, W. S. (1983). Self-recognition deficits in autism: syndrome-specific or general developmental delay? *Journal of Autism and Developmental Disorders, 13*(3), 317–324.

Fiene, L., & Brownlow, C. (2015). Investigating interoception and body awareness in adults with and without autism spectrum disorder. *Autism Research, 8*(6), 709–716. https://doi.org/10.1002/aur.1486.

Fisch, G. S. (2012). Nosology and epidemiology in autism: classification counts. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics, 160C*(2), 91–103. https://doi.org/10.1002/ajmg.c.31325.

Fisher, N., Happe, F., & Dunn, J. (2005). The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and children with moderate learning difficulties. *Journal of Child Psychology and Psychiatry, 46*(4), 409–419. https://doi.org/10.1111/j.1469-7610.2004.00371.x.

Fivush, R. (2011). The development of autobiographical memory. *Annual Review of Psychology, 62*, 559–582.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience, 10*(1), 48–58.

Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., & Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: a synthesis and meta-analysis. *Journal of Autism and Developmental Disorders, 40*(10), 1227–1240.

Foxton, J. M., Stewart, M. E., Barnard, L., Rodgers, J., Young, A. H., O'Brien, G., et al. (2003). Absence of auditory 'global interference' in autism. *Brain, 126*(12), 2703–2709.

Freeman, N. H., & Lacohée, H. (1995). Making explicit 3-year-olds' implicit competence with their own false beliefs. *Cognition, 56*(1), 31–60. https://doi.org/10.1016/0010-0277(94)00654-4.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138.

Friston, K. (2017a). Self-evidencing babies: commentary on "Mentalizing homeostasis: the social origins of interoceptive inference" by Fotopoulou & Tsakiris. *Neuropsychoanalysis, 19*(1), 43–47. https://doi.org/10.1080/15294145.2017.1295216.

Friston, K. J. (2017b). Precision psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2*(8), 640–643. https://doi.org/10.1016/j.bpsc.2017.08.007.

Friston, K. (2018). Am I self-conscious? (or does self-organization entail self-consciousness?). [Hypothesis and theory]. *Frontiers in Psychology, 9*(579), https://doi.org/10.3389/fpsyg.2018.00579.

Friston, K., & Frith, C. (2015). A duet for one. *Consciousness and Cognition, 36*, 390–405. https://doi.org/10.1016/j.concog.2014.12.003.

Friston, K., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., et al. (2012). Dopamine, affordance and active inference. *PLoS Computational Biology, 8*(1), e1002327.

Friston, K., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry, 1*(2), 148–158. https://doi.org/10.1016/S2215-0366(14)70275-5.

Frith, U., & Happé, F. (1994). Autism: beyond "theory of mind". *Cognition, 50*(1), 115–132.

Frith, U., & Happé, F. (1999). Theory of mind and self-consciousness: what is it like to be autistic? *Mind & Language, 14*(1), 82–89.

Furlano, R., & Kelley, E. A. (2019). Do children with autism Sspectrum disorder understand their academic competencies? *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-019-03988-0.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences, 4*(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5.

Gallup, G. G., Jr. (1968). Mirror-image stimulation. *Psychological Bulletin, 70*(6, Pt.1), 782–793. https://doi.org/10.1037/h0026777.

Gallup, G. G., Anderson, J. R., & Platek, S. M. (2011). Self-recognition. In S. Gallagher (Ed.), *The Oxford handbook of the self.* Oxford University Press.

Garfinkel, S. N., Tiley, C., O'Keeffe, S., Harrison, N. A., Seth, A. K., & Critchley, H. D. (2016). Discrepancies between dimensions of interoception in autism: implications for emotion and anxiety. *Biological Psychology, 114*, 117–126. https://doi.org/10.1016/j.biopsycho.2015.12.003.

Gillespie-Smith, K., Ballantyne, C., Branigan, H. P., Turk, D. J., & Cunningham, S. J. (2017). The I in autism: severity and social functioning in autism are related to self-processing. *British Journal of Developmental Psychology, 36*, 127–141. https://doi.org/10.1111/bjdp.12219.

Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological Bulletin, 131*(1), 76.

Goddard, L., Howlin, P., Dritschel, B., & Patel, T. (2007). Autobiographical memory and social problem-solving in Asperger syndrome. *Journal of Autism and Developmental Disorders, 37*(2), 291–300. https://doi.org/10.1007/s10803-006-0168-0.

Goddard, L., O'Dowda, H., & Pring, L. (2017). Knowing me, knowing you: self defining memories in adolescents with and without an autism spectrum disorder. *Research in Autism Spectrum Disorders, 37*, 31–40. https://doi.org/10.1016/j.rasd.2017.02.002.

Gonzalez-Gadea, M. L., Chennu, S., Bekinschtein, T. A., Rattazzi, A., Beraudi, A., Tripicchio, P., et al. (2015). Predictive coding in autism spectrum disorder and attention deficit hyperactivity disorder. *Journal of Neurophysiology, 114*(5), 2625–2636.

Goodenough, F. L. (1938). The use of pronouns by young children: a note on the development of self-awareness. *The Pedagogical Seminary and Journal of Genetic Psychology, 52*(2), 333–346. https://doi.org/10.1080/08856559.1938.10534320.

Gowen, E., & Hamilton, A. (2013). Motor abilities in autism: a review using a computational context. *Journal of Autism and Developmental Disorders, 43*(2), 323–344. https://doi.org/10.1007/s10803-012-1574-0.

Grainger, C., Williams, D., & Lind, S. E. (2014). Online action monitoring and memory for self-performed actions in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 44*, 1193–1206.

Grandin, T. (1996). Thinking in pictures: autism and visual thought. Thinking in pictures: my life with autism.

Greenfield, K., Ropar, D., Smith, A. D., Carey, M., & Newport, R. (2015). Visuo-tactile integration in autism: atypical temporal binding may underlie greater reliance on proprioceptive information. *Molecular Autism, 6*(1), 51. https://doi.org/10.1186/s13229-015-0045-9.

Greenfield, K., Newport, R., Smith, A. D., Carey, M., & Ropar, D. (2017). Body representation difficulties in children and adolescents with autism may be due to delayed development of visuo-tactile temporal binding. *Developmental Cognitive Neuroscience*. https://doi.org/10.1016/j.dcn.2017.04.007.

Grèzes, J., Frith, C. D., & Passingham, R. E. (2004). Inferring false beliefs from the actions of oneself and others: an fMRI study. *Neuroimage, 21*(2), 744–750. https://doi.org/10.1016/S1053-8119(03)00665-7.

Griffin, C., Lombardo, M. V., & Auyeung, B. (2015). Alexithymia in children with and without autism spectrum disorders. *Autism Research, 9*, 773–780.

Grynszpan, O., Nadel, J., Martin, J.-C., Simonin, J., Bailleul, P., Wang, Y., et al. (2012). Self-monitoring of gaze in high functioning autism. *Journal of Autism and Developmental Disorders, 42*(8), 1642–1650.

Guerra, S., Spoto, A., Parma, V., Straulino, E., & Castiello, U. (2017). In sync or not in sync? Illusory body ownership in autism spectrum disorder. *Research in Autism Spectrum Disorders, 41–42*, 1–7. https://doi.org/10.1016/j.rasd.2017.07.003.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5*(4), 382–385.

Hala, S., Rasmussen, C., & Henderson, A. M. (2005). Three types of source monitoring by children with and without autism: the role of executive function. *Journal of Autism and Developmental Disorders, 35*(1), 75–89.

Hamilton, A. F. d. C. (2009). Research review: Goals, intentions and mental states: challenges for theories of autism. *Journal of Child Psychology and Psychiatry, 50*(8), 881–892.

Hamilton, A. F. d. C., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition, 113*(1), 37–44. https://doi.org/10.1016/j.cognition.2009.07.007.

Hammond, M. (2010). *My life with Asperger's*. Sydney: New Holland.

Happé, F. (1999). Autism: cognitive deficit or cognitive style? *Trends in Cognitive Sciences, 3*(6), 216–222.

Happe, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36*(1), 5–25. https://doi.org/10.1007/s10803-005-0039-0.

Hare, D. J., Mellor, C., & Azmi, S. (2007). Episodic memory in adults with autistic spectrum disorders: recall for self- versus other-experienced events. *Research in Developmental Disabilities, 28*(3), 317–329.

Harris, C. B., Rasmussen, A. S., & Berntsen, D. (2014). The functions of autobiographical memory: an integrative approach. *Memory, 22*(5), 559–581. https://doi.org/10.1080/09658211.2013.806555.

Hart, D., & Damon, W. (1986). Developmental trends in self-understanding. *Social Cognition, 4*(4), 388–407.

Hatfield, T. R., Brown, R. F., Giummarra, M. J., & Lenggenhager, B. (2017). Autism spectrum disorder and interoception: abnormalities in global integration? *Autism*, 1362361317738392, https://doi.org/10.1177/1362361317738392.

Heasman, B., & Gillespie, A. (2017). Perspective-taking is two-sided: misunderstandings between people with Asperger's syndrome and their family members. *Autism*, 1362361317708287, https://doi.org/10.1177/1362361317708287.

Heaton, P., Hudry, K., Ludlow, A., & Hill, E. (2008). Superior discrimination of speech pitch and its relationship to verbal ability in autism spectrum disorders. *Cognitive Neuropsychology, 25*(6), 771–782. https://doi.org/10.1080/02643290802336277.

Henderson, H. A., Zahka, N. E., Kojkowski, N. M., Inge, A. P., Schwartz, C. B., Hileman, C. M., et al. (2009). Self-referenced memory, social cognition, and symptom presentation in autism. *Journal of Child Psychology and Psychiatry, 50*(7), 853–861.

Heylens, G., Aspeslagh, L., Dierickx, J., Baetens, K., Van Hoorde, B., De Cuypere, G., et al. (2018). The co-occurrence of gender dysphoria and autism spectrum disorder in adults: an analysis of cross-sectional and clinical chart data. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-018-3480-6.

Higashida, N. (2013). *The reason I jump: one boy's voice from the silence of autism: one boy's voice from the silence of autism*: Hachette UK.

Higgins, E. T., Van Hook, E., & Dorfman, D. (1988). Do self-attributes form a cognitive structure? *Social Cognition, 6*(3), 177–206. https://doi.org/10.1521/soco.1988.6.3.177.

Hill, E., & Russell, J. (2002). Action memory and self-monitoring in children with autism: self versus other. *Infant and Child Development, 11*(2), 159–170.

Hill, E., Berthoz, S., & Frith, U. (2004). Brief report: Cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. *Journal of Autism and Developmental Disorders, 34*(2), 229–235. https://doi.org/10.1023/B:JADD.0000022613.41399.14.

Hobson, P. R. (2011). Autism and the self. In S. Gallagher (Ed.), *The Oxford handbook of the self* (pp. 571–591). Oxford: Oxford University Press.

Hobson, R. P., & Meyer, J. A. (2005). Foundations for self and other: a study in autism. *Developmental Science, 8*(6), 481–491. https://doi.org/10.1111/j.1467-7687.2005.00439.x.

Hobson, H., Westwood, H., Conway, J., McEwen, F., Colvert, E., Catmur, C., et al. (2018). The impact of alexithymia on autism diagnostic assessments.

Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: a developmental lag in attribution of epistemic states. *Child Development, 57*(3), 567–582. https://doi.org/10.2307/1130337.

Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. Psyche*,* 13(2).

Hohwy, J. (2013). *The predictive mind*: Oxford University Press.

Hohwy, J., & Michael, J. (2017). Why should any body have a self? In F. de Vignemont & A. Alsmith (Eds.), *The body and the self, revisited*. Cambridge: MIT Press.

Holmes, N. P., & Spence, C. (2006). Beyond the body schema: visual, prosthetic, and technological contributions to bodily perception and awareness. Human body perception from the inside out, 15–64.

Huang, A. X., Hughes, T. L., Sutton, L. R., Lawrence, M., Chen, X., Ji, Z., et al. (2017). Understanding the self in individuals with autism spectrum disorders (ASD): a review of literature. *Frontiers in Psychology, 8*(1422), https://doi.org/10.3389/fpsyg.2017.01422.

Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: the role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin, 139*(1), 133.

Hurlburt, R. T., Happe, F., & Frith, U. (1994). Sampling the form of inner experience in three adults with Asperger syndrome. *Psychological Medicine, 24*(02), 385–395.

Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., den Ouden, H. E., et al. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron, 80*(2), 519–530.

Imafuku, M., Hakuno, Y., Uchida-Ota, M., Yamamoto, J.-i., & Minagawa, Y. (2014). "Mom called me!" Behavioral and prefrontal responses of infants to self-names spoken by their mothers. *Neuroimage, 103*, 476–484. https://doi.org/10.1016/j.neuroimage.2014.08.034.

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport, 7*(14), 2325–2330.

Izuma, K., Matsumoto, K., Camerer, C. F., & Adolphs, R. (2011). Insensitivity to social reputation in autism. *Proceedings of the National Academy of Sciences, 108*(42), 17302–17307. https://doi.org/10.1073/pnas.1107038108.

Jackson, P., Skirrow, P., & Hare, D. (2012). Asperger through the looking glass: an exploratory study of self-understanding in people with Asperger's syndrome. *Journal of Autism and Developmental Disorders, 42*(5), 697–706. https://doi.org/10.1007/s10803-011-1296-8.

Janssen, A. (2018). Gender dysphoria and autism spectrum disorders. In A. Janssen & S. Leibowitz (Eds.), *Affirmative mental health care for transgender and gender diverse youth: a clinical guide* (pp. 121–128). Cham: Springer International Publishing.

Jokisch, D., Daum, I., & Troje, N. F. (2006). Self recognition versus recognition of others by biological motion: viewpoint-dependent effects. *Perception, 35*(7), 911–920. https://doi.org/10.1068/p5540.

Jordan, R. R. (1989). An experimental comparison of the understanding and use of speaker-addressee personal pronouns in autistic children. *International Journal of Language & Communication Disorders, 24*(2), 169–179. https://doi.org/10.3109/13682828909011954.

Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B, 370*(1668), 20140169.

Kanner, L. (1943). Autistic disturbances of affective contact.

Keenan, J. P., McCutcheon, B., Freund, S., Gallup, G. G., Sanders, G., & Pascual-Leone, A. (1999). Left hand advantage in a self-face recognition task. *Neuropsychologia, 37*(12), 1421–1425.

Keenan, J. P., Nelson, A., O'connor, M., & Pascual-Leone, A. (2001). Neurology: self-recognition and the right hemisphere. *Nature, 409*(6818), 305.

Kinnaird, E., Stewart, C., & Tchanturia, K. (2019). Investigating alexithymia in autism: a systematic review and meta-analysis. *European Psychiatry, 55*, 80–89. https://doi.org/10.1016/j.eurpsy.2018.09.004.

Kiverstein, J. (2018). Free energy and the self: an ecological–enactive interpretation. *TOPOI.* https://doi.org/10.1007/s11245-018-9561-5.

Klein, S. B., Sherman, J. W., & Loftus, J. (1996). The role of episodic and semantic memory in the development of trait self-knowledge. *Social Cognition, 14*(4), 277–291. https://doi.org/10.1521/soco.1996.14.4.277.

Klein, S. B., Chan, R. L., & Loftus, J. (1999). Independence of episodic and semantic self-knowledge: the case from autism. *Social Cognition, 17*(4), 413–436.

Knoblich, G., & Prinz, W. (2001). Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: Human Perception and Performance, 27*(2), 456.

Kristen, S., Rossmann, F., & Sodian, B. (2014). Theory of own mind and autobiographical memory in adults with ASD. *Research in Autism Spectrum Disorders, 8*(7), 827–837.

Lai, M.-C., Lombardo, M. V., Chakrabarti, B., Ruigrok, A. N. V., Bullmore, E. T., Suckling, J., et al. (2018). Neural self-representation in autistic women and association with 'compensatory camouflaging'. *Autism*, 1362361318807159. https://doi.org/10.1177/1362361318807159.

Lane, R. D., Sechrest, L., & Riedel, R. (1998). Sociodemographic correlates of alexithymia. *Comprehensive Psychiatry, 39*(6), 377–385. https://doi.org/10.1016/S0010-440X(98)90051-7.

Lang, B., & Perner, J. (2002). Understanding of intention and false belief and the development of self-control. *British Journal of Developmental Psychology, 20*(1), 67–76. https://doi.org/10.1348/026151002166325.

Lawson, W. B., & Dombroski, B. A. (2015). Might we be calling problems seen in autism spectrum conditions: 'poor theory of mind, 'when actually they are related to non-generalised 'object permanence'? *Journal of Intellectual Disability-Diagnosis and Treatment, 3*(1), 43–48.

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience, 8*.

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience, 20*(9), 1293–1299. https://doi.org/10.1038/nn.4615 http://www.nature.com/neuro/journal/v20/n9/abs/nn.4615.html#supplementary-information.

Lee, A., & Hobson, R. P. (1998). On developing self-concepts: a controlled study of children and adolescents with autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*(8), 1131–1144.

Lee, A., Hobson, R. P., & Chiat, S. (1994). I, you, me, and autism: an experimental study. *Journal of Autism and Developmental Disorders, 24*(2), 155–176. https://doi.org/10.1007/bf02172094.

Leekam, S. R., & Ramsden, C. A. H. (2006). Dyadic orienting and joint attention in preschool children with autism, *Journal of Autism and Developmental Disorders., 36*(2), 185. https://doi.org/10.1007/s10803-005-0054-1.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition, 43*(3), 225–251. https://doi.org/10.1016/0010-0277(92)90013-8.

Letheby, C., & Gerrans, P. (2017). Self unbound: ego dissolution in psychedelic experience. *Neuroscience of Consciousness, 3*(1), nix016–nix016. https://doi.org/10.1093/nc/nix016.

Lewis, M., & Ramsay, D. (2004). Development of self-recognition, personal pronoun use, and pretend play during the 2nd year. *Child Development, 75*(6), 1821–1831. https://doi.org/10.1111/j.1467-8624.2004.00819.x.

Limanowski, J., & Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in Human Neuroscience, 7*, 547. https://doi.org/10.3389/fnhum.2013.00547.

Lind, S. E. (2010). Memory and the self in autism: a review and theoretical framework. *Autism, 14*(5), 430–456. https://doi.org/10.1177/1362361309358700.

Lind, S. E., & Bowler, D. M. (2009a). Delayed self-recognition in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 39*(4), 643–650. https://doi.org/10.1007/s10803-008-0670-7.

Lind, S. E., & Bowler, D. M. (2009b). Recognition memory, self-other source memory, and theory-of-mind in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 39*(9), 1231–1239.

Lombardo, M. V., & Baron-Cohen, S. (2010). Unraveling the paradox of the autistic self. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(3), 393–403.

Lombardo, M. V., Barnes, J. L., Wheelwright, S. J., & Baron-Cohen, S. (2007). Self-referential cognition and empathy in autism. *PLoS One, 2*(9), e883.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Sadek, S. A., Pasco, G., Wheelwright, S. J., et al. (2010). Atypical neural self-representation in autism. *Brain, 133*(2), 611–624. https://doi.org/10.1093/brain/awp306.

Longo, M. R., Azañón, E., & Haggard, P. (2010). More than skin deep: body representation beyond primary somatosensory cortex. *Neuropsychologia, 48*(3), 655–668. https://doi.org/10.1016/j.neuropsychologia.2009.08.022.

Lumley, M. A., Stettner, L., & Wehmer, F. (1996). How are alexithymia and physical illness linked? A review and critique of pathways. *Journal of Psychosomatic Research, 41*(6), 505–518. https://doi.org/10.1016/S0022-3999(96)00222-X.

Lyons, V., & Fitzgerald, M. (2013). Atypical sense of self in autism spectrum disorders: a neuro-cognitive perspective. In *Recent advances in autism spectrum disorders—volume I.*

Manning, C., Kilner, J., Neil, L., Karaminis, T., & Pellicano, E. (2016). Children on the autism spectrum update their behaviour in response to a volatile environment. *Developmental Science*, n/a-n/a, https://doi.org/10.1111/desc.12435.

Maras, K. L., Memon, A., Lambrechts, A., & Bowler, D. M. (2013). Recall of a live and personally experienced eyewitness event by adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*(8), 1798–1810.

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences, 8*(2), 79–86. https://doi.org/10.1016/j.tics.2003.12.008.

Mars, A. E., Mauk, J. E., & Dowrick, P. W. (1998). Symptoms of pervasive developmental disorders as observed in prediagnostic home videos of infants and toddlers. *The Journal of Pediatrics, 132*(3), 500–504. https://doi.org/10.1016/S0022-3476(98)70027-7.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience, 5*, 39. https://doi.org/10.3389/fnhum.2011.00039.

Mattan, B., Quinn, K. A., Apperly, I. A., Sui, J., & Rotshtein, P. (2015). Is it always me first? Effects of self-tagging on third-person perspective-taking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(4), 1100.

Mattan, B. D., Quinn, K. A., Acaster, S. L., Jennings, R. M., & Rotshtein, P. (2017). Prioritization of self-relevant perspectives in ageing. *The Quarterly Journal of Experimental Psychology, 70*(6), 1033–1052.

May, T., Sciberras, E., Brignell, A., & Williams, K. (2017). Autism spectrum disorder: updated prevalence and comparison of two birth cohorts in a nationally representative Australian sample. *BMJ Open,* 7(5), https://doi.org/10.1136/bmjopen-2016-015549.

McGeer, V. (2004). Autistic self-awareness. *Philosophy, Psychiatry, & Psychology, 11*(3), 235–251.

Millward, C., Powell, S., Messer, D., & Jordan, R. (2000). Recall for self and other in autism: children's memory for events experienced by themselves and their peers. *Journal of Autism and Developmental Disorders, 30*(1), 15–28. https://doi.org/10.1023/a:1005455926727.

Milton, D. E. M. (2012). On the ontological status of autism: the 'double empathy problem'. *Disability & Society, 27*(6), 883–887. https://doi.org/10.1080/09687599.2012.710008.

Milton, D. E. M. (2014a). Autistic expertise: a critical reflection on the production of knowledge in autism studies. *Autism, 18*(7), 794–802. https://doi.org/10.1177/1362361314525281.

Milton, D. (2014b). Becoming autistic: an aut-ethnography. *Cutting Edge Psychiatry in Practice, 4*, 185–192.

Mitchell, P., & O'Keefe, K. (2008). Brief report: Do individuals with autism spectrum disorder think they know their own minds? *Journal of Autism and Developmental Disorders, 38*(8), 1591–1597. https://doi.org/10.1007/s10803-007-0530-x.

Mitchell, P., & Ropar, D. (2004). Visuo-spatial abilities in autism: a review. *Infant and Child Development, 13*(3), 185–198. https://doi.org/10.1002/icd.348.

Miyakoshi, M., Kanayama, N., Iidaka, T., & Ohira, H. (2010). EEG evidence of face-specific visual self-representation. *Neuroimage, 50*(4), 1666–1675. https://doi.org/10.1016/j.neuroimage.2010.01.030.

Mizuno, A., Liu, Y., Williams, D. L., Keller, T. A., Minshew, N. J., & Just, M. A. (2011). The neural basis of deictic shifting in linguistic perspective-taking in high-functioning autism. *Brain, 134*(8), 2422–2435. https://doi.org/10.1093/brain/awr151.

Molnar-Szakacs, I., & Uddin, L. Q. (2016). The self in autism. In M. Kyrios, R. Moulding, G. Doron, S. S. Bhar, M. Nedeljkovic, & M. Mikulincer (Eds.), *The self in understanding and treating psychological disorders* (pp. 144–157). Cambridge: Cambridge University Press.

Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and Cognition, 18*(4), 1056–1064. https://doi.org/10.1016/j.concog.2009.05.004.

Moray, N. (1959). Attention in dichotic listening: affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11*(1), 56–60. https://doi.org/10.1080/17470215908416289.

Morita, T., Kosaka, H., Saito, D. N., Ishitobi, M., Munesue, T., Itakura, S., et al. (2012). Emotional responses associated with self-face processing in individuals with autism spectrum disorders: an fMRI study.

Social Neuroscience, 7(3), 223–239. https://doi.org/10.1080/17470919.2011.598945.

Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J., & Friston, K. J. (2014). Bayesian inferences about the self (and others): a review. *Consciousness and Cognition, 25*, 67–76. https://doi.org/10.1016/j.concog.2014.01.009.

Mul, C.-l., Stagg, S. D., Herbelin, B., & Aspell, J. E. (2018). The feeling of me feeling for you: interoception, alexithymia and empathy in autism. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-018-3564-3.

Mul, C. L., Cardini, F., Stagg, S. D., Sadeghi Esfahlani, S., Kiourtsoglou, D., Cardellicchio, P., et al. (2019). Altered bodily self-consciousness and peripersonal space in autism. *Autism*, 1362361319838950. https://doi.org/10.1177/1362361319838950.

Murray, D., Lesser, M., & Lawson, W. (2005). Attention, monotropism and the diagnostic criteria for autism. *Autism, 9*(2), 139–156. https://doi.org/10.1177/1362361305051398.

Nadig, A. S., Ozonoff, S., Young, G. S., Rozga, A., Sigman, M., & Rogers, S. J. (2007). A prospective study of response to name in infants at risk for autism. *Archives of Pediatrics & Adolescent Medicine, 161*(4), 378–383. https://doi.org/10.1001/archpedi.161.4.378.

Neuman, C. J., & Hill, S. D. (1978). Self-recognition and stimulus preference in autistic children. *Developmental Psychobiology, 11*(6), 571–578. https://doi.org/10.1002/dev.420110606.

Nijhof, A. D., Dhar, M., Goris, J., Brass, M., & Wiersema, J. R. (2017). Atypical neural responding to hearing one's own name in adults with ASD. *Journal of Abnormal Psychology*. https://doi.org/10.1037/abn0000329.

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: a review. *Neuroscience and Biobehavioral Reviews, 36*(2), 836–854. https://doi.org/10.1016/j.neubiorev.2011.11.008.

O'Connor, K., & Kirk, I. (2008). Brief report: Atypical social cognition and social behaviours in autism spectrum disorder: a different way of processing rather than an impairment. *Journal of Autism and Developmental Disorders, 38*(10), 1989–1997. https://doi.org/10.1007/s10803-008-0559-5.

Øien, R. A., Cicchetti, D. V., & Nordahl-Hansen, A. (2018). Gender dysphoria, sexuality and autism spectrum disorders: a systematic map review. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-018-3686-7.

Ondobaka, S., Kilner, J., & Friston, K. (2015). The role of interoceptive inference in theory of mind. *Brain and Cognition*.

Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: a study of first birthday home videotapes. *Journal of Autism and Developmental Disorders, 24*(3), 247–257. https://doi.org/10.1007/BF02172225.

Palmer, C. J., Paton, B., Hohwy, J., & Enticott, P. G. (2013). Movement under uncertainty: the effects of the rubber-hand illusion vary along the nonclinical autism spectrum. *Neuropsychologia, 51*(10), 1942–1951.

Palmer, C. J., Paton, B., Kirkovski, M., Enticott, P. G., & Hohwy, J. (2015). Context sensitivity in action decreases along the autism spectrum: a predictive processing perspective. *Proceedings of the Royal Society of London B: Biological Sciences, 282*(1802), 20141557.

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism: towards volatility, action, and behavior. *Psychological Bulletin*.

Pearson, A., Ropar, D., & Hamilton, A. F. d. C. (2013). A review of visual perspective taking in autism spectrum disorder.

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences, 16*(10), 504–510.

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the autistic child's theory of mind: knowledge, belief, and communication. *Child Development*, 689–700.

Perrin, F., Maquet, P., Peigneux, P., Ruby, P., Degueldre, C., Balteau, E., et al. (2005). Neural mechanisms involved in the detection of our first name: a combined ERPs and PET study. *Neuropsychologia, 43*(1), 12–19.

Philippi, C. L., Duff, M. C., Denburg, N. L., Tranel, D., & Rudrauf, D. (2012). Medial PFC damage abolishes the self-reference effect. *Journal of Cognitive Neuroscience, 24*(2), 475–481.

Phillips, W. (1993). *Understanding intention and desire by children with autism*. University of London 1993.

Phillips, W., Baron-Cohen, S., & Rutter, M. (1998). Understanding intention in normal development and in autism. *British Journal of Developmental Psychology, 16*(3), 337–348. https://doi.org/10.1111/j.2044-835X.1998.tb00756.x.

Plaisted, K. C. (2001). Reduced generalization in autism: an alternative to weak central coherence. In J. A. Burack, T. Charman, N. Yirmiya, & P. Zelazo (Eds.), *The development of autism: perspectives from theory and research* (pp. 135–155): Lawrence Erlbaum Associates, Inc.

Platek, S. M., Thomson, J. W., & Gallup, G. G. (2004). Cross-modal self-recognition: the role of visual, auditory, and olfactory primes. *Consciousness and Cognition, 13*(1), 197–210.

Prebble, S. C., Addis, D. R., & Tippett, L. J. (2013). Autobiographical memory and sense of self. *Psychological Bulletin, 139*(4), 815.

Prévost, P., Tuller, L., Zebib, R., Barthez, M. A., Malvy, J., & Bonnet-Brilhault, F. (2018). Pragmatic versus structural difficulties in the production of pronominal clitics in French-speaking children with autism spectrum disorder. *Autism & Developmental Language Impairments, 3*, 2396941518799643.

Quattrocki, E., & Friston, K. (2014). Autism, oxytocin and interoception. *Neuroscience & Biobehavioral Reviews, 47*, 410–430.

Raviv, A., Bar-Tal, D., Raviv, A., & Peleg, D. (1990). Perception of epistemic authorities by children and adolescents. *Journal of Youth and Adolescence, 19*(5), 495–510. https://doi.org/10.1007/BF01537477.

Reddy, V., Williams, E., Costantini, C., & Lan, B. (2010). Engaging with the self: mirror behaviour in autism, Down syndrome and typical development. *Autism, 14*(5), 531–546. https://doi.org/10.1177/1362361310370397.

Reed, T. (2002). Visual perspective taking as a measure of working memory in participants with autism. *Journal of Developmental and Physical Disabilities, 14*(1), 63–76.

Reed, T., & Peterson, C. (1990). A comparative study of autistic subjects' performance at two levels of visual and cognitive perspective taking. *Journal of Autism and Developmental Disorders, 20*(4), 555–567.

Repp, B. H., & Knoblich, G. (2007). Toward a psychophysics of agency: detecting gain and loss of control over auditory action effects. *Journal of Experimental Psychology: Human Perception and Performance, 33*(2), 469.

Robertson, C. E., Ratai, E.-M., & Kanwisher, N. (2016). Reduced GABAergic action in the autistic brain. *Current Biology*. https://doi.org/10.1016/j.cub.2015.11.019.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*(9), 677.

Rohde, M., Di Luca, M., & Ernst, M. O. (2011). The rubber hand illusion: feeling of ownership and proprioceptive drift do not go hand in hand. *PLoS One, 6*(6), e21659. https://doi.org/10.1371/journal.pone.0021659.

Root, N. B., Case, L. K., Burrus, C. J., & Ramachandran, V. (2015). External self-representations improve self-awareness in a child with autism. *Neurocase, 21*(2), 206–210.

Rubenstein, J., & Merzenich, M. M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes, Brain and Behavior, 2*(5), 255–267.

Russell, J., & Hill, E. L. (2001). Action-monitoring and intention reporting in children with autism. *Journal of Child Psychology and Psychiatry, 42*(3), 317–328. https://doi.org/10.1111/1469-7610.00725.

Russell, J., & Jarrold, C. (1999). Memory for actions in children with autism: self versus other. *Cognitive Neuropsychiatry, 4*(4), 303–331. https://doi.org/10.1080/135468099395855.

Russell, J., Hill, E. L., & Franco, F. (2001). The role of belief veracity in understanding intentions-in-action: preschool children's performance on the transparent intentions task. *Cognitive Development, 16*(3), 775–792. https://doi.org/10.1016/S0885-2014(01)00057-0.

Russo, L., Craig, F., Ruggiero, M., Mancuso, C., Galluzzi, R., Lorenzo, A., et al. (2018). Exploring visual perspective taking and body awareness in children with autism spectrum disorder. *Cognitive Neuropsychiatry, 23*(4), 254–265. https://doi.org/10.1080/13546805.2018.1486182.

Saito, N., Takahata, K., Murai, T., & Takahashi, H. (2015). Discrepancy between explicit judgement of agency and implicit feeling of agency: implications for sense of agency and its disorders. *Consciousness and Cognition, 37*, 1–7. https://doi.org/10.1016/j.concog.2015.07.011.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain, 128*(5), 1102–1111. https://doi.org/10.1093/brain/awh464.

Santiesteban, I., Shah, P., White, S., Bird, G., & Heyes, C. (2015). Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control. *Psychonomic Bulletin & Review, 22*(3), 844–849.

Sasson, N. J., Morrison, K. E., Pinkham, A. E., Faso, D. J., & Chmielewski, M. (2018). Brief report: Adults with autism are less accurate at predicting how their personality traits are evaluated by unfamiliar observers. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-018-3487-z.

Saxe, R., Moran, J. M., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Social Cognitive and Affective Neuroscience, 1*(3), 229–234.

Schauder, K. B., Mash, L. E., Bryant, L. K., & Cascio, C. J. (2015). Interoceptive ability and body awareness in autism spectrum disorder. *Journal of Experimental Child Psychology, 131*, 193–200. https://doi.org/10.1016/j.jecp.2014.11.002.

Schechtman, M. (2011). The narrative self. In S. Gallagher (Ed.), *The Oxford handbook of the self*. Oxford: Oxford University Press.

Scheeren, A. M., Begeer, S., Banerjee, R., Meerum Terwogt, M., & Koot, H. M. (2010). Can you tell me something about yourself?: self-presentation in children and adolescents with high functioning autism spectrum disorder in hypothetical and real life situations. *Autism, 14*(5), 457–473. https://doi.org/10.1177/1362361310366568.

Schwarzkopf, S., Schilbach, L., Vogeley, K., & Timmermans, B. (2014). "Making it explicit" makes a difference: evidence for a dissociation of spontaneous and intentional level 1 perspective taking in high-functioning autism. *Cognition, 131*(3), 345–354. https://doi.org/10.1016/j.cognition.2014.02.003.

Sedda, A. (2011). Body integrity identity disorder: from a psychological to a neurological syndrome. *Neuropsychology Review, 21*(4), 334–336. https://doi.org/10.1007/s11065-011-9186-6.

Seth, A. K. (2018). Being a beast machine: the origins of selfhood in control-oriented interoceptive inference. psyarxiv.com/vg5da.

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology, 2*, 395.

Shah, P., Hall, R., Catmur, C., & Bird, G. (2016). Alexithymia, not autism, is associated with impaired interoception. *Cortex, 81*, 215–220. https://doi.org/10.1016/j.cortex.2016.03.021.

Shultz, T. R., Wells, D., & Sarda, M. (1980). Development of the ability to distinguish intended actions from mistakes, reflexes, and passive movements. *British Journal of Social and Clinical Psychology, 19*(4), 301–310. https://doi.org/10.1111/j.2044-8260.1980.tb00357.x.

Silani, G., Bird, G., Brindley, R., Singer, T., Frith, C., & Frith, U. (2008). Levels of emotional awareness and autism: an fMRI study. *Social Neuroscience, 3*(2), 97–112. https://doi.org/10.1080/17470910701577020.

Silberg, J. L. (1978). The development of pronoun usage in the psychotic child. *Journal of Autism and Childhood Schizophrenia, 8*(4), 413–425. https://doi.org/10.1007/BF01538047.

Skewes, J. C., & Gebauer, L. (2016). Brief report: Suboptimal auditory localization in autism spectrum disorder: support for the Bayesian account of sensory symptoms. *Journal of Autism and Developmental Disorders, 46*(7), 2539–2547. https://doi.org/10.1007/s10803-016-2774-9.

Skewes, J. C., Jegindø, E.-M., & Gebauer, L. (2015). Perceptual inference and autistic traits. *Autism, 19*(3), 301–307. https://doi.org/10.1177/1362361313519872.

Skorich, D. P., Gash, T. B., Stalker, K. L., Zheng, L., & Haslam, S. A. (2017). Exploring the cognitive foundations of the shared attention mechanism: evidence for a relationship between self-categorization and shared attention across the autism spectrum. *Journal of Autism and Developmental Disorders*, 1–13, https://doi.org/10.1007/s10803-017-3049-9.

Smith, D., Ropar, D., & Allen, H. A. (2017). The integration of occlusion and disparity information for judging depth in autism spectrum disorder. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-017-3234-x.

Sperduti, M., Pieron, M., Leboyer, M., & Zalla, T. (2014). Altered pre-reflective sense of agency in autism spectrum disorders as revealed by reduced intentional binding. *Journal of Autism and Developmental Disorders, 44*(2), 343–352.

Spiker, D., & Ricks, M. (1984). Visual self-recognition in autistic children: developmental relationships. *Child Development, 55*(1), 214–225. https://doi.org/10.2307/1129846.

Starkweather, C. K., Babayan, B. M., Uchida, N., & Gershman, S. J. (2017). Dopamine reward prediction errors reflect hidden-state inference across time. *Nature Neuroscience, advance online publication*, https://doi.org/10.1038/nn.4520 http://www.nature.com/neuro/journal/vaop/ncurrent/abs/nn.4520.html#supplementary-information.

Strang, J. F., Kenworthy, L., Dominska, A., Sokoloff, J., Kenealy, L. E., Berl, M., et al. (2014). Increased gender variance in autism spectrum disorders and attention deficit hyperactivity disorder. *Archives of Sexual Behavior, 43*(8), 1525–1533. https://doi.org/10.1007/s10508-014-0285-3.

Sugiura, M., Kawashima, R., Nakamura, K., Okada, K., Kato, T., Nakamura, A., et al. (2000). Passive and active recognition of one's own face. *Neuroimage, 11*(1), 36–48. https://doi.org/10.1006/nimg.1999.0519.

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1105.

Summers, J. A., & Craik, F. I. (1994). The effects of subject-performed tasks on the memory performance of verbal autistic children. *Journal of Autism and Developmental Disorders, 24*(6), 773–783.

Surtees, A. D., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development, 83*(2), 452–460.

Surtees, A., Apperly, I., & Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition, 129*(2), 426–438. https://doi.org/10.1016/j.cognition.2013.06.008.

Suzuki, K., Garfinkel, S. N., Critchley, H. D., & Seth, A. K. (2013). Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia, 51*(13), 2909–2917. https://doi.org/10.1016/j.neuropsychologia.2013.08.014.

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: a meta-analysis. *Psychological Bulletin, 121*(3), 371.

Szatmari, P., Georgiades, S., Duku, E., Zwaigenbaum, L., Goldberg, J., & Bennett, T. (2008). Alexithymia in parents of children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 38*(10), 1859–1865. https://doi.org/10.1007/s10803-008-0576-4.

Tacikowski, P., & Nowicka, A. (2010). Allocation of attention to self-name and self-face: an ERP study. *Biological Psychology, 84*(2), 318–324. https://doi.org/10.1016/j.biopsycho.2010.03.009.

Taylor, G. J. (1984). Alexithymia: concept, measurement, and implications for treatment. *The American Journal of Psychiatry, 141*(6), 725–732. https://doi.org/10.1176/ajp.141.6.725.

Taylor, G. J., Michael Bagby, R., & Parker, J. D. A. (1991). The alexithymia construct: a potential paradigm for psychosomatic medicine. *Psychosomatics, 32*(2), 153–164. https://doi.org/10.1016/S0033-3182(91)72086-0.

Thaler, H., Skewes, J. C., Gebauer, L., Christensen, P., Prkachin, K. M., & Jegindø Elmholdt, E.-M. (2017). Typical pain experience but underestimation of others' pain: emotion perception in self and others in autism spectrum disorder. *Autism*, 1362361317701269.

Toichi, M., Kamio, Y., Okada, T., Sakihama, M., Youngstrom, E. A., Findling, R. L., et al. (2002). A lack of self-consciousness in autism. *American Journal of Psychiatry, 159*(8), 1422–1424. https://doi.org/10.1176/appi.ajp.159.8.1422.

Tsakiris, M., & Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of Experimental Psychology. Human Perception & Performance, 31*(1), 80–91. https://doi.org/10.1037/0096-1523.31.1.80.

Uddin, L. Q. (2011). The self in autism: an emerging view from neuroimaging. *Neurocase, 17*(3), 201–208. https://doi.org/10.1080/13554794.2010.509320.

Uddin, L. Q., Kaplan, J. T., Molnar-Szakacs, I., Zaidel, E., & Iacoboni, M. (2005). Self-face recognition activates a frontoparietal "mirror" network in the right hemisphere: an event-related fMRI study. *Neuroimage, 25*(3), 926–935. https://doi.org/10.1016/j.neuroimage.2004.12.018.

Uddin, L. Q., Davies, M. S., Scott, A. A., Zaidel, E., Bookheimer, S. Y., Iacoboni, M., et al. (2008). Neural basis of self and other representation in autism: an fMRI study of self-face recognition. *PLoS One, 3*(10), e3526. https://doi.org/10.1371/journal.pone.0003526.

Van de Cruys, S. (2017). Affective value in the predictive mind. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing*. Frankfurt am Main: MIND Group.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., et al. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychological Review, 121*(4), 649.

van Laarhoven, T., Stekelenburg, J. J., Eussen, M. L., & Vroomen, J. (2019). Electrophysiological alterations in motor-auditory predictive coding in autism spectrum disorder. *Autism Research*.

Vasudeva, S. B., & Hollander, E. (2017). Body dysmorphic disorder in patients with autism spectrum disorder: a reflection of increased local processing and self-focus. *American Journal of Psychiatry, 174*(4), 313–316. https://doi.org/10.1176/appi.ajp.2016.16050559.

Vermaat, L. E. W., van der Miesen, A. I. R., de Vries, A. L. C., Steensma, T. D., Popma, A., Cohen-Kettenis, P. T., et al. (2018). Self-reported autism spectrum disorder symptoms among adults referred to a gender identity clinic. *LGBT Health, 5*(4), 226–233. https://doi.org/10.1089/lgbt.2017.0178.

Vickerstaff, S., Heriot, S., Wong, M., Lopes, A., & Dossetor, D. (2007). Intellectual ability, self-perceived social competence, and depressive symptomatology in children with high-functioning autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 37*(9), 1647–1664.

von der Lühe, T., Manera, V., Barisic, I., Becchio, C., Vogeley, K., & Schilbach, L. (2016). Interpersonal predictive coding, not action perception, is impaired in autism. *Philosophical Transactions of the Royal Society B, 371*(1693), 20150373.

Vossel, S., Mathys, C., Daunizeau, J., Bauer, M., Driver, J., Friston, K. J., et al. (2014). Spatial attention, precision, and Bayesian inference: a study of saccadic response speed. *Cerebral Cortex, 24*(6), 1436–1450.

Warreyn, P., Roeyers, H., Oelbrandt, T., & De Groote, I. (2005). What are you looking at? Joint attention and visual perspective taking in young children with autism spectrum disorder. *Journal of Developmental and Physical Disabilities, 17*(1), 55–73. https://doi.org/10.1007/s10882-005-2201-1.

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: sources of the experience of will. *American Psychologist, 54*(7), 480.

Willey, L. H. (2014). *Pretending to be normal: living with Asperger's syndrome (autism spectrum disorder) expanded edition*: Jessica Kingsley Publishers.

Williams, D. (2009). *Nobody nowhere: the remarkable autobiography of an autistic girl*: Jessica Kingsley Publishers.

Williams, D. (2010). Theory of own mind in autism: Evidence of a specific deficit in self-awareness? *Autism, 14*(5), 474–494. https://doi.org/10.1177/1362361310366314.

Williams, D., & Happé, F. (2009a). Pre-conceptual aspects of self-awareness in autism spectrum disorder: the case of action-monitoring. *Journal of Autism and Developmental Disorders, 39*(2), 251–259.

Williams, D., & Happé, F. (2009b). What did I say? Versus what did I think? Attributing false beliefs to self amongst children with and without autism. *Journal of Autism and Developmental Disorders, 39*(6), 865–873. https://doi.org/10.1007/s10803-009-0695-6.

Williams, D., & Happé, F. (2010a). ecognising 'social' and 'non-social' emotions in self and others: a study of autism. *Autism*.

Williams, D., & Happé, F. (2010b). Representing intentions in self and other: studies of autism and typical development. *Developmental Science, 13*(2), 307–319.

Williams, D., Nicholson, T., & Grainger, C. (2017). The self-reference effect on perception: undiminished in adults with autism and no relation to autism traits. *Autism Research*, n/a-n/a, https://doi.org/10.1002/aur.1891.

Wimmer, H., & Hard, M. (1991). Against the Cartesian view on mind: young children's difficulty with own false beliefs. *British Journal of Developmental Psychology, 9*(1), 125–138. https://doi.org/10.1111/j.2044-835X.1991.tb00866.x.

Wojcik, D., Allen, R., Brown, C., & Souchay, C. (2011). Memory for actions in autism spectrum disorder. *Memory, 19*(6), 549–558.

Woźniak, M., Kourtis, D., & Knoblich, G. (2018). Prioritization of arbitrary faces associated to self: an EEG study. *PLoS One, 13*(1), e0190679.

Yamada, M., Uddin, L. Q., Takahashi, H., Kimura, Y., Takahata, K., Kousa, R., et al. (2013). Superiority illusion arises from resting-state brain networks modulated by dopamine. *Proceedings of the National Academy of Sciences, 110*(11), 4363–4367.

Yamamoto, K., & Masumoto, K. (2018). Brief report: Memory for self-performed actions in adults with autism spectrum disorder: why does memory of self decline in ASD? *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-018-3559-0.

Yang, H., Wang, F., Gu, N., Gao, X., & Zhao, G. (2013). The cognitive advantage for one's own name is not simply familiarity: an eye-tracking study. *Psychonomic Bulletin & Review, 20*(6), 1176–1180. https://doi.org/10.3758/s13423-013-0426-z.

Yoshimura, S., & Toichi, M. (2014). A lack of self-consciousness in Asperger's disorder but not in PDDNOS: implication for the clinical importance of ASD subtypes. *Research in Autism Spectrum*

Disorders, 8(3), 237–243. https://doi.org/10.1016/j.rasd.2013.12.005.

Zahavi, D. (2010). Complexities of self. *Autism, 14*(5), 547–551. https://doi.org/10.1177/1362361310370040.

Zalla, T., & Sperduti, M. (2015). The sense of agency in autism spectrum disorders: a dissociation between prospective and retrospective mechanisms? *Frontiers in Psychology, 6*, 1278. https://doi.org/10.3389/fpsyg.2015.01278.

Zalla, T., Daprati, E., Sav, A.-M., Chaste, P., Nico, D., & Leboyer, M. (2010). Memory for self-performed actions in individuals with Asperger syndrome. *PLoS One, 5*(10), e13370.

Zalla, T., Miele, D., Leboyer, M., & Metcalfe, J. (2015). Metacognition of agency and theory of mind in adults with high functioning autism. *Consciousness and Cognition, 31*, 126–138. https://doi.org/10.1016/j.concog.2014.11.001.

Zamagni, E., Dolcini, C., Gessaroli, E., Santelli, E., & Frassinetti, F. (2011). Scared by you: modulation of bodily-self by emotional body-postures in autism. *Neuropsychology, 25*(2), 270.

Zhao, S., Uono, S., Yoshimura, S., & Toichi, M. (2018). A functional but atypical self: influence of self-relevant processing on the gaze cueing effect in autism spectrum disorder. Autism Research, 0(0). https://doi.org/10.1002/aur.2019.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience, 23*(2–3), 143–152. https://doi.org/10.1016/j.ijdevneu.2004.05.001.

Zwickel, J., White, S. J., Coniston, D., Senju, A., & Frith, U. (2011). Exploring the building blocks of social cognition: spontaneous agency perception and visual perspective taking in autism. *Social Cognitive and Affective Neuroscience, 6*(5), 564–571. https://doi.org/10.1093/scan/nsq088.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

An Update on the Literature

The literature around self-cognition in autism continues to grow. Here, I cover some of the relevant literature that was not covered by the literature review, primarily due to time of publication (first submitted January 2018, published online May 2019). Where applicable for the remainder of the thesis, the associated category from Table 2 is specified in italics to contextualise the research within the schema provided in the published review above.

Wuyun et al. (2020) demonstrated a memory advantage for objects children were told they 'owned' compared to those owned by the experimenter. This effect held for both typically developing and intellectually disabled children, but the results showed no such advantage for autistic children (cross category: *self-recognition>not otherwise categorised* (object ownership) or *self-prioritisation>temporary self-association* or *memory>self-reference effect*). However, in a later experiment, when the autistic children were able to actively choose which objects they 'owned' by placing them in an appropriate basket, they did show the appropriate memory advantage for self-owned objects, similar to children in both the other groups (*memory>memory for own actions*). The authors highlight the special role of action in supporting the memory effect in autism that was not necessary in the other groups. This reiterates the importance of active inference for self-representation in autism, and provides some of the evidence I called for at the end of the active inference section in the published part of this chapter.

Memory for self was one area which did not yield clear results in the review paper. While the Wuyun et al. (2020) study supports differences in autism for passive self-association memory advantages, they suggest no difference when there is some active component in the self-attribution. A few more papers have recently come out on the side of no difference for memory for self in autism. Nijhof, Bird, Catmur, and Shapiro (2020) report findings using self-prioritisation as measured by a shape-label matching task (*self-prioritisation>temporary self-association*) and an attentional blink for own name as well as a self-reference effect paradigm (*memory>self-reference effect*). They found no relation between any of these measures and autism-spectrum quotient (AQ) scores (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). This study is important in that it explicitly studies the relationship between domains of self-cognition and as such it will be discussed in more detail in Chapter 3. Another recent study, Lind, Williams, Nicholson, Grainger, and Carruthers (2019) investigated the self-reference effect across both autistic samples and the general population with a measurement of AQ. In three experiments, they show that autistic participants demonstrate the self-reference effect in memory, and that their performance does not relate to autistic traits or autistic symptom severity. Dinulescu et al. (2020) found that performance on a self-reference effect task was positively correlated with performance on two social cognition tasks, including the reading the mind in the eyes theory of mind task and a task involving identifying the strength and valance of others' emotions. However, accuracy on the self-trait attribution task was not correlated with AQ scores. There seems then, to be increasing evidence that there is an intact self-reference effect in autism (*memory>self-reference effect*), despite the mixed results reported in the published review.

More evidence in the area of self-knowledge is also available. Pfeifer et al. (2013) looked at appraisal of sentences as applying or not applying to the self (without a memory component) in a diagnosed teenaged autistic sample (*self-knowledge>self-concept*). This study showed differences in many brain areas during this task, including absent preferential activation of the ventromedial prefrontal cortex during self-appraisals (compared to appraisals for Harry Potter) for the autistic group. Activity in the middle cingulate cortex showed opposite activation patterns for the self-other contrast between the autistic and neurotypical groups, which also related to measured social skills (autistics with better social skills had brain activation in this region more similar to the neurotypical group). Responses in the anterior insula were also different between groups, however activity in the medial posterior parietal cortex was not. Behaviourally, they found autistic participants selected more negatively valanced sentences for themselves compared others, and compared to self-appraisals in the neurotypical individuals. Overall, this study supports differences in trait attribution for autistic individuals. Further, self-report measures of quality of self-concept were overlooked in the published review (*self-knowledge>self-concept*). For example, Berna et al. (2016) found that Self-Concept Clarity Scores (Campbell et al., 1996) are negatively related to AQ score. This means that participants with high autistic traits report poorer self-concept clarity.

Since the review was published, there have been two new papers studying response to own name using electroencephalography (EEG) in young autistic children.

Thomas et al. (2019) showed that autistic preschoolers show a larger N100 auditory event related potential (ERP) in response to their own name (*self-prioritisation>orienting to own name*) that matches typically developing children and is consistent with findings from typically developing adults. Arslan et al. (2020) compared infants at high and low risk for autism and found that the high risk group showed attenuated activity in frontal electrodes to own name compared to another name, an effect that was not related to language scores obtained from these children. These neural findings support the behavioural differences in children identified in the published review, but identify different ERP components to the reported EEG studies of orienting to own name in autistic adults.

As was also evident in the review paper, a lot of the more recent work on the self in autism has focused on internal states. Huggins, Donnan, Cameron, and Williams (2020) focused on emotional self-awareness and alexithymia (*internal states*), and provide a systematic review of 47 papers. They highlight caution about the variability of definitions of emotional awareness in autism research, and suggest a multidimensional approach including at least identifying emotions, communicating emotions, imagination and externally oriented thinking, interpreting emotions, interoception, recognising emotions in others, and differentiating between own emotions. In the paper presented in this chapter, our subdivision of internal states addresses some of this conceptual confusion, focusing only on self-focused emotional appraisal in ‘alexithymia’ and separating out ‘interoceptive awareness’. However, alexithymia includes identifying, differentiating, interpreting and communicating own emotions in our classification system. A call for further conceptual clarification in this area is echoed by (Trevisan, Mehling, & McPartland, 2020).

In the review paper, I noted that the ages of participants in studies about the self in autism (and likely autism research more generally) tends to have a bimodal distribution. Interestingly, Nicholson, Williams, Carpenter, and Kallitsounaki (2019) conducted two studies on interoception in autism at different ages (*internal states>interoceptive awareness*). They show poorer interoceptive accuracy (cardiac) for autistic children, but not autistic adults (cardiac and respiratory) as compared to neurotypical peers. This suggests differences in the developmental trajectory for autistic interoceptive awareness (in some domains at least) that may match neurotypical performance by adulthood.

Trevisan, Parker, and McPartland (2021) provide an important insight into the subjective reports of differences in interoceptive awareness from autistic individuals in online forums (*internal states>interoceptive awareness*). They identify themes of hypo-sensibility, hyper-sensibility/hypochondria, poor interoceptive accuracy/confusion and alexithymia. As the published review argued, more research involving the first-person perspective in this area is always welcome.

A pictorial summary of results from the literature review, including updates given here is available in Figure 1, below. In summary, we can see that there is much evidence that autistic self-cognition is different from neurotypical self-cognition across many domains (blue), but there are still areas where the evidence is mixed (yellow). Autistic participants

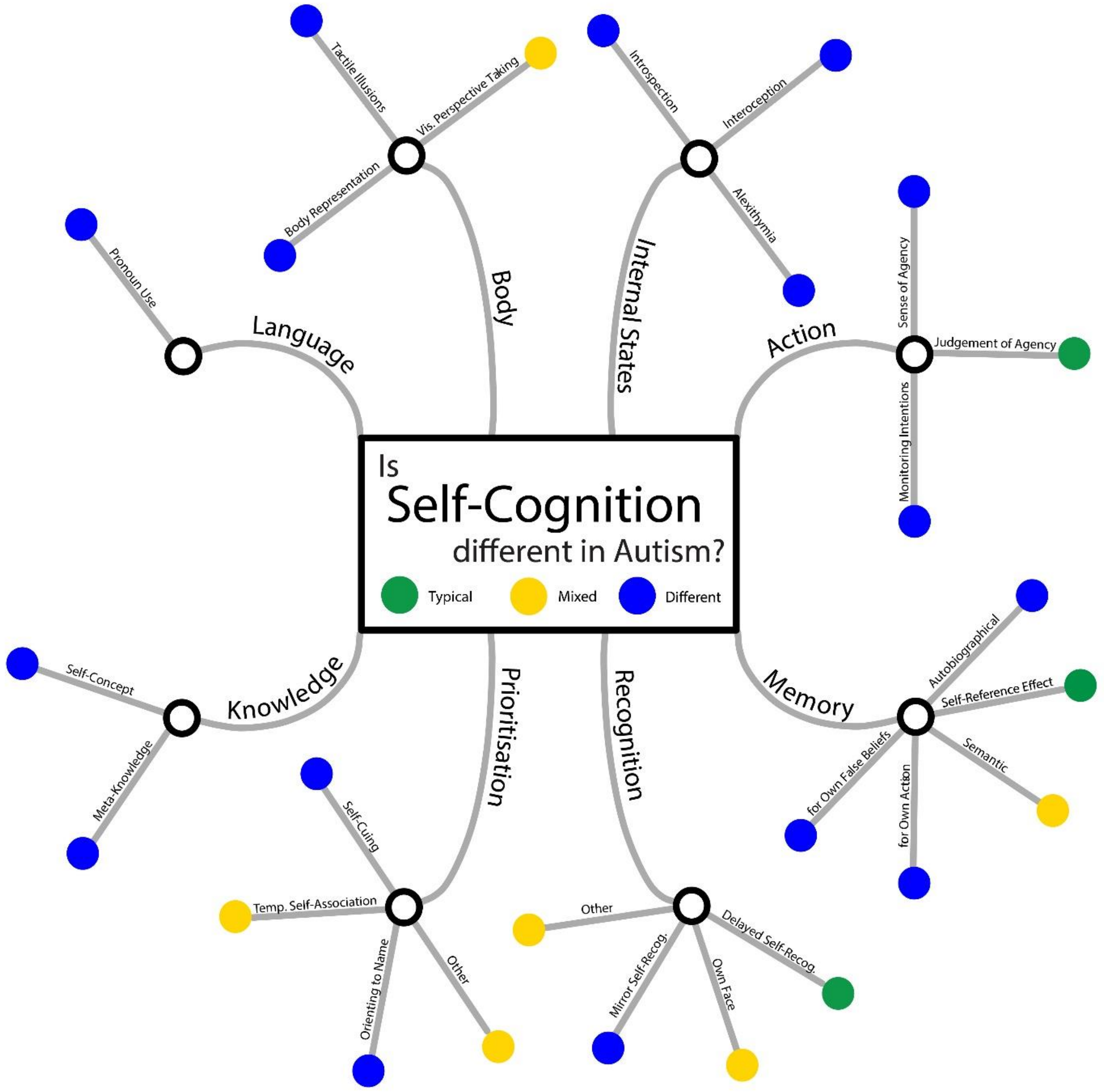


Figure 1 - Chapter 1 Review Pictorial Summary

consistently show no difference from neurotypical samples (green) in very few paradigms – namely delayed self-recognition, the self-reference effect (though some might say the evidence here is still mixed), and judgement of agency. In future chapters, I will focus on judgement of agency in detail.

A Note on Concepts Moving Forward

In the review paper, I describe the landscape of predictive processing accounts of both the self and autism. I pluralise ‘account’ because there is no consensus under predictive processing on the proper conceptual understanding/explanation of either the self, or autism. The treatment of these concepts in this paper set the stage for their use in the remainder of the thesis. As in the discussion of the published review, in the rest of this thesis I take quite a broad conceptualisation of both what the self is, and what the specific predictive processing differences are that characterise autism. None of the following chapters attempts to adjudicate between the weak priors and high inflexible precision of prediction errors accounts of autism for example. Nor is the concept of the self explicitly limited to the self as hidden cause or self as meta-model accounts (though it is more explicitly represented by the system than in the deflationary existence account). I do take the self to be constructed and integrated across many cognitive domains, and to be accessible by both introspection and third-person scientific methods.

The themes identified in this review drive the rest of the research presented in the following chapters. These include the importance of the neural hierarchy, changing environments, regularities at multiple timescales, self as cause, accumulating model evidence, and active inference. In the next chapter, I discuss in detail the core features of autism, and how a predictive processing approach can account for them in a unified way.

Chapter 2. Adaptive Behaviour and Predictive Processing Accounts of Autism

In the first chapter, we saw that previous research supports the hypothesis that self-cognition is different in autism. In this chapter, I temporarily sidestep the self, and discuss the core features of autism. I begin with a history of the diagnostic criteria for autism and motivate a theory-driven approach to understanding the condition. The chapter revolves around a commentary I wrote articulating how a predictive processing account of autism can account for symptoms which are not accounted for by the Social Motivation Theory of autism (Jaswal & Akhtar, 2019; Perrykkad, 2019).

Historically, the terminology and diagnostic criteria for autism have been highly unstable, and there are still very common characteristics of autism that are not included in the diagnostic criteria. One consequence of these changing definitions is that it contributes to fluctuations in the number of people that the label captures. In fact, the apparent and well-popularised increase in prevalence of autism is at least in part due to increasing awareness and changing definitions (Baird, Cass, & Slonims, 2003; Fisch, 2012).

The first descriptions of cases of autism come from Kanner (1943) and Asperger (1944). Since then, the Diagnostic and Statistical Manual (DSM), the most influential source of diagnostic criteria for mental conditions and disorders, has varied significantly in its description and name of this condition. In the DSM-I and DSM-II, autism was subsumed under schizophrenia, as the symptom of being withdrawn (American Psychiatric Association, 1952, 1968). In the DSM-III “Infantile Autism” was first introduced as its own diagnosis, involving lack of responsiveness to others, peculiar or absent speech and bizarre responses to the environment (Spitzer & American Psychiatric Association, 1981). The DSM-III-R introduced the term “Autistic Disorder” which included three lists of diagnostic symptoms which would be largely retained in the DSM-IV and DSM-IV-TR (American Psychiatric Association, 1987, 1995, 2000). These were roughly categorized as social interaction, communication and restricted or repetitive behaviour and interests (with slight variations between editions). Of interest here may be Wolff (2004), who provides a comprehensive though brief account of the history of autism and its diagnosis up until 2004, and Rosen, Lord, and Volkmar (2021) who discuss the evolution of autism diagnoses from the DSM-III through DSM-5.

The latest version of this document, the DSM-5, introduced some of the more major changes in conceptualizing autism (American Psychiatric Association, 2013). The new name, "Autism Spectrum Disorder", is meant to capture the heterogeneity of symptom clusters seen in autism. Note that Asperger's syndrome was a short-lived introduction to the DSM-IV that disappears again by the DSM-5, subsumed under the new heading (American Psychiatric Association, 2013; Solomon, 2017). It is also not until the DSM-5 that "Hyper or Hyporeactivity to sensory input or unusual interests in sensory aspects of the environment" is included in the core features of autism, despite being recognized as pervasive in the relevant population (Ben-Sasson et al., 2009) and even explicitly noted by Kanner (1943). He says,

> Another intrusion comes from loud noises and moving objects… Yet it is not the noise or motion itself that is dreaded… The child himself can happily make as great a noise as any that he dreads and move objects about to his heart's desire…
>
> (Kanner, 1943, p. 245).

In the DSM-5, the structure of the listed symptoms is more equally distributed between the cognitive and social domains than in previous editions. Accordingly, the symptom list is reduced to two categories: 1) "Persistent deficits in social communication and social interaction across multiple contexts" and 2) "Restricted and repetitive patterns of behaviour, interests or activities" (American Psychiatric Association, 2013).

As highlighted by the preceding chapter, some of these features are not well captured by previous theories of autism. Ideally, explanatory theories and definitions of conditions should be tightly linked. However, motivations for changing the definition of psychiatric conditions like autism are complex and socially embedded. Having an official clinical diagnosis is essential in many countries for access to financial and social supports. Pressures to define autism in a way that does not prohibit those in need from access to appropriate services can act in tension with the need for precise concepts that capture underlying natural kinds. In some cases, the solution to this tension is to define conditions differently for clinical practice and research (as Rosen et al. (2021) reports was done for the ICD-10). However, this exacerbates existing issues with application of basic research to aid relevant populations. As such, our theories of autism and our definitions of autism should be tightly linked.

As briefly covered in the last chapter and the thesis preface, many of the prominent theories of autism do not meet this aim, each tending to focus on a few symptoms from the diagnostic criteria rather than the complete set. For example, the theory of mind theory of autism suggests that autism is characterised by an inability to attribute independent mental states to others (Baron-Cohen, 2000; Baron-Cohen et al., 1985), and was one of the first causal theories that afforded specific and falsifiable hypotheses about the nature of autism (Frith & Happé, 1994). To this day, this theory continues to drive a significant proportion of autism research around both the biological basis of autism and more applied research testing supportive interventions (e.g. Andreou and Skrimpa (2020); Lecheler et al. (2020)).

From the outset, however, the theory of mind theory almost exclusively focused on difficulties with social interaction and "failure to develop normal social relationships" as the "pathognomonic symptom" of autism (Baron-Cohen et al., 1985). While my overview above of different definitions of autism shows that issues with communication and social aspects of life are always present, we have also seen that these are only ever a subset of multiple core features of autism. In restricting the focus of developing a theory of autism to just one feature, theories fail to provide a unified account of the condition as a whole. This focus also diminishes the importance of the other consistent features of autism, including particular sensory and behavioural responses to environmental stimuli and restricted and repetitive behaviours.

While less prominent than the theory of mind theory, the social motivation theory of autism is another socially focused theory of autism. It claims that the primary deficit in autism is not a lack of ability but a lack of motivation to engage with the social world (Chevallier et al., 2012). This is hypothesised to be due to differences in neural reward circuitry (Kohls, Chevallier, Troiani, & Schultz, 2012). Jaswal and Akhtar (2019) argue that the social motivation theory is false, and that the symptoms of low levels of eye contact, infrequent pointing, motor stereotypies and echolalia can be explained without reference to a lack of social motivation. Importantly, the authors also rely on autistic testimony to support their argument.

The following commentary is published in Behavioural and Brain Sciences as a response to Jaswal and Akhtar (2019) and argues that many autistic behaviours can be considered adaptive responses to the environment under the predictive processing theory of autism in an explanatorily unified way.

Adaptive behaviour and predictive processing accounts of autism

Kelsey Perrykkad

Cognition & Philosophy Lab, Philosophy Department, School of Philosophical, Historical and International Studies, Monash University, Victoria 3800, Australia.
kelsey.perrykkad@monash.edu
https://cog-phil-lab.org/people/kelsey-perrykkad/

doi:10.1017/S0140525X18002248, e108

Abstract

Many autistic behaviours can rightly be classified as *adaptive*, but why these behaviours differ from adaptive neurotypical behaviours in the same environment requires explanation. I argue that predictive processing accounts best explain why autistic people engage different adaptive responses to the environment and, further, account for evidence left unexplained by the social motivation theory.

If the behaviours described by Jaswal & Akhtar (J&A) are "adaptive responses to a particular situation" (sect. 2.5, para. 2), then the crucial question is this: Why are the adaptive responses to the environment different in autism than in a neurotypical population? Or, if many of these behaviours are used by the neurotypical population, then why is the frequency of their use different in autism? Given the same environment, what is different about autistic individuals that makes their behaviours distinct, yet still adaptive?

In evolutionary ecology, adaptive behaviour consists of responses to the demands of the environment that promote survival and reproductive success. While originally related to phenotypic strategies of whole populations, it has been extended to individual differences (Buss & Greiling 1999; Wilson 1998) and co-opted by clinical psychology to refer to abilities that conform to social expectations for age-appropriate independent living (Coulter & Morrow 1978; cf. Sohn 1976). J&A repeatedly state that characteristic autistic behaviours are adaptive (10 occurrences). This should be taken to mean that the behaviours have cognitive utility (or constitute a cognitive phenotype with evolutionary success; Montague et al. 2012). We should agree with J&A that many distinctively autistic behaviours are adaptive in this way. This observation is, however, best framed in terms of predictive processing theories of autism.

Predictive processing accounts of autism are promising in that they explicitly account for differences in adaptive strategy and thereby are able to address the question I posed for J&A at the outset (Brock 2012; Lawson et al. 2014; 2017; Palmer et al. 2017; Pellicano & Burr 2012; Van de Cruys et al. 2014). Predictive processing is a general and unifying explanation of brain function with growing application to psychiatry (Friston et al. 2014; 2017). These accounts argue that, as the brain seeks to model current and future states of the world, incoming sensory information is weighted differently in autism than in the neurotypical case. Action and perception become tools for inference about the causal origins of sensory inputs, and these theories can thereby explain differences in both domains in autism. The purported difference in general processing in autism generates different responses from neurotypicals because superficially identical environments are mentally represented differently. For example, an adaptive response as an autistic person may be to exploit highly predictable affordances (Constant et al. 2018), whereas for neurotypical individuals, it may be to engage in more exploration. Note that our actions shape our environment, and so this challenges the purported equality of the environments experienced by individuals in these two groups, further giving reason for why the adaptive response to it might differ.

J&A are correct to say that insofar as the social motivation theory is meant to be a unified explanation of autistic cognition and behaviour, it fails to explain all the available evidence (sect. 3 introduction). This includes not just the (very important) firsthand testimony, but also other findings not discussed by J&A. Predictive processing theories account for the tendency for autistic individuals to perceive small elements of the sensed world particularly precisely, therefore accounting for differences in sensitivity to sensory information (Ben-Sasson et al. 2009), as evidenced by superior performance in visual search. Weaker prior expectations for stimulus qualities (Pellicano & Burr 2012), higher sensory precision (Brock 2012; Lawson et al. 2014), or inflexibly high sensitivity to the differences between expectations and outcomes (prediction error; Van de Cruys et al. 2014) are potential specifications of this learning rate difference in autism (Palmer et al. 2017). Increased interest in highly regular domains due to the tendency to construct a prediction-satisfying environment (Constant et al. 2018) may also account for autistic savant skills (Meilleur et al. 2015).

Furthermore, predictive processing accounts of autism offer plausible explanations of the four key pieces of behavioural evidence discussed by J&A.

Predictive processing explains why it may be necessary for autistic people to engage in calming, self-regulatory behaviour in social situations, such as avoiding eye contact. Social situations involve some of the most complicated interacting causes in our environment, and so learning from social stimuli (and thereby participating in successful interaction) requires integrating information over many instances to learn what actions and stimuli might yield the clearest social signal. It is hard to predict another person's behaviour, partly because each social interaction is, in many ways, completely novel, and partly because social interactions are interpreted against a rich tapestry of background information. Reduced eye contact during highly demanding social contexts may be related to decreased precision of social cues (from failing to learn these over many instances), which thereby decreases the ability to reduce uncertainty overall (Palmer et al. 2017). Predictive processing accounts of autism also explain repetitive motor stereotypies as active ways of making incoming sensory information more precise (Palmer et al. 2017).

A similarly complex social action is pointing. One must *learn* to use actions like pointing to reduce uncertainty by controlling and predicting the flow of an interaction based on one's social history. Reduction in pointing may be explained by a weaker understanding of what states in the interlocutor are influenced by the autistic person's actions and how to achieve desired states.

Echolalia too can be understood as an adaptive behaviour in that it reduces prediction error. Oral participation in conversation is made more predictable by reusing heard utterances to communicate similar meanings. This plausibly makes the interlocutor's response more predictable, as the same situation is repeated over multiple events. Predictive processing theories are also compatible with firsthand accounts that social situations are not less appealing, but potentially less accessible to autistic individuals due to the many inferred interacting causes which must be modelled.

Predictive processing accounts of autism suggest that differences in updating mental representations of the self and the environment lead to differences in strategies of inference. This includes perception and action selection which may account for differences in adaptive behaviours between neurotypical individuals and autistic individuals.

References

[The letters "a" and "r" before author's initials stand for target article and response references, respectively]

Ben-Sasson A., Hen L., Fluss R., Cermak S. A., Engel-Yeger B. & Gal E. (2009) A meta-analysis of sensory modulation symptoms in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders* **39**(1):1–11. [KP]

Brock J. (2012) Alternative Bayesian accounts of autistic perception: Comment on Pellicano and Burr. *Autism* **14**:209–24. [KP]

Buss D. M. & Greiling H. (1999) Adaptive individual differences. *Journal of Personality* **67**(2):209–43. [KP]

Constant A., Bervoets J., Hens K. & Cruys S. V. d. (2018) Precise worlds for certain minds: An ecological perspective on the relational self in autism. *Topoi.* doi:10.1007/s11245-018-9546-4. [KP]

Coulter W. A. & Morrow H. W. (1978) *Adaptive behavior: Concepts and measurements*: Grune & Stratton. [KP]

Friston K. J. (2017) Precision psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **2**(8):640–43. doi:10.1016/j.bpsc.2017.08.007. [KP]

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):148–58. Available at: https://doi.org/10.1016/S2215-0366(14)70275-5. [KP]

Lawson R. P., Mathys C. & Rees G. (2017) Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience* **20**:1293–99. doi:10.1038/nn.4615. [KP]

Lawson R. P., Rees G. & Friston K. J. (2014) An aberrant precision account of autism. *Frontiers in Human Neuroscience* **8**:302. doi: 10.3389/fnhum.2014.00302. [KP]

Meilleur A.-A. S., Jelenic P. & Mottron L. (2015) Prevalence of clinically and empirically defined talents and strengths in autism. *Journal of Autism and Developmental Disorders* **45**(5):1354–67. doi:10.1007/s10803-014-2296-2. [KP]

Montague P. R., Dolan R. J., Friston K. J. & Dayan P. (2012) Computational psychiatry. *Trends in Cognitive Sciences* **16**(1):72–80. doi:10.1016/j.tics.2011.11.018. [KP]

Palmer C. J., Lawson R. P. & Hohwy J. (2017) Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological Bulletin* **143**(5):521–42. doi: 10.1037/bul0000097. [KP]

Pellicano E. & Burr D. (2012) When the world becomes 'too real': A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences* **16**(10):504–10. Available at: http://dx.doi.org/ 10.1016/j.tics.2012.08.009. [aVKJ, KP]

Sohn D. (1976) Two concepts of adaptation: Darwin's and psychology's. *Journal of the History of the Behavioral Sciences* **12**(4):367–75. [KP]

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 processing in autism. *Psychological Review* **121**(4):649. [KP]

Wilson D. S. (1998) Adaptive individual differences within single populations. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **353** (1366):199–205. [KP]

In Chapter 1, the predictive processing framework was used to generate hypotheses about self-cognition in autism. The review demonstrated that there are broad differences in self-cognition, which are not just restricted to the interpersonal self, but also present in the bodily and interoceptive domains, for example. In this chapter, we took a closer look at the core features of autism, both historically and under the current diagnostic manual. I also showed that predictive processing can be a useful tool not only for understanding the self in autism, but perhaps autism more generally, as it explains the core features in a unified way.

The general empirical promise of the predictive processing account of autism is also reinforced by findings of a recent systematic review by Cannon, O'Brien, Bungert, and Sinha (2021). While being motivated by predictive processing accounts of autism, this review uses a broad and non-specific definition of ‘predictive processes’ as inclusion criteria. For this paper, neural or behavioural predictive processes are those that are 1) based on a learned association between an antecedent and a consequent, 2) occur in response to the antecedent event, and 3) impact responses to the consequent. Importantly, this characterisation does not rely on the complex hierarchical, prediction error based architecture of the predictive processing framework. This review included 47 original studies in diagnosed autistic participants using paradigms investigating either the process of learning antecedent-consequent pairings and/or responding to the antecedent event in a predictive way. Cannon et al. (2021) report consistent differences in both predictive learning and predictive responses.

The cumulative evidence suggests, therefore, that predictive processing theories of autism are proving fruitful, for both understanding the condition as a unified psychiatric construct, and for understanding how the self in particular might come to be different in autism. Predictive processing accounts are also showing promise in understanding the self across psychiatric contexts, including for borderline personality disorder (Fineberg, Stahl, & Corlett, 2017; Fineberg, Steinfeld, Brewer, & Corlett, 2014), schizophrenia, and drug-related states of consciousness (Corlett, 2017; Letheby & Gerrans, 2017). In the next chapter, I ask whether self-cognitive differences, such as those from Chapter 1, should be added to the list of defining symptoms of autism, and whether these differences are specific to autism, or apply to psychiatric conditions more generally.

Chapter 3. Are differences in self-cognition a characteristic feature of autism? Evidence from psychiatric traits, self-concept and shape-label matching

In Chapter 1, I reviewed evidence from self-cognition across many domains to show that construction and maintenance of the self is likely different in autism. The predictive processing framework shaped my interpretation of the available evidence. In Chapter 2, I showed how this framework can be used to understand core features of autism spectrum conditions. In this chapter, I focus in on these core features and take an empirical approach to the conceptual clarification of ‘autism’.

This chapter is the first experimental chapter in the thesis. In the manuscript that follows, I set aside predictive processing and ask whether cross-domain differences in self-cognition in autism should be added to the diagnostic criteria for autism. To answer this question, I use two self-concept questionnaires (Chapter 1: *self-knowledge>self-concept*) and a self-prioritisation task involving temporary self-association (Chapter 1: *self-prioritisation>temporary self-association*). By using data from both of these very cognitively different self-domains, the research method reflects the domain-general motivations from predictive processing as described in earlier chapters without deploying a predictive processing framework.

Diagnostic criteria for psychiatric conditions play multiple functional roles, and the interests of the involved parties (diagnosed individuals, clinicians, researchers, the public etc.) are often in conflict. As such, it can be difficult to agree on thresholds or principles for inclusion of a feature of a condition to its diagnostic criteria. Here, I took existing diagnostic criteria that *do* include self-cognitive features as exemplars which autism would have to emulate in order to justify including similar features to its diagnostic criteria. Whether or not this approach falls prey to existing problems with symptom-based definitions of psychiatric disorders is covered by the discussion.

Are differences in self-cognition a characteristic feature of autism? Evidence from psychiatric traits, self-concept and shape-label matching

Kelsey Perrykkad[1] and Jakob Hohwy[1]

1. Cognition and Philosophy Lab, Philosophy Department, School of Philosophical, Historical and International Studies, Monash University

ORCID:

Perrykkad: 0000-0001-5876-9136

Hohwy: 0000-0003-3906-3060

Abstract

Growing evidence suggests that a diagnosis of autism spectrum disorder is associated with differences in self-cognition in many domains. The question thus arises whether these differences warrant including self-cognition in the diagnostic criteria for autism. It is also unclear whether measures of implicit and explicit self-cognition are associated within individuals. This online study aims to answer both of these questions. Data was collected from 328 participants from the general population measuring psychiatric traits for autism, in addition to two psychiatric conditions which are partly defined in terms of self-cognitive features (borderline personality disorder and schizophrenia) and two conditions that are not (depression and anxiety). Further, participants completed two self-concept questionnaires and a shape-label matching task to measure self-cognition across the cortical hierarchy. Unexpectedly, results suggest that while autistic traits are significantly correlated with explicit self-concept, this relationship is weaker than between explicit self-concept and most of the other psychiatric traits, including for depression. Further, the task-based implicit self-cognition measures were not significantly correlated with the explicit self-concept measures nor any of the psychiatric trait measures. While these results support previous findings that autism traits are related to self-cognition differences, they suggest that the strength of these differences do not distinguish autism from conditions that are not defined in terms of self-cognition. This may also imply that self-cognition serves as a transdiagnostic dimension of clinical relevance.

Keywords: autism spectrum condition, self-concept, self-cognition, shape-label matching, psychiatric traits

How we build and maintain representations of ourself – our own body, dispositions, name, history etc – requires cognitive resources from across sensory domains and different levels of the cognitive hierarchy. Previous research suggests that autism spectrum condition (autism) is associated with differences in many aspects of self-representation (Frith & Happé, 1999; Hobson, 2011; Huang et al., 2017; Lombardo & Baron‐Cohen, 2010; Lyons & Fitzgerald, 2013; Molnar-Szakacs & Uddin, 2016; Perrykkad & Hohwy, 2020; Uddin, 2011; Williams, 2010). Most of these reviews suggest that the reason for this is homologous with the characteristic social difficulties in autism, but in our recent review we suggest they may stem from differences in distinctively autistic, domain general, cognitive architectures (Perrykkad & Hohwy, 2020).

Differences in self-representation in autism appear to span the cortical hierarchy. There is evidence that low-level self-representation involving early attentional and sensory processes are affected. These are *implicit* processes – not necessarily consciously accessible to the individual. Differences at these early stages of neural processing include diminished behavioural and neural responses to one's own name in autistic children and adults, which are present in neurotypical samples (Cygan, Tacikowski, Ostaszewski, Chojnicka, & Nowicka, 2014; Leekam & Ramsden, 2006; Mars, Mauk, & Dowrick, 1998; Nadig et al., 2007; Nijhof, Dhar, Goris, Brass, & Wiersema, 2018; Osterling & Dawson, 1994; Zwaigenbaum et al., 2005). Two studies have investigated the source of this effect using neural measurements. Nijhof et al. (2018) reported reduced activation in the right temporo-parietal junction and increased activation in the right inferior frontal gyrus in autistic participants when hearing their own name compared to that of others'. The authors interpreted this finding as evidence of differences in self-referential attentional mechanisms in autism compared to neurotypical participants. Findings by Cygan et al. (2014) for visual name and face processing using electroencephalography also suggest attention allocation to self in autism is not different to close others (parent, sibling, grandparent, best friend) unlike in their neurotypical sample.

However, Cygan et al. (2014) additionally propose that since the relevant neural signal was a later one (P300), an alternative explanation is that autistic participants had similar levels of higher-order person specific knowledge for self and close other, or “a poorly developed or even absent ‘I-concept’” (Cygan et al., 2014, p. 11). Developing this kind of higher-order knowledge is an *explicit* self-cognitive process, in that it relates to the conscious, self-reflective aspects of the self. While self-concept would be generally considered to be on the higher end of the cortical hierarchy, involving integrating information across many modalities over long periods of time, it too, appears to be different in autism (Perrykkad & Hohwy, 2020). For example, in contrast to neurotypical self-report, autistic participants have claimed that their own self-knowledge is not as accurate as the perception others have of them (Dritschel, Wisely, Goddard, Robinson, & Howlin, 2010). So, differences in autistic self-representation may involve implicit attentional mechanisms and explicit conceptual representations.

It is unclear whether self-representations at different levels of the cortical hierarchy or across cognitive domains and modalities generally have similar qualities. Recent evidence from Nijhof, Bird, Catmur, and Shapiro (2020) shows that the magnitude of two low level measures of self-prioritisation, namely reduced attentional blink for own name and increased association between arbitrarily-paired self-labels and shapes, were not correlated within individuals. Comparing within explicit self-representations, Nowicka, Wójcik, Kotlewska, Bola, and Nowicka (2018) found that participants with high self-esteem showed greater neural self-preference when evaluating traits for self-attribution. A third study by Krol, Thériault, Olson, Raz, and Bartz (2019) investigated the relationship between implicit and explicit self-representations by comparing self-concept clarity and illusory experiences of the body (rubber hand illusion and body swap illusion) within individuals. This study demonstrated that participants with poorer self-concept were more susceptible to the body-swap illusion and were more likely to feel that the rubber hand was theirs in the asynchronous condition (contrary to classic patterns) (Krol et al., 2019). As such, a less well established self-concept may be associated with a less stable sense of the bodily self. Taken together, these three studies raise questions about how self-representational processes both within and across levels of the cognitive hierarchy relate to one another.

In this study, we aim to answer two research questions. First, whether self-representations across the cortical hierarchy are related within individuals; specifically the quality of relatively low-level attentional self-prioritisation (implicit self-representation) and relatively high-level self-concept (explicit self-representation). Second, whether these measures of self-cognition relate to autism strongly enough to motivate the inclusion of self-cognitive features to its diagnostic criteria. Psychiatric conditions already in this category include borderline personality disorder (BPD) and schizophrenia, which both involve the self in their characteristic features as defined in the ICD-11 (see Table 1) (World Health Organisation, 2018). Conversely, conditions such as depression and anxiety do not involve self-cognition in their ICD or DSM characterisations (American Psychiatric Association, 2013; World Health Organisation, 2018). However, it should be noted that there is some evidence that issues with self-cognition, such as identity disturbances, are transdiagnostically relevant, including for non-self-defined conditions like depression and anxiety (Neacsiu, Herr, Fang, Rodriguez, & Rosenthal, 2015).

Previous research using some of the same measures used here support the idea that differences in explicit self-representation are associated with psychiatric diagnoses and their traits more broadly. Poorer explicit self-representation as measured by the Self-Concept Clarity Scale (SCCS) has been established in individuals with schizophrenia (Cicero, Martin, Becker, & Kerns, 2016) and BPD (Roepke et al., 2011). Lower SCCS scores have also been associated with more depressive symptoms as reported by the Beck Depression Index (BDI) short (Wong, Dirghangi, & Hart, 2019), and BDI-II (Chiu, Chang, & Hui, 2017); anxiety symptoms as measured by the Beck Anxiety Index (BAI) (Chiu et al., 2017); and autism traits measured by the autism spectrum quotient (AQ) (Berna et al., 2016). Lower scores on a different measure of self-concept, the Self-Concept and Identity Measure (SCIM), have also been associated with BPD, depression (Kaufman, Cundiff, & Crowell, 2015; Kaufman, Puzia, Crowell, & Price, 2019) and the depression-anxiety scale (Vanden Poel & Hermans, 2019).

For the measure of self-prioritisation, we chose a task originally reported by Sui, He, and Humphreys (2012). In this task, participants must respond to a presented shape and label and decide whether they match a learned mapping. Participants are faster and more discriminant in response to shapes learned as representing oneself. Previous research investigating associations between performance on this shape-label matching task and the psychiatric traits of interest here shows differences depending on mood inductions related to depression and anxiety (Qian, Wang, Li, & Gao, 2020; Sui, Ohrling, & Humphreys, 2016), but no relationship with autism diagnosis or AQ score (Nijhof et al., 2020; Williams, Nicholson, & Grainger, 2017). In a similar associative learning task involving self-labels, Zhao, Uono, Yoshimura, and Toichi (2018) found that similarly to typically developing participants, responses were faster to self-relevant stimuli than other stimuli. Interestingly, the authors argued that this had a different effect on the attentional system in autism based on differences in a cueing manipulation that were sensitive to autistic symptom severity. As such, there reason to think that behaviour in this implicit task will be associated with the some of the psychiatric conditions of interest.

We have established that differences exist in both implicit and explicit self-representational domains for a range of conditions, including those in both our self-defined and non-self-defined categories. Therefore, to properly answer our second research question of whether it is reasonable to include these characteristics in the diagnostic criteria, we directly compare whether the strength of the relationship between autism and self differences more closely mirrors conditions for which self differences are seen as a characteristic feature, or conditions for which they are not. A clean way to address both of these research questions simultaneously is in a within-subjects design looking at transdiagnostic psychiatric traits, implicit, and explicit self measures in the general population.

For our first research question, if self-prioritisation and self-concept are built on the same domain general architecture in the general population, we would expect quality of self-concept as measured by the SCCS and SCIM to be correlated with implicit self-prioritisation measures from the shape-label matching task. If the explicit and implicit self measures are not correlated, this might suggest that self-cognition is not seamlessly integrated across low and high processing levels.

For the second research question, based on Perrykkad and Hohwy (2020) and the findings reported above, we expected autistic traits to correlate with explicit self measures. While evidence from attentional processes for own name suggest that autism traits should also correlate with implicit self measures, previous findings using the shape-label matching task suggest otherwise, so a correlation between autistic traits and implicit self measures is less expected. The final answer as to whether autism should be characterised by self-cognitive features will lie in direct comparisons of correlations between the self measures and psychiatric traits across conditions. If autistic traits correlate with self measures to a similar degree as BPD and schizophrenia traits correlate with self measures, and more than the self measures correlate with depression and anxiety severity scores, then that would provide good evidence that self-cognitive differences should be added to the diagnostic criteria of autism. On the other hand, if the strength of the relationship between autistic traits and self measures is more similar to the non-self-defined conditions (ie. depression and anxiety), then this would suggest there is less justification for this kind of change.

Table 1 -- Self-Disorder Classifications and ICD-11

Psychiatric Condition	**Classification for this Study**	**Relevant ICD-11 Description Excerpt (World Health Organisation, 2018)**
Personality Disorder: Borderline Pattern (*Borderline Personality Disorder*)	Characterized by Self-Disturbances	*"Personality disorder is characterised by problems in functioning of aspects of the self (e.g., identity, self-worth, accuracy of self-view, self-direction), and/or interpersonal dysfunction…" … "The borderline pattern descriptor may be applied to individuals whose pattern of personality disturbance is characterised by a pervasive pattern of instability of interpersonal relationships, self-image… identity disturbance, manifested in markedly and persistently unstable self-image or sense of self;…"*
Schizophrenia	Characterized by Self-Disturbances	*"Schizophrenia is characterized by disturbances in multiple mental modalties, including… self-experience (e.g., the experience that one's feelings, impulses, thoughts or behaviour are under the control of an external force)…"*
Depressive Disorders (*Depression*)	**Not** Characterized by Self-Disturbances	*"Depressive disorders are characterised by depressive mood (e.g., sad, irritable, empty) or loss of pleasure accompanied by other cognitive, behavioural, or neurovegetative symptoms that significantly affect the individual's ability to function."*
Anxiety	**Not** Characterized by Self-Disturbances	*"Apprehensiveness or anticipation of future danger or misfortune accompanied by a feeling of worry, distress, or somatic symptoms of tension. The focus of anticipated danger may be internal or external."*
Autism Spectrum Disorder (*Autism*)	**Not** Characterized by Self-Disturbances	*"Autism spectrum disorder is characterised by persistent deficits in the ability to initiate and to sustain reciprocal social interaction and social communication, and by a range of restricted, repetitive, and inflexible patterns of behaviour, interests or activities that are clearly atypical or excessive for the individual's age and sociocultural context."*

Methods

This study was approved by Monash University Human Research Ethics Committee (Project Number 23583) and was conducted in accordance with the relevant guidelines and regulations. All participants gave informed consent upon commencing the protocol.

Participants

A total of 328 participants successfully completed the study posted on Amazon Mechanical Turk using the Cloud Research platform (formerly TurkPrime (Litman, Robinson, & Abberbock, 2017)), with an overall completion rate of 70% (30% accepted but did not complete the posting) and a bounce rate of 9% (decided not to complete the study after viewing the description). Data was collected between June 26 and July 30, 2020. Participants were paid $9 USD for completing the task, which took an average of 64 minutes to complete (including consent process and self-timed breaks to a maximum of 180 min total task duration). A total of 40 participants were excluded for the following reasons: uncorrected issue with vision (n=2), previous head injury which resulted in temporary unconsciousness (n=2), more than one missed manipulation check (>10%, n=4)(Oppenheimer, Meyvis, & Davidenko, 2009), performance on self-prioritisation task which was more than two standard deviations below mean (ie. <31% overall accuracy, n=9), more than 50% of self-prioritisation task trials removed or an overall mean greater than two standard deviations above the average for reaction time on the shape-label matching task (details below, n=31). Participant demographic information for the final dataset from 288 participants is available in Table 2.

Procedure

Psychiatric traits for the five conditions were measured using the Autism-Spectrum Quotient (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), Borderline Personality

Questionnaire (Poreh et al., 2006), Schizotypal Personality Questionnaire (Raine, 1991), Beck Anxiety Inventory (Beck, Epstein, Brown, & Steer, 1988) and Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). Explicit self-concept was measured by two questionnaires, the Self-Concept and Identity Measure (SCIM) (Kaufman et al., 2015) and the Self-Concept Clarity Scale (SCCS) (Campbell et al., 1996). Implicit self-prioritisation was measured using a label-shape matching task (Sui et al., 2012). The demographic information, AQ, BPQ and SPQ were completed in that order before the self-prioritisation task. The SCCS, SCIM, BAI and BDI were completed following the task. This order did not change across participants.

Psychiatric Trait Survey Measures

Autism-Spectrum Quotient (AQ)

The AQ is a 50-item questionnaire measuring autistic traits in the general population. Items are rated on a four-point Likert scale and scored on a two-point scale (all responses of the same valence are collapsed for scoring). The questionnaire covers the autistic features related to social skills, attention switching, attention to detail, communication and imagination (Baron-Cohen et al., 2001). While we acknowledge that some uses of this scale in making conclusions about core features of autism have recently been criticised (Perrykkad & Hohwy, 2019; Ridley, 2019), it remains the most widely used for measuring autistic traits in a general population. Cronbach's alpha for the internal consistency of AQ in our sample was 0.80 (good).

Borderline Personality Questionnaire (BPQ)

The BPQ is an 80-item questionnaire measuring borderline personality traits as defined by the DSM-IV criteria. Items cover features of borderline personality disorder including impulsivity, affective instability, abandonment, relationships, self image, suicide or selfmutilation, emptiness, intense anger and quasi-psychotic states (Poreh et al., 2006). Participant responses consist of true/false judgements. Cronbach's alpha for the internal consistency of the sum score for BPQ in our sample was 0.94 (excellent).

Schizotypal Personality Questionnaire (SPQ)

The SPQ is a 74-item questionnaire measuring schizotypy based on the DSM-III-R criteria for Schizotypal Personality Disorder. Features of schizophrenia which are covered by the questionnaire include ideas of reference, social anxiety, odd beliefs and magical thinking, unusual perceptual experiences, eccentric or odd behaviour and appearance, no close friends, odd speech, constricted affect and suspiciousness or paranoid ideation (Raine, 1991). Participants respond to each item with Yes or No. Cronbach's alpha for the internal consistency of the sum score for SPQ in our sample was 0.95 (excellent).

Beck Anxiety Inventory (BAI)

The BAI is a 21-item questionnaire measuring the recent presence and severity of anxiety symptoms. Items cover symptoms such as fear, inability to relax, numbness, sweating, dizziness, and heart-racing (Beck et al., 1988). Participants report how often they have been bothered by each symptom of anxiety in the last month on a four-point likert scale ("Not at all", "Mildly…", "Moderately…", "Severely…"). Cronbach's alpha for the internal consistency of BAI in our sample was 0.92 (excellent).

Beck Depression Inventory (BDI)

The BDI is a 21-item questionnaire measuring the recent presence and severity of depressive symptoms. The items address features including sadness, pessimism about the future, sense of failure, lack of satisfaction/pleasure, guilt, sense of punishment, self-hatred, self-blame, suicidal thoughts, crying, irritability, social interest, indecision, body image, work, sleep disturbance, fatigue, appetite, weight loss, health concerns and libido (Beck et al., 1961). Participants choose one of four options for each item with increasing severity of descriptions for depression symptoms with reference to the last few weeks. Cronbach's alpha for the internal consistency of BDI in our sample was 0.94 (excellent).

Explicit Self-Concept Survey Measures

Self-Concept and Identity Measure (SCIM)

The SCIM is a 27-item questionnaire measuring dimensions of healthy and disturbed identity as understood as a core component of personality pathology in the DSM-5. Items are rated on a seven-point Likert scale ("Strongly Agree"… "Neither agree nor disagree"… "Strongly Disagree"). Higher scores are indicative of "greater identity disturbance" (Kaufman et al., 2015). Cronbach's alpha for the internal consistency of total scores on the SCIM in our sample was 0.93 (excellent). In its initial development, the SCIM was found to have a three factor structure, and can be broken down into measures of Disturbed Identity (11-items), Consolidated Identity (11-items) and Lack of Identity (6-items). A confirmatory factor analysis of this structure did not yield strong evidence for this factor structure in our sample (CFI = 0.81, RMSEA = 0.10), however, Cronbach's alpha was good-excellent within each factor (Disturbed Identity: $\alpha = 0.88$, Consolidated Identity: $\alpha = 0.85$, Lack of Identity: $\alpha = 0.91$). Only the full score was used in subsequent analysis.

Self-Concept Clarity Scale (SCCS)

The SCCS is a 12-item questionnaire measuring structural aspects of self-concept. Each item is rated on a five-point Likert scale ("Strongly Agree"… "Neutral"… "Strongly Disagree"). Higher scores are related to increased clarity of self-concept, including temporal stability, certainty and perceived internal consistency of beliefs about oneself. Low scores were independently associated with chronic self-analysis, lower internal state awareness and ruminative self-focused attention (Campbell et al., 1996). Cronbach's alpha for the internal consistency of SCCS in our sample was 0.87 (good).

It should be highlighted that high scores on the SCCS and low scores on the SCIM relate to better quality self-representation, and thus are expected to be anti-correlated.

Self-Prioritisation Task

Implicit, perceptual self-prioritisation was measured using a shape-label matching task (Sui et al., 2012). The self-prioritisation task was run using Inquisit Web (2020, Retrieved from: https://www.millisecond.com). In this task, participants were presented with three pairings of a shape and label. The labels used were "self", "friend" and "stranger" which were paired with a circle, triangle and square (mappings counterbalanced across participants). In each trial, participants are asked whether a briefly presented (100ms) shape and label matched. Trials falling into the six possible pairings (self-match, self-mismatch, friend-match, friend-mismatch, stranger-match, stranger-mismatch) were equally probable and randomly ordered within three blocks of 120 trials for a total of 360. All stimuli were white and presented on a grey background following a 500ms central fixation cross. After each trial, participants receive feedback as to whether they were correct and further percent accurate feedback was given at the end of each block. The response was speeded, and if participants were too slow (random window of 800-1200ms), a warning appeared following that trial.

From performance on this task, two measures of implicit self-representation were computed. The first measure was based on reaction time, which is the time in milliseconds from the offset of the stimuli until the response. The second measure is based on signal detection sensitivity, or d'. This measure combined matching and nonmatching shape trials to give an unbiased measure of the separation between distributions between signal and noise in units of standard-deviation for the signal distribution. For our analysis, we focus on self-advantage measures (self trials minus the average of friend and stranger trials) for both d' and reaction time. This gives us a measure of self-prioritisation, called *self-advantage*, in both average reaction time and sensitivity for each individual.

Statistical Analysis

Where possible, statistical analyses are reported in both traditional null-hypothesis significance testing (NHST) and Bayesian form using JASP v0.9.0.1 (Marsman & Wagenmakers, 2017) through Jamovi v1.1.9.0 (The Jamovi Project, 2019), and R v3.6.3 software (R Core Team, 2018). Bayes factor interpretation follows Jeffreys (1998).

To answer the first research question about the relationship between implicit and explicit self-representation, we conduct Pearson's correlations between the two explicit questionnaires and the two self-advantage measures. For the matching Bayesian correlations we use a stretched beta prior of width one. The NHST correlations are Bonferroni corrected for the six pairwise comparisons both within and between explicit and implicit self measures.

We answer the second research question in three stages. This question addresses whether the strength of the relationship between autism traits and the self measures is more similar to self-defined or non-self-defined conditions. The first step is to look at the relationships between the traits scores and the self measures in their simplest form - using pairwise correlations as we did for the first question. Significance of the NHST correlations are Bonferroni corrected for the 20 pairwise comparisons.

The second step compares the strength of relationships between self measures and psychiatric traits indirectly, by quantifying variance in the trait score which is explained by the self measures. To do this, we perform a multiple linear regression with each psychiatric trait score as the predicted variable and each of the four self measures as predictors. Bayesian regressions are run with a JZS prior with r scale of 0.354, uniform model priors and use the BAS sampling method, all of which are default in Jamovi. Comparing the R-squared values across these models further illuminates whether the cluster of self measures as a whole best predicted the traits for conditions defined by self differences.

The final step is to directly statistically test whether the strength of the relationship between AQ and the self measures is significantly different from the other conditions. There was no known Bayesian method for this part of the analysis. To minimise the number of comparisons, we perform principal component analysis (PCA) with varimax rotation on the collection of self measures (total scores for each measure as above) as a dimensionality-reduction technique. The top components are selected at an eigenvalue threshold of one. From the loadings variables based on each selected component are created for each participant for further analysis using weighted sum scores (DiStefano, Zhu, & Mindrila, 2009). The cocor toolbox (Diedenhofen & Musch, 2015) is used to directly statistically compare correlations between the psychiatric trait scores and each of the components defined by the PCA analysis, accounting for the dependent (same participants) and overlapping (a shared variable in each comparison) features of the data using ten NSHT methods. As this only allows for pairwise comparisons of correlations, the significance threshold for this family of results is Bonferroni corrected for eight comparisons. In the vast majority of cases for our data all ten methods agree, and so we report only Pearson and Filon’s z statistic in the reported results. Rare disagreements between methods are also noted, and the more conservative outcome favoured in our interpretation.

Finally, as a control to check that the self measures distinguish between conditions that are defined by self differences and those that are not, we use the same method to compare correlations between our self-defined conditions and non-self-defined conditions across both components. The significance threshold for this family of results was also Bonferroni corrected for eight comparisons.

Table 2 – General Demographic Information

Demographic	Category	N	% (Total N = 288)
Gender	Male	155	54.2
	Female	128	44.4
	Other	5	1.7
Age	18-24	46	16.0
	25-31	79	27.4
	32-38	95	33.0
	39-45	55	19.1
	46-50	25	8.7
Country of Residence	USA	283	98.3
	Canada	5	1.7
First Language	English	276	95.8
	Other – Fluent in English	12	4.2
Highest Completed Education	Highschool or equivalent including Vocational Training	71	24.7
	Bachelors, Honours or Associate Degree	164	56.9
	Masters or Doctorate	53	18.4
Employment Status	Unemployed or Not Working	36	12.5
	Student or Intern	19	6.6
	Employed	233	80.9
Official Diagnoses	Autism Spectrum Disorder/Autism/Autistic Disorder/Aspergers' Syndrome/Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS)	3	1.0
	Borderline Personality Disorder	2	0.7
	Schizophrenia	0	0
	Depression	43	14.9
	Anxiety	49	17.0
	Attention-Deficit/Hyperactivity Disorder	3	1.0
	Bipolar Disorder	2	0.7
	Obsessive Compulsive Disorder	1	0.3
	Posttraumatic Stress Disorder	1	0.3
	None	224	77.8

Results

Descriptive statistics for all measures are available in Table 3. Data used for statistical analysis presented below is freely available on Figshare (DOI: 10.26180/14214464).

For the shape-label matching task, while not the focus of the study here, we replicate Sui et al. (2012) insofar as d' is greater for self than friend which is greater than stranger stimuli ($F(2,524) = 303.78$, $p_{greenhouse\text{-}geisser} = 4.23x10^{-83}$), and for congruent trials, participants respond faster ($F(2,478) = 430.79$, $p_{greenhouse\text{-}geisser} = 2.35x10^{-96}$) and more accurately ($F(2,532) = 295.02$, $p_{greenhouse\text{-}geisser} = 1.39x10^{-82}$) for self than others and for friend than stranger.

Table 3 – Descriptive Statistics Summary

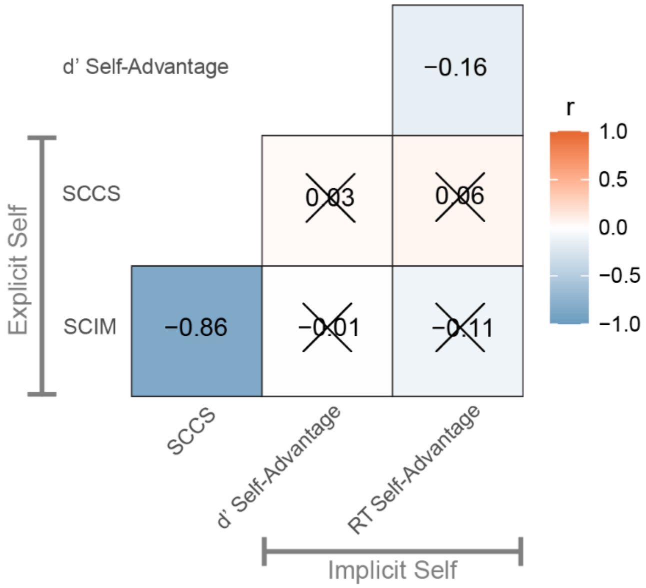
Questionnaire	Mean	Range	1st Qu.	3rd Qu.
Autism-Spectrum Quotient	21.3	4:38	16.0	26.0
Borderline Personality Questionnaire	19.3	0:64	6.0	28.0
Schizotypal Personality Questionnaire	21.4	0:74	10.0	30.3
Beck Depression Inventory	9.6	0:45	2.0	15.0
Beck Anxiety Inventory	8.2	0:40	2.0	12.0
Self-Concept Clarity Scale	43.5	15:60	35.0	52.0
Self-Concept and Identity Measure	68.2	27:145	50.8	81.3
Self-Prioritisation Task *(Overall Accuracy - %)*	76.1	41.7:97.2	66.3	86.7
Self-Prioritisation Task *(d' Self-Advantage)*	0.76	-0.81:3.1	0.40	1.1
Self-Prioritisation Task *(Reaction Time (ms) Self-Advantage)*	245.6	129.5:408.2	215.1	274.2

Explicit and Implicit Self-Representation

To investigate the relationship between the explicit self-concept survey scores and the implicit self-prioritisation measures obtained from the task, we performed pairwise correlations. This analysis showed significant relationships within the explicit and implicit measures, but not across (*Figure 1*). The explicit SCIM and SCCS scores were strongly negatively correlated ($r = -0.86$, $p = 1.39x10^{-87}$, $BF_{10} = 7.37x10^{83}$) suggesting they measure very similar underlying constructs as their high scores indicate opposite self-concept quality. The implicit self measures, d’ self-advantage and reaction time self-advantage, were weakly negatively correlated (r = -0.16, p = 6.81x10^{-3}, BF_{10} = 2.82). This suggests that participants who had a greater difference in sensitivity to self (vs other) had a smaller difference in reaction time advantage to self (vs other). All four contrasts between implicit and explicit measures were non-significant by NHST statistics. Bayesian Pearson correlations show that there is evidence for no relationship between explicit and implicit measures (d’ self-advantage and SCCS, BF_{10} = 0.083 (strong); d’ self-advantage and SCIM, BF_{10} = 0.075 (strong); reaction time self-advantage and SCCS, BF_{10} = 0.11 (strong); reaction time self-advantage and SCIM, BF_{10} = 0.40 (anecdotal)). This suggests that in answer to our first research question, it is not merely that there is a lack of evidence for a relationship between our implicit and explicit self measures, but our data provides evidence against such a relationship.

Figure 1 - Correlation Matrix Self Measures

Stronger negative correlations are given in an increasingly darker blue shade, and stronger positive correlations in increasingly darker orange. Non-significant Bonferroni corrected (six comparisons) Pearson correlations with Bayesian evidence for the null hypothesis are indicated by an X.



Psychiatric Traits and Self Measures

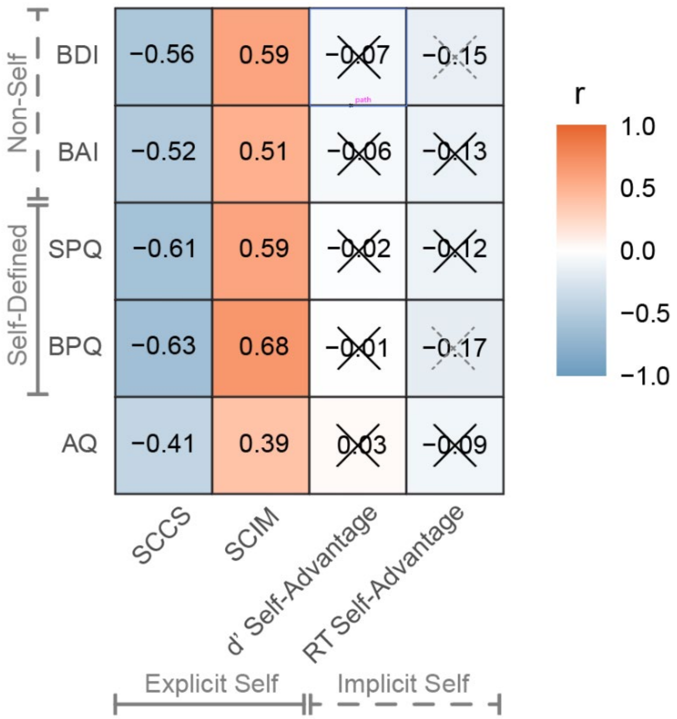
Simple Correlations

Our next aim was to investigate the strength of the relationship between traits for the psychiatric conditions and self measures. The initial analysis showed that all of the psychiatric trait measures were significantly correlated with the explicit self-concept measures (Figure 2). Higher psychiatric traits in general are associated with poorer explicit self-concept as measured by both the SCCS and SCIM. Further, the correlations between AQ and SCCS (r = -0.41, p = 5.12×10^{-13}, BF_{10} =1.36×10^{10}) and SCIM (r = 0.39, p = 6.28×10^{-12}, BF_{10} =1.18×10^{9}) are numerically weaker than those between the non-self-conditions and SCCS (BAI: r = -0.52, p = 1.77×10^{-21}, BF_{10} = 2.73×10^{18} ; BDI: r = -0.56, p = 1.01×10^{-24}, BF_{10} =4.29×10^{21}) and SCIM (BAI: r = 0.51, p = 2.07×10^{-20} , BF_{10} =2.43×10^{17}; BDI: r = 0.59, p = 7.60×10^{-29}, BF_{10} = 4.98×10^{25}). Further, as would be expected by their classifications, the non-self-defined conditions have a numerically weaker relationship with the explicit measures than between the self-conditions and SCCS (BPQ: r = -0.63, p = 8.08×10^{-34}, BF_{10} = 4.06x1050 ; SPQ: r = -0.61, p = 1.82×10^{-30}, BF_{10} =1.99x1027) and SCIM (BPQ: r = 0.68, p = 4.74×10^{-40}, BF_{10} = 5.88×10^{36} ; SPQ: r = 0.59, p = 7.16×10^{-29}, BF_{10} = 5.29×10^{25}).

By traditional NHST correlations, neither of the implicit self measures were correlated with any of the trait measures. Bayesian Pearson correlations indicate anecdotal evidence for a weak negative correlation between BDI and reaction time self-advantage (r = -0.15, BF_{10} = 1.59) and moderate evidence for a weak negative correlation between BPQ and reaction time self-advantage (r = -0.17, BF_{10} =3.99). There was anecdotal evidence against a correlation between reaction time self-advantage for SPQ (BF_{10} = 0.52) and BAI (BF_{10} = 0.74), and moderate evidence against a relationship with AQ (BF_{10} = 0.24). There was strong evidence against a correlation between d’ self-advantage and AQ (BF_{10} = 0.083), BPQ (BF_{10} =0.076) and SPQ (BF_{10} = 0.078), and moderate evidence against a correlation between d’ self-advantage and BAI (BF_{10} = 0.13) and BDI (BF_{10} = 0.14).

Figure 2 – Correlation Matrix Psychiatric Traits and Self Measures

Psychiatric trait measures are on the y-axis, and self measures along the x-axis. Stronger negative correlations are given in an increasingly darker blue shade, and stronger positive correlations in increasingly darker orange. Non-significant Bonferroni corrected (20 comparisons) NHST correlations are indicated by an X. Dashed, grey Xs indicate that while NHST statistics showed a non-significant relationship, Bayesian equivalents showed evidence for a relationship.



Regression Models

The next step was to compare how well the self measures predicted variance in the trait scores. All of the tested regression models were significant (with extreme evidence for H_1), demonstrating that at least some of the self measures account for some of the variance in all measured psychiatric trait scales. Results show that the self measures explain the most variance for BPQ with 47% variance explained, followed by SPQ which is similar to BDI, followed by BAI and lastly, AQ, which has only 17% variance explained. One or both explicit self measures is a significant predictor in all models, consistent with the correlation results above. Of the two implicit self measures, only for the BPQ model is the reaction time self-advantage a significant predictor across both statistical methods (BF$_{\text{inclusion}}$ = 1.10). Otherwise, the implicit self measures did not contribute to the regression models. A summary including significant predictors for each model can be found in Table 4.

Table 4 - Summary of Multiple Linear Regressions

Significant predictors in NHST regressions matched winning Bayesian model in all cases except for variable marked with * indicating that it was not present in Bayesian winning model.

NHST							Bayesian	
Trait	Adjusted R-squared	F-statistic (2,285)	p-value	Significant Predictors	t-value	p-value	BF$_{10}$ winning model	P(M|data)
AQ	0.17	15.2	2.84x 10^{-11}	Intercept SCCS	5.27 -2.71	2.78x10^{-7} 0.0071	1.26x10^{10}	0.47
BAI	0.29	29.9	9.58x 10^{-21}	Intercept SCCS SCIM	3.13 -3.31 2.13	0.0020 0.0011 0.034	2.90x10^{18}	0.29
BDI	0.37	42.0	1.18x 10^{-27}	SCIM RT Self-Advantage*	4.56 -2.11	7.68x10^{-6} 0.036	3.27x10^{25}	0.38
SPQ	0.38	45.8	1.23x 10^{-29}	Intercept SCCS SCIM	4.01 -4.12 2.76	7.74x10^{-5} 4.94x10^{-5} 0.0062	7.58x10^{27}	0.58
BPQ	0.47	65.0	5.93x 10^{-39}	Intercept SCCS SCIM RT Self-Advantage	2.12 -2.40 5.67 -2.40	0.029 0.017 3.53x10^{-8} 0.017	4.23x10^{36}	0.29

Direct Statistical Comparison of Correlations

Numerically comparing Pearson’s r from the simple correlations and adjusted r-squared from the regression models suggest that the self measures were least related to AQ score, most related to the self-defined conditions, with the non-self-conditions falling in between. This final part of the analysis directly compares the strength of relationship between the self measures and the traits scores. Following the PCA, the top two components were selected and with a cumulative variance of 75.9%. Component loadings are available in Table 5. While PCA components are not interpretable in themselves, these two components neatly map onto our distinction between explicit (Component 1) and implicit (Component 2) self measures. From these loadings, the variables *C1:Explicit* and *C2:Implicit* were created for each participant for further analysis. It should be noted that Multiple Linear Regression models as above using these components as predictors yielded comparable results to those in Table 4, except that C2:Implicit was not a significant predictor for the BDI model (it remained a significant predictor of BPQ).

Table 5- PCA Analysis Details for Self-Measure Dimensionality Reduction

	Component 1	**Component 2**
Variable	**Loadings**	
SCCS	0.963	
SCIM	-0.963	
d’ Self-Advantage		0.772
RT Self-Advantage		-0.751
	Eigenvalue	
	1.88	1.55
	% of Variance	
	47.01	28.88

Results from the correlation comparison analysis can be found in Table 6. Only in comparing the relationship between AQ and BAI and C1:Explicit did any of the methods reported by the toolbox disagree (30% of the methods indicate that the null hypothesis should be rejected). In summary, AQ has a weaker correlation with C1:Explicit than BPQ, SPQ and BDI do, but the strength of the correlation is not significantly different from that with BAI. The correlation between AQ and C2: Implicit is not significantly different to that of the other trait measures (none of which were significant to begin with, see Figure 2).

Table 6 - Comparing Correlations between AQ and Simplified Self Measures with Other Trait Measures and Simplified Self Measures

Pearson and Filon's z: *** = p<0.0001, ** = p<0.0005, * = p<0.00625, X = p>0.00625

AQ *compared to* *correlated with*	**BPQ**		**SPQ**		**BDI**		**BAI**	
	z	*p*	*z*	*p*	*z*	*p*	*z*	*p*
C1: Explicit	5.64	<0.00001	4.63	<0.00001	3.59	0.0003	2.08	0.0380
	***		***		**		X	
C2: Implicit	-1.2283	0.2193	-0.4699	0.6385	-0.8301	0.4065	-0.5026	0.6152
	X		X		X		X	

Our control comparisons between correlations within and between our self-defined and non-self-defined categories are reported in Table 7. In summary, C1:Explicit successfully distinguishes between BPQ and both the non-self-defined conditions, but SPQ shows statistically equivalent relationships to the non-self-defined conditions across all measures. Again, comparisons between relationships with C2:Implicit are not significant, but neither are any of the first order correlations (Figure 2).

Table 7 - Comparing Correlations between Self-defined and Non-self-defined Psychiatric Traits with Simplified Self Measures

Pearson and Filon's z: *** = p<0.0001, ** = p<0.0005, * = p<0.00625, X = p>0.00625

		Self-defined Conditions							
		BPQ				**SPQ**			
		C1:Explicit		*C2:Implicit*		*C1:Explicit*		*C2:Implicit*	
		z	*p*	*z*	*p*	*z*	*p*	*z*	*p*
Non-self-defined Conditions	**BDI**	-2.7915	0.0052	0.5180	0.6045	-0.4312	0.6663	-0.5489	0.5831
		*		X		X		X	
	BAI	-4.0888	<0.00001	0.8179	0.4134	-2.1207	0.0339	-0.1769	0.8596
		***		X		X		X	

Discussion

For this study, we used self-report and task data indexing the quality of implicit and explicit self-representations and the level of psychiatric traits for self-defined and non-self-defined conditions including autism. Participants completed two self-concept surveys, five psychiatric traits surveys and completed a shape-label matching task based on Sui et al. (2012). We did so to answer two independent research questions and will discuss each of these in turn in what follows. In brief, our results suggest a dissociation between explicit and implicit self-representations within an individual. We also find that differences in self-cognition are not sufficiently strongly associated with autistic traits to warrant their inclusion as diagnostic criteria, even though self-concept clarity scores did significantly predict autism traits.

Is there coherence across implicit and explicit self-cognition?

The first research question was restricted to the self measures and asked whether there was a relationship between the implicit and explicit measures of self-representation. Our data provides evidence for no relationship between explicit self-concept, as measured by the SCCS and SCIM, and implicit self-prioritisation, as measured by d' self-advantage and reaction time self-advantage from the shape-label matching task. This is consistent with findings of dissociation in self-cognitive measures from Nijhof et al. (2020).

Conclusions that can be drawn from this data are limited to the specific cognitive domains studied. Self-cognition has numerous facets across the cortical hierarchy that have not been incorporated in the current study, including the bodily self, self in action, memory for self, self-recognition and self-related language use (Perrykkad & Hohwy, 2020). As we reported earlier, studies such as Krol et al. (2019) have found intra-individual relationships between self-cognitive domains. While it is still plausible that attentional mechanisms which typically lead to a prioritisation of the self at low level cognitive processing also impact on downstream integrated self-representations at the explicit level, these data suggest that it is not always an easy line to draw from one domain to the other. This may be especially true when comparing the lowest and highest levels, as we ostensibly did here, between which there are many intervening factors.

While not one of our aims, it is interesting to note that our results show that explicit self-concept is more closely related to traits of psychiatric conditions than implicit, low-level, perceptual and attentional self-prioritisation. This is borne out by the lack of significant correlations between the implicit self measures and any of the psychiatric traits, the rare appearance of implicit self measures in the regression models predicting psychiatric traits and the absence of implicit measures in contributing to discriminability of self-defined conditions from non-self-defined conditions. In contrast, both explicit self measures are correlated with all psychiatric trait measures, at least one of them significantly contributes to the regression model for each psychiatric trait score, and the combination of explicit measures successfully distinguishes BPQ from both BDI and BAI. This might be because psychiatric diagnostic criteria (either including or excluding self-cognition) comes from years of clinically observable features. Explicit self-concept is more clinically observable than differences in attentional differences as measured by discriminability or reaction time in the task. This is because it is, qua explicit, the kind of thing that is acknowledged and reportable by the individual, and thereby more accessible to a clinician. This is not to say that implicit measures cannot be clinically relevant, just that distinctions which are based on years of clinical observations (as between self-defined and non-self-defined conditions) are likely to be reflected by features that are easily observable in a clinical setting.

Nevertheless, we would encourage further research into the possible importance of the reaction time self-advantage measure from the shape-label matching task. The Bayesian analysis suggested that there was some support for a weak negative relationship between this measure and BDI and BPQ, despite its non-significance in the conservative NHST statistics. Further, it was a significant predictor for the regression models for BDI and BPQ. These analyses together suggest that there may be more to investigate on this measure of self-bias for depression and borderline personality disorder, despite the relative dismissal of the implicit self measures for psychiatric conditions in this study.

Is self-cognition characteristic of autism?

The second research question was whether autism was more similar to self-defined conditions or non-self-defined conditions. Our results suggest that the relationship between subclinical autism traits and explicit self measures is more similar to their relationship with subclinical traits of conditions that are not characterised by differences in self cognition, than those that are. As such, based on this data alone, differences in self-cognition are not characteristic enough of autism to warrant their inclusion as diagnostic criteria.

Evidence for this comes from several of our analyses. First, the amount of variance in psychiatric traits explained by the self measures is the lowest in autism of all studied conditions, at 17%, and only SCCS scores significantly contributed to this regression (for all other conditions at least two self measures contributed to the prediction in NHST models). Second, while autism traits were moderately correlated with explicit self measures, the strength of these correlations was less than those between the same explicit self measures and both the self-defined condition traits (BPQ and SPQ) and one of the non-self-defined-conditions (BDI) and not significantly different from the other (BAI). This places the relevance of these differences for autism squarely in line with the non-self-defined conditions, and below the self-defined conditions.

This is not to say, however, that self differences are not related to autistic traits in our measures. It is important to highlight that there was extreme evidence for significant correlations between AQ and both SCCS and SCIM, however, the strength of this correlation was not significantly different to their links with anxiety symptom severity (BAI). There was also extreme evidence that SCCS scores contribute to a prediction of autistic traits. All this to say that our data does suggest that there are self-cognitive differences related to autism. One of the primary goals of diagnostic criteria however, is to be not only sensitive to differences in a population, but choose features which are specific to that population. As we discussed in the introduction, self-cognition differences are broadly associated with psychiatric conditions (Neacsiu, Herr, Fang, Rodriguez, & Rosenthal, 2015), and so they should be particularly strong to warrant their inclusion in diagnostic criteria. Our threshold here for ‘particularly strong’ was ‘at least as strong as other conditions which are currently defined by self-differences’. Based on our data, autism does not meet such a threshold.

It is important to highlight that our control analysis directly contrasting self-defined and non-self-conditions did not show any differences between SPQ and BAI or BDI on our self measures. One possible reason is that Schizophrenia, while defined as having self-features in the ICD-11 (World Health Organisation, 2018), is not defined by self-specific features in the DSM-5, despite involving symptoms which often relate to the self in presentation (for example, ‘delusions’ often involve delusions of control as in the ICD-11 classification, see Table 1)(American Psychiatric Association, 2013). Borderline personality disorder, on the other hand, involves identity disturbances in both the ICD-11 and DSM-5 criteria, and our results show that it is significantly more strongly related to our explicit self measures than either BDI or BAI. This makes it the more prototypical self-defined condition in our study, while schizophrenia seems to sit between borderline personality disorder and the non-self-defined-conditions.

One can, of course, also disagree with the assumption that there is a clear distinction between self-defined psychiatric conditions and non-self-defined psychiatric conditions at the outset. It may be more appropriate to conceive of a transdiagnostic multi-axial spectrum of self-cognition. In this sense self-cognition is an important feature of both our ‘self-defined’ and ‘non-self-defined’ conditions because self-cognitive measures correlated with and predicted traits for *all* of them. Along such a spectrum, high borderline personality disorder traits appear to fall to the furthest extreme of the self-concept dimension (see Table 7). There is still conceptual room below the autism traits correlation on these axes – for a condition that has no correlation with self-cognitive measures. We did find an association between explicit self-cognition and autism in our dataset, and thus, it would be hasty to dismiss the importance of self-cognition for understanding autism outright. Our data also suggests that the explicit measures used here are a better candidate for a trans-diagnostic dimension than are responses to the shape-label matching task.

There is a now long history of considering dimensional approaches both within and across mental conditions as opposed to merely relying on traditional categories based on the presence or absence of symptoms (Goldberg, 1996; Haslam, 2003; Helzer, Kraemer, & Krueger, 2006; Kendell, 1975; Kessler, 2002; Rosen, Lord, & Volkmar, 2021). More recent frameworks which endorse the move towards dimensional, diagnostically agnostic, research projects include the National Institute of Mental Health's Research Domain Criteria (RDOC). This is an alternative research program to the traditional ICD and DSM diagnostic classification systems in which multidimensional neuro-cognitive data drives psychopathological research (see Clark, Cuthbert, Lewis-Fernández, Narrow, and Reed (2017)). The RDOC includes aspects of self-cognition as a proposed dimension in two of its domains (systems for social processes and sensorimotor systems). This allows for more flexibility and scepticism of the existing diagnostic categories given psychophysiological evidence.

Even with the pitfalls of diagnostic rigidity in mind, it is still important to note that there are important limitations of using trait-based measures of our psychiatric conditions. Further research should be done comparing self measures in diagnosed populations and in participants with no diagnosis of any psychiatric condition. We chose trait-based measures to enable a within-subjects design, but our choice of psychiatric conditions in each category was also limited by the need for comparable measures. For self-defined conditions, borderline personality disorder and schizophrenia were the optimal choice regardless, but more varied non-self-defined conditions may have been preferable. It is possible that anxiety and depression are less amenable to self differences because they are both sometimes transient conditions, while our other conditions are developmental or lifelong. Future research in diagnosed populations should also consider using cognitive conditions, such as attention deficit hyperactivity disorder, as additional contrasts to self-defined conditions. These might be the kind of condition which shows no relationship with self-cognition at all.

Conclusion

In this study, we were interested in two questions. First, whether self-cognition measures across the cognitive hierarchy are associated within an individual; and second, whether or not it was warranted to add self-cognitive differences to the diagnostic criteria of autism. In summary, data presented here from 288 participants suggests no relationship between low-level, implicit, attentional self-biases in sensitivity and response time to self-stimuli and higher-order, explicit, self-concept clarity and stability. Further, our results indicate that the relationship between self-cognition and autistic traits is of a degree that is more similar to non-self-defined conditions than those that are defined by self-differences. As such, based on the data presented here, it is not warranted to add self-cognitive differences to the diagnostic criteria for autism.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*: American Psychiatric Pub.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5-17. doi:10.1023/A:1005653411471

Beck, A., Epstein, N., Brown, G., & Steer, R. (1988). An Inventory for Measuring Clinical Anxiety: Psychometric Properties.

Beck, A., Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. (1961). Beck depression inventory (BDI). *Arch Gen Psychiatry, 4*(6), 561-571.

Berna, F., Göritz, A. S., Schröder, J., Coutelle, R., Danion, J.-M., Cuervo-Lombard, C. V., & Moritz, S. (2016). Self-Disorders in Individuals with Autistic Traits: Contribution of Reduced Autobiographical Reasoning Capacities. *Journal of Autism and Developmental Disorders, 46*(8), 2587-2598. doi:10.1007/s10803-016-2797-2

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of personality and social psychology, 70*(1), 141.

Chiu, C.-D., Chang, J.-H., & Hui, C. M. (2017). Self-concept integration and differentiation in subclinical individuals with dissociation proneness. *Self and Identity, 16*(6), 664-683. doi:10.1080/15298868.2017.1296491

Cicero, D. C., Martin, E. A., Becker, T. M., & Kerns, J. G. (2016). Decreased Self-Concept Clarity in People with Schizophrenia. *The Journal of nervous and mental disease, 204*(2), 142-147. doi:10.1097/NMD.0000000000000442

Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychol Sci Public Interest, 18*(2), 72-145. doi:10.1177/1529100617727266

Cygan, H. B., Tacikowski, P., Ostaszewski, P., Chojnicka, I., & Nowicka, A. (2014). Neural Correlates of Own Name and Own Face Detection in Autism Spectrum Disorder. *PloS one, 9*(1), e86020. doi:10.1371/journal.pone.0086020

Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PloS one, 10*(4), e0121945. doi:10.1371/journal.pone.0121945

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assessment, Research & Evaluation, 14*(20), 1-11.

Dritschel, B. M., Wisely, M., Goddard, L., Robinson, S., & Howlin, P. (2010). Judgements of self-understanding in adolescents with Asperger syndrome. *autism, 14*(5), 509-518. doi:10.1177/1362361310368407

Frith, U., & Happé, F. (1999). Theory of mind and self-consciousness: What is it like to be autistic? *Mind & Language, 14*(1), 82-89.

Goldberg, D. (1996). A dimensional model for common mental disorders. *The British Journal of Psychiatry, 168*(S30), 44-49.

Haslam, N. (2003). Categorical Versus Dimensional Models of Mental Disorder: The Taxometric Evidence. *Australian & New Zealand Journal of Psychiatry, 37*(6), 696-704. doi:10.1080/j.1440-1614.2003.01258.x

Helzer, J. E., Kraemer, H. C., & Krueger, R. F. (2006). The feasibility and need for dimensional psychiatric diagnoses. *Psychological medicine, 36*(12), 1671.

Hobson, P. R. (2011). Autism and the self. In S. Gallagher (Ed.), *The Oxford Handbook of the Self* (pp. 571-591). Oxford, UK: Oxford University Press.

Huang, A. X., Hughes, T. L., Sutton, L. R., Lawrence, M., Chen, X., Ji, Z., & Zeleke, W. (2017). Understanding the Self in Individuals with Autism Spectrum Disorders (ASD): A Review of Literature. *Frontiers in psychology, 8*(1422). doi:10.3389/fpsyg.2017.01422

Inquisit Web. (2020). Self-Referential Processing. Retrieved from https://www.millisecond.com

Jeffreys, H. (1998). *The theory of probability*: OUP Oxford.

Kaufman, E. A., Cundiff, J. M., & Crowell, S. E. (2015). The Development, Factor Structure, and Validation of the Self-concept and Identity Measure (SCIM): A Self-Report Assessment of Clinical Identity Disturbance. *Journal of Psychopathology and Behavioral Assessment, 37*(1), 122-133. doi:10.1007/s10862-014-9441-2

Kaufman, E. A., Puzia, M. E., Crowell, S. E., & Price, C. J. (2019). Replication of the Self-Concept and Identity Measure (SCIM) Among a Treatment-Seeking Sample. *Identity, 19*(1), 18-28. doi:10.1080/15283488.2019.1566068

Kendell, R. E. (1975). *The role of diagnosis in psychiatry*. Oxford, England: Blackwell Scientific Publications.

Kessler, R. C. (2002). The Categorical versus Dimensional Assessment Controversy in the Sociology of Mental Illness. *Journal of Health and Social Behavior, 43*(2), 171-188. doi:10.2307/3090195

Krol, S. A., Thériault, R., Olson, J. A., Raz, A., & Bartz, J. A. (2019). Self-Concept Clarity and the Bodily Self: Malleability Across Modalities. *Personality and Social Psychology Bulletin*, 0146167219879126. doi:10.1177/0146167219879126

Leekam, S. R., & Ramsden, C. A. H. (2006). Dyadic Orienting and Joint Attention in Preschool Children with Autism. *Journal of Autism and Developmental Disorders, 36*(2), 185. doi:10.1007/s10803-005-0054-1

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior research methods, 49*(2), 433-442.

Lombardo, M. V., & Baron-Cohen, S. (2010). Unraveling the paradox of the autistic self. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(3), 393-403.

Lyons, V., & Fitzgerald, M. (2013). Atypical Sense of Self in Autism Spectrum Disorders: A Neuro-Cognitive Perspective. In *Recent Advances in Autism Spectrum Disorders - Volume I*.

Mars, A. E., Mauk, J. E., & Dowrick, P. W. (1998). Symptoms of pervasive developmental disorders as observed in prediagnostic home videos of infants and toddlers. *The Journal of Pediatrics, 132*(3), 500-504. doi:https://doi.org/10.1016/S0022-3476(98)70027-7

Marsman, M., & Wagenmakers, E.-J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology, 14*(5), 545-555. doi:10.1080/17405629.2016.1259614

Molnar-Szakacs, I., & Uddin, L. Q. (2016). The Self in Autism. In M. Kyrios, R. Moulding, G. Doron, S. S. Bhar, M. Nedeljkovic, & M. Mikulincer (Eds.), *The Self in Understanding and Treating Psychological Disorders* (pp. 144-157). Cambridge, UK: Cambridge University Press.

Nadig, A. S., Ozonoff, S., Young, G. S., Rozga, A., Sigman, M., & Rogers, S. J. (2007). A prospective study of response to name in infants at risk for autism. *Archives of Pediatrics & Adolescent Medicine, 161*(4), 378-383. doi:10.1001/archpedi.161.4.378

Nijhof, A., Bird, G., Catmur, C., & Shapiro, K. (2020). No evidence for a common self-bias across cognitive domains. *Cognition, 197*.

Nijhof, A., Dhar, M., Goris, J., Brass, M., & Wiersema, R. (2018). Atypical neural responding to hearing one's own name in adults with Autism Spectrum Disorder. *Journal of Abnormal Psychology, 127*(1), 129-138.

Nowicka, M. M., Wójcik, M. J., Kotlewska, I., Bola, M., & Nowicka, A. (2018). The impact of self-esteem on the preferential processing of self-related information: Electrophysiological correlates of explicit self vs. other evaluation. *PloS one, 13*(7), e0200604.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of experimental social psychology, 45*(4), 867-872.

Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday home videotapes. *Journal of Autism and Developmental Disorders, 24*(3), 247-257. doi:10.1007/BF02172225

Perrykkad, K., & Hohwy, J. (2019). When big data aren't the answer. *Proceedings of the National Academy of Sciences, 116*(28), 13738. doi:10.1073/pnas.1902050116

Perrykkad, K., & Hohwy, J. (2020). Modelling Me, Modelling You: the Autistic Self. *Review Journal of Autism and Developmental Disorders, 7*, 1-31. doi:10.1007/s40489-019-00173-y

Poreh, A. M., Rawlings, D., Claridge, G., Freeman, J. L., Faulkner, C., & Shelton, C. (2006). The BPQ: a scale for the assessment of borderline personality based on DSM-IV criteria. *Journal of personality disorders, 20*(3), 247-260.

Qian, H., Wang, Z., Li, C., & Gao, X. (2020). Prioritised self-referential processing is modulated by emotional arousal. *Quarterly Journal of Experimental Psychology, 73*(5), 688-697.

R Core Team. (2018). R: A Language and environment for statistical computing. Retrieved from https://cran.r-project.org/

Raine, A. (1991). The SPQ: A Scale for the Assessment of Schizotypal Personality Based on DSM-III-R Criteria. *Schizophrenia Bulletin, 17*(4), 555-564. doi:10.1093/schbul/17.4.555

Ridley, R. (2019). Some difficulties behind the concept of the 'Extreme male brain' in autism research. A theoretical review. *Research in Autism Spectrum Disorders, 57*, 19-27. doi:https://doi.org/10.1016/j.rasd.2018.09.007

Roepke, S., Schröder-Abé, M., Schütz, A., Jacob, G., Dams, A., Vater, A., . . . Lammers, C. H. (2011). Dialectic behavioural therapy has an impact on self-concept clarity and facets of self-esteem in women with borderline personality disorder. *Clinical psychology & psychotherapy, 18*(2), 148-158.

Rosen, N. E., Lord, C., & Volkmar, F. R. (2021). The Diagnosis of Autism: From Kanner to DSM-III to DSM-5 and Beyond. *Journal of Autism and Developmental Disorders*, 1-18.

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1105.

Sui, J., Ohrling, E., & Humphreys, G. W. (2016). Negative mood disrupts self-and reward-biases in perceptual matching. *Quarterly Journal of Experimental Psychology, 69*(7), 1438-1448.

The Jamovi Project. (2019). jamovi (Version 0.9). Retrieved from https://www.jamovi.org

Uddin, L. Q. (2011). The self in autism: An emerging view from neuroimaging. *Neurocase, 17*(3), 201-208. doi:10.1080/13554794.2010.509320

Vanden Poel, L., & Hermans, D. (2019). Narrative Coherence and Identity: Associations With Psychological Well-Being and Internalizing Symptoms. *Frontiers in psychology, 10*(1171). doi:10.3389/fpsyg.2019.01171

Williams, D. (2010). Theory of own mind in autism: Evidence of a specific deficit in self-awareness? *autism, 14*(5), 474-494. doi:10.1177/1362361310366314

Williams, D., Nicholson, T., & Grainger, C. (2017). The Self-Reference Effect on Perception: Undiminished in Adults with Autism and No Relation to Autism Traits. *Autism Research*, n/a-n/a. doi:10.1002/aur.1891

Wong, A. E., Dirghangi, S. R., & Hart, S. R. (2019). Self-concept clarity mediates the effects of adverse childhood experiences on adult suicide behavior, depression, loneliness, perceived stress, and life distress. *Self and Identity, 18*(3), 247-266. doi:10.1080/15298868.2018.1439096

World Health Organisation. (2018). International classification of diseases for mortality and morbidity statistics (11th Revision). Retrieved from https://icd.who.int/browse11/l-m/en

Zhao, S., Uono, S., Yoshimura, S., & Toichi, M. (2018). A functional but atypical self: Influence of self-relevant processing on the gaze cueing effect in autism spectrum disorder. *Autism Research, 0*(0). doi:doi:10.1002/aur.2019

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience, 23*(2–3), 143-152. doi:https://doi.org/10.1016/j.ijdevneu.2004.05.001

Acknowledgements

The authors would like to thank the participants of this study for their efforts during a global pandemic. Thanks also to Manja Engel and Andrew Corcoran who provided feedback on an earlier draft of this manuscript. Kelsey Perrykkad is supported by an Australian Government Research Training Program Scholarship. This project was supported by the Australian Research Council (DP190101805) through Jakob Hohwy.

While it seems there is overwhelming evidence that self-cognition is different in autism as we saw in Chapter 1, the results of this chapter suggest that the relationship between different aspects of self-cognition, especially at vastly different levels of the neural hierarchy may be more complicated than I implied in earlier Chapters.

At this point, the direction of the thesis takes a turn, to focus primarily on one aspect of self-cognition that presented itself as an anomaly in the Chapter 1 - Judgement of Agency. The next chapter sets the stage for this transition, giving a predictive processing account of fidgeting (and autistic ‘stimming’ or repeated and repetitive behaviours) as a response to rising uncertainty that reaffirms the self-model.

Chapter 4. Fidgeting as Self Evidencing: A predictive processing account of non-goal-directed action

In this chapter, we transition from a non-specific notion of the self to focusing specifically on evidence for the self-model that comes from *action*. Following this chapter, in a preface to the series of experiments that follow, I will discuss in more detail why action is particularly important for modelling the self. This chapter will act to lay the ground for the concepts that are important going forward. The paper presented is not experimental, and while it does review the very limited existing literature in the area, it is not fundamentally a review paper. It rather takes the form of a philosophical argument.

This paper argues that fidgeting is best understood as a response to unexpected uncertainty about the effectiveness of ones' actions in the world, and provides predictable sensory consequences across environments, which act in a *self-evidencing* way to reaffirm the self-model. In simple terms, we fidget to reassure ourselves that we exist as we always have before. This explanation is extended to autistic 'stimming', which might be understood as a more extreme, more frequent, or socially less acceptable form of neurotypical fidgeting. In this way, the paper provides a detailed predictive processing account for one of the core features of autism – restricted and repetitive behaviours. It also demonstrates how this core feature is intimately tied to autistic self-cognition and may stem from differences in the self-model.

Self-evidencing is a technical concept from philosophy of science, usually used to understand what makes a good explanation (Hempel, 1965; Lipton, 2004). The basic idea is that the best hypothesis that explains some evidence is itself supported by evidence that was generated under it. The classic example is of a set of footprints in the snow outside one's window, the explanation for which is hypothesised to be a burglar. However, if someone were to ask you what evidence you had that there was a burglar (how do you know that the hypothesis is true), the response would be to reference the footprints (evidence). And thus, the explanation is somewhat circular – I believe there is a burglar because there are footprints, and because there are footprints I believe there is a burglar

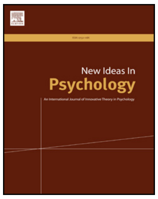
(Lipton, 2004, p. 24). While this seems methodologically flawed, it is in fact a benign and very common form for scientific explanations. For more detail about how this form of reasoning can remain unproblematic in the specific case of the predictive brain, see Hohwy (2016).

In the following paper, I use the term ‘self-evidencing’ in this classic technical way. However, it is also used in a way that more strongly related to the self. In this paper, I argue that fidgeting actions are self-evidencing about the self. The hypotheses generated by the self-model are fulfilled by evidence from the world in the form of sensory consequences, which in turn provide evidence for the self.

Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: www.elsevier.com/locate/newideapsych



Fidgeting as self-evidencing: A predictive processing account of non-goal-directed action

Kelsey Perrykkad[*], Jakob Hohwy

Cognition and Philosophy Lab, Philosophy Department, School of Philosophical Historical and International Studies, Monash University, Victoria, Australia

ARTICLE INFO

Keywords:
Fidgeting
Self-evidencing
Non-goal-directed action
Prediction error minimisation
Autism spectrum conditions

ABSTRACT

Non-goal-directed actions have been relatively neglected in cognitive science, but are ubiquitous and related to important cognitive functions. Fidgeting is seemingly one subtype of non-goal-directed action which is ripe for a functional account. What's the point of fidgeting? The predictive processing framework is a parsimonious account of brain function which says the brain aims to minimise the difference between expected and actual states of the world and itself, that is, minimise prediction error. This framework situates action selection in terms of active inference for expected states. However, seemingly aimless, idle actions, such as fidgeting, are a challenge to such theories. When our actions are not obviously goal-achieving, how can a predictive processing framework explain why we regularly do them anyway? Here, we argue that in a predictive processing framework, evidence for the agent's own existence is consolidated by self-stimulation or fidgeting. Endogenous, repetitive actions reduce uncertainty about the system's own states, and thus help continuously maintain expected rates of prediction error minimisation. We extend this explanation to clinically distinctive self-stimulation, such as in Autism Spectrum Conditions, in which effective strategies for self-evidencing may be different to the neurotypical case.

1. Introduction: fidgeting and non-goal directed actions

In investigating the psychological and neuroscientific bases of movement, cognitive science has primarily focused on goal-directed action. These actions can be generally defined as those whose function is to achieve some instrumental aim for the individual. For example, reaching for a glass of water to quench one's thirst. However, there is a relatively neglected category of action that has not been well explained under cognitive theories of action. This is non-goal-directed action, which incorporates behaviours such as fidgeting. This common distinction between *goal-directed action* and *non-goal-directed action* is often subtly made, without definition, but rather by attributing a certain cognitive mechanism to goal-directed action specifically. Since goal-directed action is defined in opposition to non-goal-directed action, the latter cannot be about achieving a personal-level end. Other than this negative definition, it is not well explicated (Dayan, 2009). Sometimes, this contrast class is labelled *habits* or *reflexes* (Barfield et al., 2017; Friston et al., 2016).

We label our target class of non-goal-directed actions in this paper, *fidgeting.* Fidgeting is defined partly in opposition to other action types, such as habits. Habits colloquially include automatically performed sequences of productive action (for example brushing your teeth before bed), whereas, superficially at least, fidgeting appears to have no epistemic or pragmatic value. Examples of fidgeting thus excludes brushing one's teeth or always walking the dog along a certain route but includes tapping the table with one's finger or twirling one's hair. The primary outcome of fidgeting, as we define it here, can often be described as reflexive stimulation; caused by the individual, to the individual. Importantly, stimulation of the kind intended here is usually repetitive or patterned and is both self-initiated and self-sustained. Agents may be unaware or aware that they are fidgeting. Typically, agents have not consciously decided to fidget but fidgeting behaviours can be intentionally (and consciously) terminated, resisted or permitted. Note that fidgeting is still distinct from reflexive, goal directed stimulation (such as scratching an itch), because there is no obvious person-level goal that is achieved through the action. Stimulation can occur in multiple sensory domains and is often primarily visual, tactile or auditory (see Table 1 for examples of the target behaviour in different modalities). It can be performed with the body alone, or with a prop, such as a piece of string, a pen and paper or a light source.

There is little consensus among the scientific community about the reason we fidget. There is a possibility that it serves no purpose. Some philosophers use it as the paradigm case of sub-intentional action or mere movement, which, by definition, serves no end for the agent

[*] Corresponding author. 20 Chancellors Walk, Monash University, Clayton, VIC, 3800, Australia.
E-mail address: kelsey.perrykkad@monash.edu (K. Perrykkad).

https://doi.org/10.1016/j.newideapsych.2019.100750
Received 16 April 2019; Received in revised form 26 July 2019; Accepted 30 July 2019

Table 1
Examples of fidgeting in different modalities for typically developing adults.

Primary Modality	*Examples*
Visual	Doodling
	Visually tracking a rotating fan
	Absentmindedly arranging objects on desk
Vestibular	Rocking on a chair
	Absentminded head nodding
Tactile	Playing with own hair
	Touching own face
	Rubbing a soft sweater
Auditory	Clicking a pen
	Tapping foot
	Humming
Taste	Chewing gum
	Sucking on a toothpick
Proprioception	Bouncing one's leg

(Hommel, 2015, pp. 307–326; O'shaughnessy, 1980). However, upon further examination there is growing evidence that fidgeting is importantly and systematically involved in many psychological processes, which suggests a deeper story. Fidgeting is commonly thought to be indicative of a lack of attention, and is increased in individuals with Attention Deficit Hyperactive Disorders (Lis et al., 2010). In the context of lecture based learning, fidgeting was shown to predict recall independently of attention, with increased fidgeting associated with decreased memory (Farley, Risko, & Kingstone, 2013). However, unlike other tasks performed alongside a primary task, doodling has been shown to aid incidental memory (Andrade, 2010). Exaggerated and more frequent fidgeting behaviours are also implicated in psychiatric and neurological conditions such as the stereotypies found in Autism Spectrum Conditions, Rett syndrome, schizophrenia and people who are blind (Barry, Baird, Lascelles, Bunton, & Hedderly, 2011; Morrens, Hulstijn, Lewi, De Hert, & Sabbe, 2006). Together, these findings suggest that fidgeting is not purposeless and, accordingly, several functional accounts of fidgeting have been proposed.

One prominent theory regards fidgeting as *bodily regulation*, such that we fidget to release excess stored energy or reduce fat mass gain (Johannsen & Ravussin, 2008). However, though spontaneous physical activity (including but not limited to fidgeting) is inversely correlated with obesity and was associated with changes in energy expenditure, perturbations in physical activity or weight do not change the amount of spontaneous physical activity (Johannsen & Ravussin, 2008) so the explanation cannot be this simple. Many of the studies under this theory consider fidgeting to be just one of many similar, non-exercise forms of energy expenditure, called non-exercise activity thermogenesis (NEAT), and includes walking and standing along with fidgeting. Thus, the explanandum of these proposals is broader than the current project.

In certain clinical cases in particular, fidgeting may help the body regulate its own function. Fidgeting in patients with autonomic failures has been shown to counteract symptomatic drops in blood pressure (Cheshire, 2000), and may also prevent endothelial dysfunction by improving blood flow during prolonged sitting (Morishima et al., 2016). Perhaps then it is natural to think that the usual instance of fidgeting is associated with autonomic self-regulation where movements would always increase bloodflow and blood pressure. However, fidgeting often occurs in situations of high stress, where the self-regulatory autonomic goal should be to decrease these cardiac parameters (assuming one is regulating automatic fight or flight responses). Further, these purely homeostatic and bodily explanations do not explain the cognitive effects associated with fidgeting, or why stereotypies occur in psychiatric conditions without clear autonomic dysfunction. As such, these explanations do not capture the whole story. While it should be noted that there may be cases of fidgeting that are restricted to specific mechanistic failures of the body (as, perhaps, in patients with specific autonomic failures), we suggest that an explanation of fidgeting in the general case should account for both autonomic regulation and cognitive elements of the phenomenon. In this paper, we shall appeal to active inference as the best framework for explaining fidgeting, and this notion subsumes homeostatic self-regulation in a broader (allostatic) framework, addressing these misgivings about the homeostatic account of fidgeting specifically.

Another theory regards fidgeting as *cognitive regulation*. As we briefly reviewed earlier, evidence shows that fidgeting is associated with cognitive states such as attention and memory, and varied psychiatric conditions. If fidgeting is associated with so many cognitive functions, it is plausible to think the best explanation will be a cognitive or neurological one. One theory is that fidgeting is bodily *mind wandering* because the tendency to fidget and mind wander are correlated between individuals (Carriere, Seli, & Smilek, 2013). However, this fails to illuminate either notion. Further, the usual approach of correlating fidgeting behaviour with contextual differences fails to explain which features are significant to these contexts, and ignores explanatorily fruitful individual differences in behaviour. A comprehensive account of fidgeting behaviour should specify both why and when we do it. More empirical research on the causal relations between fidgeting and cognitive processes such as attention, memory and mind-wandering is needed.

Lovaas, Newsom, and Hickman (1987), argue that self-stimulatory activity occurs as a form of operant conditioning, in which the reinforcer is the resulting perceptual stimuli. This is the most similar theory to the one presented in this paper, however, it does not account for *why* self-stimulation and particular perceptual consequences should be rewarding, which our proposal aims to explain. In general, an account that relies on reinforcement learning would be attractive, because it integrates fidgeting with other forms of action.

Our explanation will rather employ the *active inference* account of action from the predictive processing framework. Briefly, this account situates action as a form of prediction error minimisation. It is a contemporary alternative to reinforcement learning that would also unify the account of fidgeting with an account of action in general. Active inference arguably has computational advantages over reinforcement learning explanations because the former does not rely on an independent value system (Friston, Daunizeau, & Kiebel, 2009; Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2017). Further, since reinforcement learning requires an action (or series of actions) to be fully completed and yield an outcome before one can learn from them, it raises an issue in explaining how and why we might choose to switch to fidgeting in the middle of another task (as the initial policy is not yet completed) (see also Friston (2017b)). The active inference account allows learning at all stages of the policy execution, and so the experience of changing to fidgeting mid-task is more amenable to our account.

The problem for this strategy is, however, that fidgeting does not obviously seem to specify any expected low prediction error state. Conversely put, fidgeting becomes a challenge for those who believe active inference is a unifying account of action. This sets the challenge for this paper, and we will argue that active inference does in fact have the resources to accommodate fidgeting, in a way that interestingly reveals fidgeting's function and throws light on the nature of self-representation.

2. Predictive processing framework

In this section, we will introduce the specific elements of the predictive processing framework relevant to fidgeting behaviour including active inference. We will explain why fidgeting poses a challenge to this framework, and will lay the groundwork for our solution.

The predictive processing account of brain function asserts that the brain hierarchically compares incoming information from the senses with descending prior expectations. Where these signals mismatch, a prediction error is propagated up the neural hierarchy. The brain's overall function is to minimise this error by making contextually appropriate adjustments of its internal expectations and thus creating a dynamic model of the hidden causes of sensory input (both environmental and bodily). This theoretical framework parsimoniously explains both perception and action as methods of prediction error minimisation (Clark, 2015; Friston, 2010; Hohwy, 2013).

2.1. Policy selection

Each individual's particular internal model of the world influences their perception and action selection enabling prediction error minimisation over their lifespan. A *policy* is a set of possible actions (or individual control states) that have been grouped together by the individual for its history of success as a strategy to reduce prediction error when faced with situations with learned commonalities, which cue success for that policy.

Policies are considered under uncertainty, in the sense that each policy has a certain probability that it will lead to a desired outcome, defined in terms of the expected sensory outcomes. Since there is not a one-to-one relation between actions and outcomes, policy selection presents an inference problem, hence the label *active inference.* This implies that, in some circumstances, agents can misrepresent causal relations between actions and outcomes when there are none. For example, Skinner (1948) describes pigeons who learn to perform "superstitious" actions for rewards given at random intervals. Policies influence the dynamics of states of the world, each of which is associated with a likelihood whose *precision* describes the fidelity of the mapping from those states to their sensory consequences. Many extremely precise policies are performed automatically without any conscious awareness, such as the movements associated with walking for an ambulatory adult. The precision of a likelihood mapping is inversely related to its *ambiguity*, where high ambiguity means it is unclear which sensory outcome should be predicted. As such, alternative policies are associated with different levels of ambiguity in virtue of the alternative states (and therefore likelihoods) whose dynamics they prescribe. Active inference is therefore sensitive to precision and ambiguity.

Active inference can minimise prediction error in two ways, for epistemic value or for utility. Epistemic value refers to actions that implicate causally intervening on modelled causal processes (e.g., shaking a present to find out what is inside) and can provide insight into the accuracy of that model in the same way as experimenting in science can provide insight into causal processes. Policies can then be inferred based on their expected epistemic utility, in the set of actions expected to maximise model updating (i.e., how much the existing model improves upon acting), and thereby minimise uncertainty (Friston et al., 2015, 2016). This reveals a prior expectation of agents, namely that they expect to act to occupy states where they minimise uncertainty. Action in this framework can also just be about changing the world such that it conforms with the agent's expected states, and thereby reduces prediction error (e.g., opening a present to obtain the reward inside) (Friston et al., 2015). As such, polices are *selected* by being estimated to be the most probable course of action to induce state transitions which minimise prediction error in one of these two ways. Notice that, for both types of action, the inferred policies provide information to the agent about what kind of agent they are; policies reveal how an agent is likely to act in different contexts, and thereby help the agent infer their own traits and develop a sense of identity and selfhood.

Active inference concretely relates to prediction error minimisation in the sense that the inferred policy predicts sensory input that is not actually occurring, which induces prediction error that will be minimised by moving the body. That is, bodily movement arises in the minimisation of prediction error generated from an inferred policy that specified some expected sensory input (servicing either epistemic value or utility). Under the predictive processing framework, this is the *only* explanation for action, and is proposed as an alternative to traditional motor command accounts (Adams, Shipp, & Friston, 2013).

All this implies that fidgeting arises as an inferred policy. However, under the active inference account, it is a challenge to understand why our inferred policy would be to fidget given that it doesn't seem likely to reduce prediction error in any straightforward, meaningful way.

2.2. Volatility

Agents subsist in the actual, noisy, changeable and uncertain world, which is characterized by dynamic changes in sensory input as causal factors at various spatiotemporal levels (including causes associated with the agent's own actions) interact with each other creating nonlinear and unexpected fluctuations in the sensory input. Overall, this means that the expected uncertainty changes dependent on context. Under a predictive processing framework, human brains represent such change as an estimate of *volatility* – the expected variability (or variance) in the dynamics of the states causing sensory input (i.e., changes to the standard deviation across contexts).

Volatility reflects an expectation about the rate at which our models change in perceptual inference, and how much they can change when acting for epistemic value. This expectation can be implicitly represented in the brain in terms of the overall uncertainty of what is being inferred together with the precision (or fidelity) with which states or actions generate outcomes. This would provide a dynamic estimate of how much information would be gained by a certain action. Intuitively, the distance with which the agent's beliefs move following an observation is the information gain, and this reduction in uncertainty corresponds to the degree to which prediction errors are minimised.

Some types of agents might have evolved explicit representation of this rate of prediction error minimisation, represented in the brain as a hyperprior – a deep hierarchical prior about the dynamic nature of the agent's own model. Technically, the rate at which a prediction error is minimised is partially determined by the precision (or confidence) ascribed to that prediction. Intuitively, we update our beliefs faster and more dramatically when we believe our data to be more reliable and when we suspect the world is frequently liable to change (for a more formal account of the dynamic effect of variability and volatility on learning rate, see Mathys, Daunizeau, Friston, and Stephan (2011)). This means that those creatures whose internal models allow them to make predictions about this precision implicitly hold (sub-personal) beliefs about the expected rate of error-minimisation, conditioned on the data they choose to sample. Given the inverse relationship between precision and ambiguity, we can associate beliefs about the ambiguity expected under a given policy with beliefs about the expected rate of uncertainty (or error) minimisation.

Expectations for volatility and rate of prediction error minimisation will impact on policy selection. Policies that will give the expected prediction error minimisation should be selected but this inference is sensitive to beliefs about the extent to which there will be underlying change in the world during execution of the policy. Actions which are effective in one context will likely be rendered less optimal when the statistics of the environment change, as various hidden causes both enter and leave the causal chain.

2.3. Hypothesis decay

The expectation for change and volatility will exert pressure on the current best evidenced hypothesis about the causes of sensory input, decreasing its strength with time. The longer one sticks to a winning hypothesis, in the presence of higher-level information about causal interactions, the less likely it is that this hypothesis will remain efficient at minimising prediction error (Hohwy, Paton, & Palmer, 2016, p. 320). In short, the strength of evidence for a particular hypothesis decays over time, contingent on the inferred volatility. A given prior, which was efficient to guide inference at some time may soon be less efficient as hidden causes in the environment change. This carries over to policy selection via the fidelity of the mapping from states and actions to outcomes, which is sensitive to underlying change in the causal landscape – as time passes the agent must begin to anticipate that ambiguity will increase, that is, that uncertainty will increase and rate of prediction error minimisation will decrease. As the mapping between actions and outcomes is anticipated to change under volatility and thus becomes more ambiguous with time, the precision of the estimation for the belief that "I can effectively minimise prediction error" decreases. This leads us to our discussion of self-evidencing.

2.4. Self-evidencing and self-models

Biological systems, under predictive processing, are embodied models of the statistical regularities in the world which resist entropy by occupying a limited number of possible homeostatic states, relative to which they minimise prediction error by active and perceptual inference (Friston, 2013; Friston & Ao, 2012). Minimising prediction error under a model is equivalent to maximising the evidence for that model. If the model is embodied as the agent, then the reduction in prediction error becomes evidence for the existence of the agent; this is called *self-evidencing* (Friston, 2010, 2013; Hohwy, 2016). This means that an agent will model its own existence implicitly in terms of a belief that it is self-evidencing. Meeting an expected rate of prediction error minimisation (or information gain) in perceptual and active inference essentially reassures the agent of its own existence.

In modelling the causes of sensory input, one relatively stable cause of changes to the sensed environment is oneself. As the agent's own body, via its actions, causes endogenously generated changes to sensory information, the model that best explains all of the agent's sensory experiences will include a model of the agent itself (Apps & Tsakiris, 2014); this is the *self-model.*

The self-model is hierarchical, spanning basic body parameters (limb size and reach), occurrent control states (beliefs about how desire for sip of coffee leads to reaching movements), habits (having coffee every morning), and character traits (introvert coffee connoisseur). Interactions among the self-models' elements help minimise prediction error caused by the agent's own actions (Hohwy & Michael, 2017).

The agent represents itself as a (complex) cause that is cyclically perturbed as the system interacts with the world and deals with the uncertainty inherent in spatially or temporally changing environments. This representation then is also implicated in policy selection, as the scope of possible intervention by the agent and the threshold for irreducible uncertainty is inferred through this model.

The self-model represents the internal states of the agent from which action outcomes are generated. Therefore, what the self-model represents is centrally implicated in the fidelity of the mapping from actions to outcomes, and thus for the expected reduction of uncertainty or rate of prediction error minimisation. This creates a conceptual link between the self-model ("what kind of cause in the world am I?") to existence ("are my actions self-evidencing?"). This link is captured in the ambiguity of the action-outcome mapping – a highly ambiguous mapping predicts poor self-evidencing. As already briefly discussed, some agents may be able to represent the rate of self-evidencing explicitly, as a belief capturing the expected rate of prediction error minimisation. Further, sensitivity to dynamic changes in the rate of prediction error minimisation may function similarly to explicit estimations of volatility, improving stability following volatile changes in statistical contexts and allowing quicker and more differentiated changes in response to volatility (Joffily & Coricelli, 2013).

In this section, we have reviewed relatively subtle aspects of the predictive processing framework, which we will recruit in order to meet the challenge of explaining why fidgeting policies should be inferred.

3. Fidgeting as a solution to self-model decay

Since the representation of self is central to the agent's model of the world, it too is sensitive to dynamic estimates of environmental volatility. The hypotheses generated from the self-model will then also have decaying evidence given the right (or wrong) circumstances. Self-hypothesis decay implies that the self-model is deprived of evidence, undermining the agent's confidence that it is minimising error at the expected rate – implicitly, its confidence that it is self-evidencing and will continue to exist. When there is hypothesis decay (contingent on expectations for volatility) about the very mapping between our actions and their consequences we will increasingly violate our expectation for the rate at which we can minimise prediction error. In that case, our current policy for interacting with the world no longer accurately predicts how much or how little prediction error it can minimise. This raises the question of what we should do when there are unexpected rates of prediction error minimisation, that is, when belief in self-evidencing is threatened.

In practice, it is likely that the actual rate of prediction error minimisation deviates from expected rates of prediction error minimisation (for the given context) during the extremes of psychological arousal. When the agent is bored, the available policies are not expected to have much epistemic value, even though the ambiguity is low – the model is already largely optimised for the environment. This can be interpreted as the model performing better than expected in terms of achieving states with high utility, eliminating prediction error faster than expected in the context, in spite of relative sparsity of sensory perturbations and enacted policies. Consider boredom during on-task behaviour such as when reading dull administrative emails. In this situation, the individual may have an expectation that opening a new email should produce some level of prediction error (i.e., verify a model with somewhat ambiguous mapping between action and sensory effect), however, on finding that they perfectly predict the content they will be bored – reducing prediction error faster than expected based on the modelled ambiguity. The modelled ambiguity has failed to adhere with the expected result (greater prediction error) of the most precise policy. Note that in overwhelming and overly complex situations one might also become bored, such as in a very complicated lecture. In these cases, the most precise policy (which is often to disengage from the inordinate task) reduces the total set of presumed model-able hidden causes, making the bound on estimated irreducible prediction error higher, but enabling the rate of prediction error minimisation to be more precisely estimated.

At the other extreme, when the agent is stressed, the model reduces uncertainty at a slower rate than expected, leaving the system with much unaccounted-for prediction error (see Peters, McEwen, and Friston (2017) for an account of stress in terms of irreducible uncertainty). In this scenario, the administrative email which was expected to have a small range of content may have contained disturbing news requiring a whole new set of policies and goal states for the day with increased complexity and ambiguity.

These deviations in expected rate of prediction error minimisation could come about for different reasons, such as in the case of boredom, when the world may be less complex (in terms of quantity of interacting hidden causes), less volatile and/or the agent is more efficacious than expected. In the case of stress, the world may be more complex, more volatile and the agent may be have less effective causal powers than expected (for reasons such as tiredness). It is in these situations that the self-model is most susceptible to the processes of hypothesis decay. They represent instances where the fidelity of the action-outcome mapping deteriorates in different ways. This is because there is increasing ambiguity in policies learned over a long history of perception and action, which have reciprocally constructed the inferred self-model. If your actions no longer do what they have always done, then this will motivate the inference that your self-model is itself subject to volatility. Because your understanding of what kind of agent you are in the world is based on beliefs about how you act in general and how those actions should affect prediction error minimisation, you begin to lose evidence that you exist (at least as the same kind of agent that you always have been).

This is the context in which, we argue, we should understand fidgeting. The problem, described above, was to explain why we might infer a fidgeting policy when such actions seem to provide no utility – agents obtain no obvious utility for fidgeting. The notion of active inference also allows action for epistemic value but it is also difficult to see how fidgeting could provide epistemic value, since it is not obviously a matter of exploring the world to reduce uncertainty about the model. These actions are precisely ones that the agent has performed many times before with unchanging results (compare eye movement to look more closely at something to determine what it is – a clear case of acting for epistemic value). Nevertheless, we propose that, in a setting where there is self-hypothesis decay, fidgeting can in fact be conceived as action for epistemic value.

The idea is that fidgeting reinforces the fidelity of the action-outcome mapping and thereby solidifies the agent's belief that it is efficiently self-evidencing. Our actions more or less reliably have certain expected consequences. This reliability is the precision of our inferred policy, which determines how efficiently sensory prediction errors are minimised. At the most basic level, even aimless moving will reliably increase the signal in proprioceptive inputs relative to the noise. So fidgeting should always reduce ambiguity, that is, speeding up the error minimisation process in relation to, for this particular example, the skeletomotor system. Perhaps, because this will always the case, this is the type of behaviour we default to if other behaviours cannot reduce uncertainty (either because there is no other relatively informative policy available in the context (i.e. boredom), or because the environment is so ambiguous that efficient policy selection is impossible (i.e. stress)). In this way, fidgeting brings the rate of prediction error minimisation back to what we expect it to be, while also changing the expected sensory input to include this repetitive self-stimulation. The proposal is thus that fidgeting is action for epistemic value in the foundational sense of self-evidencing. In other words, when I can't resolve uncertainty about anything else, I can still resolve uncertainty about myself.

The evidence provided by fidgeting is particularly strong and precise because fidgeting policies are relatively simple, involving few interacting hidden causes between the movement and the sensory effects of the movement. In this way, fidgeting polices are robust to volatile changes in the external environment. In the clearest examples, such as bouncing one's leg or playing with one's hair, the only hidden causes are part of the agent's own body. When other causes are involved, they are reliable and familiar, such as the sounds generated from clicking a pen. The particular ways an individual chooses to fidget will depend on their learned history about what fidgeting policies are most precise in each context, which will also be individually delimited based on learned cues to relevant similarities across environments.

The proposal captures several important features of fidgeting. Fidgeting is often self-stimulatory and reflexive, reflected in the tight causal loop and involvement of few hidden causes, with good robustness across environmental contexts. The stimulation is patterned because this increases the predictability within the current context and provides temporal rhythm that also increases the dimensions across which the stimuli are predictable. As a precise, expertly completed policy, fidgeting can be performed consciously or unconsciously but is usually initiated non-deliberatively because it does not fulfil person-level goals, but rather is an automatic way to reaffirm the existence of the entity to which person-level goals are attributed. The sensory modality over which fidgeting is completed does not matter for its self-evidencing role, so policies may span these domains within an individual (Table 1). It can involve external objects, such as clicking a pen,visually tracking a rotating fan or rocking on a chair; or be limited to touching or moving parts of ones' own body, for example bouncing one's leg, playing with one's hair or biting one's nails.

In some ways, this may seem similar to the bodily-regulation accounts of fidgeting reviewed earlier, in that fidgeting policies are a way to stay within a relatively small range of expected states. However, our proposal differs for two primary reasons. First, the states that cause the deviation from expected states need not be primarily bodily (homeostatic) states, but are related to cognitive states. So the prediction error being eliminated by fidgeting is not necessarily directly to do with states like heart rate and blood pressure. Second, the account allows for fidgeting in *anticipation* of prediction error, inferred from a deviation from the expected trend. In this way, where a classic homeostatic regulation account would place fidgeting as a response to prediction error, our account is more allostatic, in that fidgeting may occur before being in the unexpected bodily state (Corcoran & Hohwy, 2019). The prediction error it addresses is related to unexpected changes in the modelled transitions between states rather than the result of being in a particular state.

Even though fidget policies are very reliable and provide strong self-evidencing when executed, (neurotypical) agents do not over-indulge in fidgeting. This is because in the real, volatile world, a fidget-only strategy would accumulate prediction error in the long run, that is, there is hypothesis-decay even for the hypotheses leading to fidgeting. Concretely, feeding, exploring, and paying the bills are precluded by fidgeting, meaning that the agent eventually strays from its expected states. Self-evidencing agents will therefore dynamically recruit from a varied repertoire of policies. We have argued that fidgeting belongs in this repertoire.

The extent to which fidgeting policies should be inferred raises the question of maladaptive fidgeting, especially in psychiatric conditions. It is far from a trivial inference problem to continuously vary the inference of different types of policies from one's repertoire. Active inference requires a good sense for which kinds of expected states can actually be achieved – the world does not always co-operate with one's desires. Confident policy inference depends on quite high-level, abstract representations of the underlying statistics, at various time-scales, that might have bearing on the current context. Learning of this type may be compromised in some mental and developmental conditions, which could in some cases lead to increased inference of fidgeting policies. We turn to this topic next.

4. Fidgeting in psychiatric conditions: the case of autism

One of the diagnostic criteria for Autism Spectrum Conditions in the DSM-5 is "Stereotyped or repetitive motor movements, use of objects, or speech" (American Psychiatric Association, 2013). Autistic individuals' fidgeting is pathologised, as it is often done in a socially inappropriate way (for example larger, more noticeable, more unique actions), and seemingly happens more frequently. It is commonly called *stimming,* which is short for self-stimulation.

Our proposal can make sense of stimming in the autistic case. Previous research has shown that self-cognition may be different in autism (Frith & Happé, 1999; Hobson, 2011; Huang et al., 2017; Lombardo & Baron-Cohen, 2011; Molnar-Szakacs & Uddin, 2016; Perrykkad & Hohwy, 2019; Uddin, 2011; Williams, 2010). From the predictive processing perspective, research suggests that the autistic self-model has less hierarchical depth (Perrykkad & Hohwy, 2019), and autistic people have a high expectation for volatility (Lawson, Mathys, & Rees, 2017), meaning autistic people would accumulate uncertainty faster than neurotypicals. Relatively shallow models with high expectations for volatility would consider much sensory input as unexpected surprise, leading to high uncertainty overall. Autistic individuals could then be expected to recruit fidgeting policies more and for longer, in order to establish and maintain some fidelity of their action-outcome mapping and reduce uncertainty quickly. With a relatively shallow self-model, they will have fewer cognitive resources to model volatile changes in the sensory input, which means that fidgeting policies may remain attractive for longer, even as prediction error arises.

It might be revealing to note how stereotypies, as in the case of autism, differ from tics, as in the case of Tourette syndrome. The characteristics of stereotypies are more similar to the characteristics of fidgeting as we have defined it, than are the characteristics of tics. These differences are outlined nicely by Mills and Hedderly (2014) (in a table based on Barry et al. (2011)), such that stereotypies, concordantly with our approach to fidgeting, are "fixed, identical, foreseeable … rhythmic", whereas tics are "variable".

Additionally, while tics are often vehemently avoided, stereotypies are commonly described by autistic people as enjoyable. To our knowledge, the only other explicit functional account of the rate of prediction error minimisation is in determining emotional valence, such that, in general, a higher rate of prediction error minimisation is associated with positive affect (Joffily & Coricelli, 2013; Van de Cruys, 2017; Wilkinson, Deane, Nave, & Clark, 2019). In this way, switching to a policy with a reliably steep prediction error minimisation rate, as in fidgeting, should be enjoyable under a predictive processing explanation. In a similar vein, increased prevalence of anxiety in autism (van Steensel, Bögels, & Perrin, 2011) might also be explained by higher ambiguity for overall expected states which leads to "'faster and faster' increase in the violation of the expectations about … existential causes of sensations, eliciting fear at these levels" (Joffily & Coricelli, 2013, p. 12).

Along with what is traditionally considered autistic stimming, some further diagnostic criteria are also relevant here. Specifically, "insistence on sameness, inflexible adherence to routines, or ritualised patterns of verbal or nonverbal behaviour … highly restricted, fixated interests that are abnormal in intensity or focus … hyper- or hyporeactivity to sensory input or unusual interests in sensory aspects of the environment." (American Psychiatric Association, 2013). These activities can be conceived as self-evidencing behaviours that proactively (if unconsciously) construct an environmental niche in which modelling hidden causes is less complex (Constant, Bervoets, Hens, & Cruys, 2018). This is a longer-term way of building an environment in which the individual is more likely to have an unambiguous action-outcome mapping, meet their expected rate of prediction error minimisation and avoid the need to engage in short term policies for prediction error minimisation like stimming.

In some cases, fidgeting involves self-harm, including some of the more worrisome and clinically problematic versions of stimming in autism, such as head-banging or cutting. This may seem inconsistent with our proposal: if the fidgeting literally destroys the boundaries of the agent's own body, or damages the vehicle of this self-model, how can it be self-evidencing? We speculate that pain, in this case, may be functioning as a precise source of information about the self because the presence of acute pain is more certain than the presence of a touch or movement of a limb. In spite of their averse nature, actions that involve self-inflicting pain spring from policies with precise action-outcome mappings, which could explain why they are sometimes inferred. In a clinical context, this suggests that instead of focusing on just stopping the self-harm outright, a good strategy may be to replace it with a less harmful stim. The struggle for clinicians will be finding a stimming policy that yields equally precise self-evidencing, given the hierarchical structure of the self-model.

Using autism as a case study, we can then see how impaired fidgeting might arise. Fidgeting maximises reliability of the action at the expense of learning changing and complex environmental regularities. In this sense, it is like pushing Occam's razor too far. While fidgeting begets precise prediction error, it is too simplistic to capture much of the causal structure of the world.

For the case of autism, we think it is likely that some individuals have quite profound experiences relating to self-evidencing, as illustrated in this conversation:

> Mukhopadhyay: *Rules are formed by an Autistic person to simplify the ongoing uncertainty which is taking place around him. The uncertainty may lead the Autistic person to lose his identity. And because that would be a total chaotic situation, he tends to take the shelter of his rules, which he has created, choosing certain phenomenon from the greater uncertainty surrounding him. …*
>
> Biklen: *I would think that if a rule is known only to you, this could cause difficulty to those around you?*
>
> Mukhopadhyay: *Rules are somewhat the very proof to an Autistic person that he exists. He would have guidelines about these rules, which rule would be performed by him to the extremities of forming a rigid system of ritual. I am no exception and I get a sort of self existing sense when I have followed a routine set of activities.*
>
> (Tito Rajarshi Mukhopadhyay, who is autistic, speaks with ethnographer Biklen (2005, p. 126)).

5. Mental fidgeting?

Given that self-evidencing and minimising decay of the self-model relies on effective active inference, this raises a question about what happens in cases where an individual cannot move, such as in locked-in syndrome, paralysis, or imprisonment. In these cases, we would expect that the sense of self transforms or even fades. However, we also must consider mental actions such as imagining and thinking, subtle movements such as eye movements, and homeostatic regulation of bodily states for which the brain receives feedback, all of which might provide some self-evidencing. Mental fidgeting may include rumination or repetitive thoughts, or lapses of concentration and frequent sojourns into mindwandering. We would expect such processes to be more frequent in conditions of constricted or absent agency.

Temporary fading of the belief in bodily existence can reportedly be achieved by the practices of meditation or sensory deprivation. One interpretation of these cases is that they still involve self-hypothesis decay but that for some time period it is possible to maintain a belief that the body is dispersing, rather than choosing to act to re-confirm bodily existence. This may lead to common experiences of the dissolution of the ego boundary under these conditions (see also Letheby and Gerrans (2017)), as self-model decay is not actively resisted. These states can be pleasurably and convincingly upheld until bodily states create endogenous volatility outside of the agent's control – for example, the feeling of hunger. Conversely, in some disorders of the self where self-evidencing is compromised, a tranquil absence of action may be too hard to maintain, and we should see increased mental and bodily fidgeting.

Similarly, there may be cases where repetitive actions induce a trance-like state which has analogous ego-dissolving characteristics. In these cases, individuals deliberatively engage in fidget-like behaviours in order to *create* self-model decay. Here, these individuals are not naturally in a context of growing uncertainty, but are intentionally ignoring the complexities in the world and purposefully inducing longer-term prediction errors. While initially these actions might serve the usual fidgeting function of self-evidencing, when the usual drive to switch policies in order to occupy the most probable states for the agent (as discussed earlier, such as feeding, exploring, paying bills) is *intentionally* delayed, these actions begin to violate the expected rate of prediction error, and serve the opposite function. They make the world too simple/predictable for too long, which leads to ego-dissolution due to (in this case) intentional violation of the expected volatility.

6. Fidgeting in animals and babies?

On the predictive processing framework, self-evidencing is a necessary characteristic of any biological, active inference system. This means that any biological system will implicitly or explicitly represent the expected reduction in uncertainty from its mapping of actions to outcomes, or its expected prediction error minimisation. Some creatures may even have explicit representations of themselves, that is, have a sense of self. This implies that many if not all organisms will experience occasional deviations from their expected rate of self-evidencing. This in turn predicts that fidgeting should be observed in many species.

This prediction is modulated by details about the timescale over which such organisms expect to self-evidence and the depth of their hierarchical model such that the overall expected rate of prediction error minimisation will interact with how often a policy like fidgeting will be learned as likely to reduce uncertainty. Organisms with a higher threshold for expected irreducible prediction error and a slower expected rate of prediction error minimisation (e.g., due to a less complex model) may find fidgeting not a very useful policy, as it is too short term. However, it is conceivable that examples of animal fidgeting include a dog chasing its tail, an elephant rocking in a zoo and a Tasmanian devil pacing along circular tracks in its enclosure. Note too that the latter two cases are situations of reduced agency as discussed at the beginning of section 5.

Hommel (2017) suggests that newborn babies do not perform goal-directed actions, but rather start out with many *reflex* behaviours, which are slowly replaced by goal directed actions in development, during which time babies develop a sense of self (Verschoor & Hommel, 2017). In our framework, we conceive of these babies as self-evidencing from their first active inference, and with a marginal self-model at this point representing how they can reduce prediction error over time (see also Friston (2017a)). Their behaviours can then be likened to proto-fidgeting, that is, the first steps in setting up and exploring mappings between actions and outcomes.

7. Conclusion

We have provided a novel understanding of fidgeting as action for self-oriented epistemic value. Fidgeting leads to uncertainty reduction using a policy for action that involves only few hidden causes and which therefore furnishes a highly precise mapping of actions to outcomes. These policies are therefore rationally inferred in some contexts, despite being unfit for bringing about reward in a traditional sense. Fidgeting thus conforms with a form of self-evidencing, and helps the agent confirm its own existence in situations where evidence for its self-model might be waning. Impaired fidgeting can then be seen to arise for individuals who have compromised learning of expected levels of self-evidencing.

Acknowledgements

The authors would like to thank Thomas Parr for useful discussion and generous comments on this paper. On behalf of all authors, the corresponding author states that there is no conflict of interest. Kelsey Perrykkad is supported by an Australian Government Research Training Program (RTP) Scholarship. Jakob Hohwy is supported by the Australian Research Council (DP160102770, DP190101805).

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function, 218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®).* American Psychiatric Pub.

Andrade, J. (2010). What does doodling do? *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 24*(1), 100–106.

Apps, M. A., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews, 41*, 85–97. https://doi.org/10.1016/j.neubiorev.2013.01.029.

Barfield, E. T., Gerber, K. J., Zimmermann, K. S., Ressler, K. J., Parsons, R. G., & Gourley, S. L. (2017). Regulation of actions and habits by ventral hippocampal trkB and adolescent corticosteroid exposure. *PLoS Biology, 15*(11), e2003000.

Barry, S., Baird, G., Lascelles, K., Bunton, P., & Hedderly, T. (2011). Neurodevelopmental movement disorders–an update on childhood motor stereotypies. *Developmental Medicine and Child Neurology, 53*(11), 979–985.

Biklen, D. (2005). *Autism and the myth of the person alone.* NYU Press.

Carriere, J. S., Seli, P., & Smilek, D. (2013). Wandering in both mind and body: Individual differences in mind wandering and inattention predict fidgeting. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 67*(1), 19.

Cheshire, W. P. (2000). Hypotensive akathisia: Autonomic failure associated with leg fidgeting while sitting. *Neurology, 55*(12), 1923–1926.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind.* Oxford University Press.

Constant, A., Bervoets, J., Hens, K., & Cruys, S. V.d. (2018). *Precise worlds for certain minds: An ecological perspective on the relational self in autism. Topoi(The relational self - basic forms of self-awareness).* https://doi.org/10.1007/s11245-018-9546-4.

Corcoran, A. W., & Hohwy, J. (2019). Allostasis, interoception, and the free energy principle: Feeling our way forward. In M. Tsakiris, & H. De Preester (Eds.). *The interoceptive mind* (pp. 272–292). Oxford, UK: Oxford University Press.

Dayan, P. (2009). Goal-directed control and its antipodes. *Neural Networks, 22*(3), 213–219. https://doi.org/10.1016/j.neunet.2009.03.004.

Farley, J., Risko, E. F., & Kingstone, A. (2013). Everyday attention and lecture retention: The effects of time, fidgeting, and mind wandering. *Frontiers in Psychology, 4*, 619. https://doi.org/10.3389/fpsyg.2013.00619.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138.

Friston, K. (2013). Life as we know it. *Journal of The Royal Society Interface, 10*(86).

Friston, K. (2017a). Self-evidencing babies: Commentary on "Mentalizing homeostasis: The social origins of interoceptive inference" by Fotopoulou & Tsakiris. *Neuro-psychoanalysis, 19*(1), 43–47. https://doi.org/10.1080/15294145.2017.1295216.

Friston, K. (2017b). *The variational principles of action. Geometric and numerical foundations of movements*. Springer International Publishing207–235.

Friston, K., & Ao, P. (2012). Free energy, value, and attractors. *Computational and mathematical methods in medicine, 2012*.

Friston, K., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS One, 4*(7), e6421. https://doi.org/10.1371/journal.pone.0006421.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., J, O. D., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews, 68*, 862–879. https://doi.org/10.1016/j.neubiorev.2016.06.022.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation, 29*(1), 1–49. https://doi.org/10.1162/NECO_a_00912.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience, 6*(4), 187–214.

Frith, U., & Happé, F. (1999). Theory of mind and self-consciousness: What is it like to be autistic? *Mind & Language, 14*(1), 82–89.

Hobson, P. R. (2011). Autism and the self. In S. Gallagher (Ed.). *The oxford handbook of the self* (pp. 571–591). Oxford, UK: Oxford University Press.

Hohwy, J. (2013). *The predictive mind.* Oxford University Press.

Hohwy, J. (2016). The self-evidencing brain. *Noûs, 50*(2), 259–285.

Hohwy, J., & Michael, J. (2017). Why should any body have a self? In F. de Vignemont, & A. Alsmith (Eds.). *The body and the self, revisited.* MIT Press.

Hohwy, J., Paton, B., & Palmer, C. (2016). Distrusting the present. *Phenomenology and the Cognitive Sciences, 15*(3), 315–335. https://doi.org/10.1007/s11097-015-9439-6.

Hommel, B. (2015). *Action control and the sense of agency.* The sense of agency307–326.

Hommel, B. (2017). Goal-directed actions. *Handbook of Causal Reasoning,* 265–278.

Huang, A. X., Hughes, T. L., Sutton, L. R., Lawrence, M., Chen, X., Ji, Z., et al. (2017). Understanding the self in individuals with autism spectrum disorders (ASD): A review of literature. *Frontiers in Psychology, 8*(1422), https://doi.org/10.3389/fpsyg.2017.01422.

Joffily, M., & Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS Computational Biology, 9*(6), e1003094.

Johannsen, D. L., & Ravussin, E. (2008). Spontaneous physical activity: Relationship between fidgeting and body weight control. *Current Opinion in Endocrinology Diabetes and Obesity, 15*(5), 409–415.

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*. https://doi.org/10.1038/nn.4615.

Letheby, C., & Gerrans, P. (2017). Self unbound: Ego dissolution in psychedelic experience. *Neuroscience of Consciousness, 3*(1), https://doi.org/10.1093/nc/nix016 nix016-nix016.

Lis, S., Baer, N., Stein-en-Nosse, C., Gallhofer, B., Sammer, G., & Kirsch, P. (2010). Objective measurement of motor activity during cognitive performance in adults with attention-deficit/hyperactivity disorder. *Acta Psychiatrica Scandinavica, 122*(4), 285–294. https://doi.org/10.1111/j.1600-0447.2010.01549.x.

Lombardo, M. V., & Baron-Cohen, S. (2011). The role of the self in mindblindness in autism. *Consciousness and Cognition, 20*(1), 130–140.

Lovaas, I., Newsom, C., & Hickman, C. (1987). Self-stimulatory behaviour and perceptual reinforcement. *Journal of Applied Behavior Analysis, 20*(1), 45–68. https://doi.org/10.1901/jaba.1987.20-45.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience, 5*, 39. https://doi.org/10.3389/fnhum.2011.00039.

Mills, S., & Hedderly, T. (2014). A guide to childhood motor stereotypies, tic disorders and the tourette spectrum for the primary care practitioner. *Ulster Medical Journal, 83*(1), 22.

Molnar-Szakacs, I., & Uddin, L. Q. (2016). The self in autism. In M. Kyrios, R. Moulding,

G. Doron, S. S. Bhar, M. Nedeljkovic, & M. Mikulincer (Eds.). *The self in understanding and treating psychological disorders* (pp. 144–157). Cambridge, UK: Cambridge University Press.

Morishima, T., Restaino, R. M., Walsh, L. K., Kanaley, J. A., Fadel, P. J., & Padilla, J. (2016). Prolonged sitting-induced leg endothelial dysfunction is prevented by fidgeting. *American Journal of Physiology - Heart and Circulatory Physiology, 311*(1), H177–H182.

Morrens, M., Hulstijn, W., Lewi, P. J., De Hert, M., & Sabbe, B. G. C. (2006). Stereotypy in schizophrenia. *Schizophrenia Research, 84*(2), 397–404. https://doi.org/10.1016/j.schres.2006.01.024.

O'shaughnessy, B. (1980). *The will, Vol. 2*. Cambridge: Cambridge University Press.

Perrykkad, K., & Hohwy, J. (2019). Modelling Me, modelling you: The autistic self. *Review Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s40489-019-00173-y.

Peters, A., McEwen, B. S., & Friston, K. (2017). Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. *Progress in Neurobiology, 156*, 164–188. https://doi.org/10.1016/j.pneurobio.2017.05.004.

Skinner, B. F. (1948). Superstition' in the pigeon. *Journal of Experimental Psychology, 38*(2), 168–172. https://doi.org/10.1037/h0055873.

van Steensel, F. J. A., Bögels, S. M., & Perrin, S. (2011). Anxiety disorders in children and adolescents with autistic spectrum disorders: A Meta-analysis. *Clinical Child and Family Psychology Review, 14*(3), 302. https://doi.org/10.1007/s10567-011-0097-0.

Uddin, L. Q. (2011). The self in autism: An emerging view from neuroimaging. *Neurocase, 17*(3), 201–208. https://doi.org/10.1080/13554794.2010.509320.

Van de Cruys, S. (2017). Affective value in the predictive mind. In T. K. Metzinger, & W. Wiese (Eds.). *Philosophy and predictive processing*. Frankfurt am Main: MIND Group.

Verschoor, S. A., & Hommel, B. (2017). Self-By-Doing: The role of action for self-acquisition. *Social Cognition, 35*(2), 127–145. https://doi.org/10.1521/soco.2017.35.2.127.

Wilkinson, S., Deane, G., Nave, K., & Clark, A. (2019). Getting warmer: Predictive processing and the nature of emotion. In L. Candiotto (Ed.). *The value of emotions for knowledge* (pp. 101–119). Cham: Springer International Publishing.

Williams, D. (2010). Theory of own mind in autism: Evidence of a specific deficit in self-awareness? *Autism, 14*(5), 474–494. https://doi.org/10.1177/1362361310366314.

In this chapter, we encountered some more technical discussion using the predictive processing framework. I suggested that individuals will use prediction-error minimising policies (in this case, fidgeting) to respond to unexpected uncertainty in action-outcome mappings. This theoretical hypothesis is borne out in the following experimental chapters, where I demonstrate that participants do in fact respond to rising prediction error by using policies in a way that is effective in prediction error minimisation. This paper also suggested an important role for environmental volatility. In all of the coming experiments, I independently manipulate the uncertainty in the environment, either by randomly changing the experienced variability (within trials in Chapter 5, within blocks in Chapter 6) or by providing participants with environments comprising distinct regimes of uncertainty which they can freely move between (Chapter 7). I will now introduce these chapters collectively, and how they speak to these notions and the ideas presented in the earlier parts of the thesis.

Preface to The Squares Task Trio: Chapters 5-7

The following three chapters focus on judgement of agency as a sub-category of self-cognitive processes (Chapter 1: *action>judgement of agency*). In Chapter 1, we saw that previous experiments investigating judgement of agency in autistic populations have shown no difference in performance to neurotypical populations. Results from Chapter 3 redirected the course of research from trying to understand the entirety of the self in autism to focusing on just one domain. In Chapter 4 (and littered throughout Chapters 1 and 2), we saw that action might be a particularly interesting avenue when applying the tools of predictive processing to self-cognition that is relatively understudied in autism.

A judgement of agency is equivalent to an explicit “I did that” response. Paradigms investigating judgement of agency usually have a participant perform an action (or not), observe some sensory consequence of the action (often expected or unexpected), and then report whether or not the consequence was a result of their action. This is conceptually and empirically distinct from the sense of agency (Chapter 1: *action>sense of agency*), which is the feeling that the action was self-caused. However, it is generally thought that in the usual case a sense of agency informs judgements of agency (borne out in Chapter 7).

In this preface, I will provide further justification for the particular focus on action to understand the self in autism. I will also provide an overview of the similarities and differences in the experiments reported in the following three chapters.

Why action is important to self-cognition

If we accept that the self-model is built through an inferential process, then this inference must be made on the basis of some data. In a folk psychological sense, when we think about what kind of person we are, or describe our*selves*, we usually describe general character traits. These include various abstract properties; for instance, about the way we choose to interact with others (“I am kind”), the roles we fill (“I am an interdisciplinary scientist”), observations about the way we are received (“I am well-respected”), the valence of our reactions to certain stimuli (“I like chocolate”), the effectiveness of our goal-directed actions in various contexts ("I am a poor basketball player"), our likely moods ("I am easily frustrated"), and so on. These can all be reconceptualised as expectations about the likely states of our bodies and our preferred policies based on a past history of occupied states and the policies that got us there. What grounds my inference that I am a self with the property of kindness? A history of performing kind actions. What grounds my inference that liking chocolate is part of my self-model? Many times when I have previously eaten chocolate (an action), I have enjoyed it (an expected outcome) and this sets up the expectation that I am the kind of self that will both choose to perform that action again (policy), to successfully occupy a likely future state of my organism. The data that informs our inferences and therefore the selected model of the self is learned from a history of active exploration of the environment we inhabit and observations of the consequences of these actions. It is also plausible that the truth about ourselves is also grounded in our actions – if I model myself as kind, but do not act in kind ways, then I am mis-representing myself to myself.

Under an active inference account, our selected model also determines what actions we take. In the previous chapter, we saw that the construction and maintenance of the self-model is done in a circular loop with data from the outside world. This is part of the self-fulfilling prophecy nature of active inference. When I choose kind actions, it is because that is my preferred policy – it leads to the most likely next state of my organism. In other words, I do kind things because I am a kind person and I am a kind person because I do kind things (self-evidencing). Of course, some of our self-model that guides action is subconscious, so this will not always be one's conscious rationale for the action. We act, perceive the consequences, update our model, and act again based on that model. Rinse and repeat. This is the action-perception loop.

The point here is that the action-perception loop is integral for building and maintaining the self-model. Self-representations are dynamic. Recall too from Chapter 1 that the self in predictive processing can be understood either as a hidden cause of sensory input (through observation of consequences consistently attributable to agent-originating actions) or the model of the model (including its active elements). Therefore, investigating the self-model in experimental paradigms that close the action-perception loop captures many of the most essential aspects of the self from both predictive processing and the folk psychological notion.

Judgement of agency as spanning the neural hierarchy

In Chapter 3, I tested the self at multiple levels of the neural hierarchy, with the shape-label matching task yielding a primarily low-level perceptual effect, and the self-concept questionnaires giving us an explicit, integrated, higher-level conceptual difference in self-representation. So where, then, does the judgement of agency fall in this neural hierarchy?

When operationalised for scientific study, both judgements of agency and sense of agency are usually measured in response to very particular sensory events, and when predictability of the sensory event is involved, the properties that are predicted are usually very low-level sensory properties (e.g. pitch of a tone). In this sense, cognitive processes underlying agency might be based in very low-level sensory processing. On the other hand, judgements of agency in particular sometimes involve conscious reasoning and integrating beliefs about causation (including high level beliefs about the self as cause). For instance, think about the case where you are stumbling around in the dark to get a glass of water in the middle of the night. You hear a crashing noise as you hit something with your foot. You may, in the moment, feel a sense of agency over the crashing noise. However, after an instant of reflection, you realise that the sound came from the bins outside, and judge that agency should rather be attributed to a possum. This involved early processing of the unpredicted sensory information and higher-order integration with other beliefs and contextual information to accurately judge agency. Findings from intentional binding paradigms in EEG (Chapter 1: *action>sense of agency*) also give us reason to think that both early sensory processing, as measured by attenuation of the N100 component, and later, more integrated processes involving contextual factors influencing predictability, as in the P3b component, may be involved in determining agency (Bednark, Poonian, Palghat, McFadyen, & Cunnington, 2015; Poonian, McFadyen, Ogden, & Cunnington, 2015).

As such, judgement of agency seems a suitable arena to target differences in self-cognition that span the cognitive hierarchy.

Action and autism

It may be of interest here to note too that many autistic individuals show differences in motor processes generally (Fournier, Hass, Naik, Lodha, & Cauraugh, 2010; Torres & Donnellan, 2015). These processes include gait, balance, coordination of locomotor skills (e.g. running and jumping), slower hand and foot movements, poorer manual dexterity, impairment in coordinating antagonistic movements in quick succession, poor movement planning and low muscle tone (Fournier et al., 2010; Gowen & Hamilton, 2013). Dziuk et al. (2007) demonstrated that dyspraxia (impaired performance of skilled motor gestures) in autism was associated with but not reducible to basic motor skills deficits. These findings imply that motor ability in general seems to be affected in autism and indeed correlates with symptom severity. There is also evidence of greater noise and jerk in autistic reaching movements compared to neurotypical peers (Palmer, Paton, Kirkovski, Enticott, & Hohwy, 2015; Torres et al., 2013). While not currently included in the diagnostic criteria, based on a meta-analysis, Fournier et al. (2010) suggest that motor impairments could be added to the core features of autism, given that they are widespread and have a large effect size when comparing autistic and neurotypical performance. This often has effects on quality of life, for example, on early academic performance when it impacts the ability to handwrite (Verma & Lahiri, 2021). In many cases, autistic individuals may employ compensatory mechanisms or modified processes to perform unimpaired movements (Gowen & Hamilton, 2013).

From an active inference perspective, this wealth of evidence implies that there may be basic differences in the precision of the models and the prediction error that influences how individuals with autism move. As such, the evidence from autistic populations discussed in this section support the idea that autistic participants may show differences in action-oriented self-modelling that depend specifically on inferences and deployment of policies and uncertainty.

In the last chapter, I discussed the core feature of restricted and repetitive behaviours, a subset of which are the ‘stimming’ behaviours. This suggests that autism may be associated with the frequent use of self-evidencing policies as a ready solution to rising uncertainty. The following three chapters using the judgement of agency task will provide further empirical evidence for this claim.

The importance of uncertainty

In both Chapters 1 and 4, I emphasised the importance of the precision of expectations and how these reflect environmental uncertainty. As a reminder, uncertainty is the inverse of precision, and can be modelled at multiple levels. At the most basic level, variability represents the width of the distribution of expected inputs. At the next level up, volatility represents how frequently the variability distribution changes. In scenarios with high volatility, the environment is highly unstable. On the other hand, in low volatility environments, the variability is stable and the system can rely on relatively unchanging priors about variability.

A brief comparison of the Squares Task experiments in the thesis

In the following three tasks, I manipulate environmental uncertainty in a judgement of agency task. In this task, based on *The Squares Task* (Grainger et al., 2014; Russell & Hill, 2001; Williams & Happé, 2009), participants move squares on the screen using the mouse, and must identify which, if any, of the squares they control. By selecting a particular square, the participants are making a judgement of agency. When they select that they did not control any of the squares, they are making a judgement of no agency.

This task is the most frequently used judgement of agency paradigm in the autistic population (see Chapter 1). While previous studies found no differences between autistic and neurotypical participants in this task, previous versions did not include environmental uncertainty, nor did they look at fine grained action selection and policies employed to complete the task.

This task closes the action-perception loop – participants can freely move in the environment and dynamically respond to the ongoing stream of sensory information that results directly from their movement. In the series of experiments that follow, for the first time, I measure particular policies used by participants, and add structured and unstructured, changing and unchanging, environmental variability to see if these particular features of the loop inspired by predictive processing give us a better insight into how autistic individuals judge agency. I also developed a way of measuring a moment-to-moment behavioural proxy for prediction error using eye-tracking data (and button presses for Chapter 7), which allows us to test some of the hypotheses from predictive processing more directly.

For a tabular summary of some of the similarities and differences across the three studies, see Table 1.

In the first version of this experiment, presented in the next chapter (Chapter 5), I manipulate variability and volatility and measure its effect on movement and strategy as well as prediction error over time. In this version, in each trial, the distribution of experienced variability changes either a few times or many times depending on the level of volatility. In this study, I found broad effects of uncertainty on action selection policies and on the dynamics of prediction error. Not least of which, I show that participants with more autism traits appear to switch hypotheses about which square they control more readily in the face of prediction error as compared to those with fewer autistic traits. I also show differences in acceleration, time spent moving, dominant policy use, prediction error in different levels of variability and prediction error minimisation as it relates to agency judgements.

Volatility, as it was manipulated in Chapter 5, did not, however, show very many main effects across our dependent variables. As such, in Chapter 6, I aimed to increase the effect of volatility by increasing the amount of time over which volatility was manipulated. Perhaps the lack of effect in Chapter 5 was due to the brain interpreting the changes in variability at such short time scales to be merely variability, rather than modelling the volatility at the higher level. In Chapter 6, I manipulated volatility over blocks instead. In half the blocks, the variability was the same across all trials in that block. These were stable blocks, which had low volatility. In the other half, half the trials had high variability and half had low, in a randomly presented order. These blocks were unstable, and thus had high volatility. The completion of this study was halted by the COVID-19 pandemic in 2020-2021, and so a pilot sample of twelve final participants is included for this thesis. Many of the early results I present here replicate the findings of Chapter 5, and suggest that the changed volatility manipulation was successful in bringing out the effects of volatility on the dependent variables investigated. While data from twelve neurotypical participants with measured autistic traits is not enough to draw conclusions about the autism spectrum, future directions for analysis are discussed.

The final study in this trio is presented in Chapter 7. In this chapter, I gave control over the environmental uncertainty to the participants, and investigated how they would use changing environments as a policy. It is called the Beach task because I called the two environments participants could pick between ‘sand’ and ‘water’. While the two environments differ in the structure of their variability, it seems incorrect to attribute the difference to volatility explicitly since both are highly predictable and unchanging given the right model. In the ‘sand’ environment, the variability is random and unstructured. In the ‘water’ environment, periodic waves limit the experienced variability to the left or the right of the square’s heading. In this way, I could use this paradigm to investigate how participants weigh up model complexity (additionally modelling the waves in the ‘water’) and model fit (wider expected variability at each time point in the ‘sand’). Findings show that participants do have a significant preference to the sand environment, despite the higher irreducible uncertainty, which is associated with greater accuracy and confidence. I also show that participants with more autistic traits choose to switch environments earlier in response to rising uncertainty, in a way that mimics the findings about use of the hypothesis switch policy in Chapter 5.

In all three studies I report correlations with AQ in a community sample and do not compare a diagnosed autistic population and a neurotypical population. There are limitations to this of course, and we must be careful in how we interpret the results with respect to the self in autism. This will be the focus of Chapter 8. At the very least, this collection of studies provides promising avenues for future research in diagnosed populations.

Table 1 – *Comparison of Squares Task Design Across Chapters 5-7*

Experimental Feature	**Chapter 5:** *The Effect of Prediction Error in the Action Perception Loop*	**Chapter 6:** *Judgements of Agency and Block-wise Volatility: A Pilot Study*	**Chapter 7:** *The Beach Task: Environmental Niche Selection Under Uncertainty*
Variability Manipulation	**Within Trial**	**Within Block**	**Within Trial**
Location of Testing	**In Person**	**In Person**	**Online**
Participant-Terminated Trial Length		**X**	
Coloured Squares (vs. B&W pattern)	**X**		**X**
Luminance Matched Squares		**X**	**X**
Number of Squares per Trial	**8**	**8**	**4**
Eye-Tracking Hypothesis Estimate	**X**	**X**	
Pre-trial Fixation Cross		**X**	**X**
Measure Sense and Judgement of Agency		**X**	**X**
% No-Control Trials	**11%**	**25%**	**50%**

Chapter 5. The Effect of Uncertainty on Prediction Error in the Action-Perception Loop

The published work in this chapter reports the first of the squares task experiments in the thesis. In this study, I manipulate variability and volatility in the mapping between actions and sensory outcomes within a trial. I measured the effects of this uncertainty on policies (both action selection and broader strategy use), hypothesis selection as a particular policy of interest measured by eye position, and behavioural prediction error (that is, not taking into account changing priors in different conditions).

Cognition 210 (2021) 104598

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

The effect of uncertainty on prediction error in the action perception loop

Kelsey Perrykkad[a,*], Rebecca P. Lawson[b], Sharna Jamadar[c], Jakob Hohwy[a]

[a] Cognition and Philosophy Lab, Philosophy Department, School of Philosophy, Historical and International Studies, Monash University, Clayton, Australia
[b] Department of Psychology, University of Cambridge, United Kingdom
[c] Turner Institute for Brain and Mental Health, Monash University, Clayton, Australia

ARTICLE INFO

Keywords:
Agency
Prediction error
Action-perception loop
Uncertainty
Autism traits

ABSTRACT

Among all their sensations, agents need to distinguish between those caused by themselves and those caused by external causes. The ability to infer agency is particularly challenging under conditions of uncertainty. Within the predictive processing framework, this should happen through active control of prediction error that closes the action-perception loop. Here we use a novel, temporally-sensitive, behavioural proxy for prediction error to show that it is minimised most quickly when volatility is high and when participants report agency, regardless of the accuracy of the judgement. We demonstrate broad effects of uncertainty on accuracy of agency judgements, movement, policy selection, and hypothesis switching. Measuring autism traits, we find differences in policy selection, sensitivity to uncertainty and hypothesis switching despite no difference in overall accuracy.

A significant challenge to an agent's perceptual and decision-making processes is to distinguish between sensations that it can control, and those out of its control. For example, imagine you are working on your computer and it beeps. How do you know if you caused it, as opposed to a colleague emailing you? Influential theoretical work on predictive processing and active inference suggests that the brain relies on prediction errors to assess and test hypotheses about agency (Friston et al., 2013), but empirical evidence for this suggestion is lacking.

Inferring the relations between actions and their sensory consequences is riddled with uncertainty due to the complexities involved in deconstructing sensory evidence from the non-linear confluence of hidden causes. Sometimes when you click, the ensuing beep occurs later because the computer is updating its virus-software; other times, it happens straight away. The brain must represent this uncertainty at numerous hierarchical levels to identify when it is appropriate to attribute agency to oneself. In this example, the breadth of the distribution representing how long it takes for the beep to occur is the *variability* and the frequency of the virus-updates is the *volatility* (how often does the variability distribution change). Crucially, we do not yet know how this uncertainty changes ongoing decisions about *which* actions to perform when trying to explore and infer agency; thus, we have yet to explore how agents close the action-perception loop under uncertainty.

A *judgement of agency* is the verdict that the agent was herself the source of a sensory event – the conscious "I did that" response. It is often (but not always) based on a *sense of agency* (or a feeling of authorship) during the movement. Predictability is often investigated in sense and judgement of agency paradigms by manipulating whether or not the identity (Bednark, Poonian, Palghat, McFadyen, & Cunnington, 2015; Engbert & Wohlschlager, 2007; Hughes, Desantis, & Waszak, 2013; Kuhn et al., 2011; Majchrowicz & Wierzchoń, 2018), timing (Hughes et al., 2013; Majchrowicz & Wierzchoń, 2018) and/or presence (Moore & Haggard, 2008) of a sensory outcome meets some prediction set up by the block-wise probability of each outcome. However, very few studies consider a more continuous distribution of deviations from the expected outcome (e.g., Zalla, Miele, Leboyer, and Metcalfe (2015)) and, to our knowledge, no previous studies have considered volatility (changes to such a distribution) in an agency paradigm.

In classic agency experiments, there are so few actions available to participants that action-selection strategies (or *policies*) cannot easily change in response to changes in prediction error or uncertainty. In some designs, such as Desantis, Hughes, and Waszak (2012), specific actions trigger specific outcomes, but the participants are instructed to equally perform each action. This does not allow participants to explore and attempt to optimally vary policy-selection. In other studies, participants do have freedom to change strategy, and have online action outcome mismatches, but the dependent variables are not sensitive to these strategies and so the temporal dynamics of online decisions with respect to this error are unknown (Zama, Takahashi, & Shimada, 2017). This gap in knowledge is crucial for understanding how we distinguish self-generated and externally-caused sensations in the real world. The

* Corresponding author.
E-mail address: kelsey.perrykkad@monash.edu (K. Perrykkad).

https://doi.org/10.1016/j.cognition.2021.104598
Received 9 July 2020; Received in revised form 7 January 2021; Accepted 8 January 2021
Available online 23 January 2021

current study sought to close this gap using a novel judgement of agency task that dynamically closed the action-perception loop while independently manipulating variability and volatility.

To understand these missing components in the process of inferring agency, we turn to recent accounts of agency from predictive processing - an explanatory framework whose fundamental claim is that the brain's function is to minimise the long-term average error between its expected and actual sensory input (*prediction error*) (Clark, 2015; Friston, 2010; Hohwy, 2013). In doing so it reduces uncertainty by refining models of the hidden causes of sensory input in the environment and in the agent itself.

Prediction error can be minimised by updating expectations while passively receiving sensory input (*perceptual inference; perception*). Another way to minimise prediction error is through action, by selectively sampling sensory input to satisfy beliefs about sensory input in future states of the world and the agent's own body, given certain actions (*active inference; action*) (Friston, 2017). Previous agency research has focused on perceptual inference in the context of agency, and has not interrogated the ongoing process of active inference.

Under an active inference account, agency attribution would occur by minimising the divergence between the predicted outcomes of available policies for action and the most probable future sensory states; in other words, when there is a belief that goals can be reached from the agent's current state (Friston et al., 2013; Friston, Samothrakis, & Montague, 2012; Hohwy, 2015). Thus, precision (i.e., the inverse of uncertainty) of these inferences is important (Friston et al., 2013) and lead us to investigate the effect of such variability on actions, prediction error and inferred agency.

According to active inference, the very purpose of action is then to minimise expected prediction error. To understand how this plays out in the action-perception loop it is then essential to reveal the interplay between action selection and the magnitude of prediction error at a given time, under a given policy. For the critical case of agency attribution, it is not known how an agent infers policies that may help reduce uncertainty about agency; this is mainly because thus far its magnitude has been under the control of experimenters, not participants themselves. Here, rather than dictating the magnitude of prediction error and measuring effects on behaviour and neural processes, we instead measure the prediction error itself and allow participants to control it with their actions.

The most straightforward expectation for active interrogation in an action-perception loop is that, where possible, policies are inferred which minimise prediction error. Part of the difficulty in testing this prediction is finding an appropriate way to measure prediction error. Here, we operationalise prediction error using eye position to calculate the evolving divergence between hand-movement and stimulus trajectories. Eye-tracking indicates moment-to-moment beliefs about agency which can be tested by mouse-movement. We predict that variability and volatility will have independent effects on movement patterns and policy selection, as well as on prediction error minimisation and subsequent judgements of agency. Specifically, high variability allows less precise representation of control states, which predicts more repetitive policy selections (Perrykkad & Hohwy, 2020a), more prediction error and less accurate judgements of agency. High volatility suggests potentially discoverable interfering hidden causes, predicting more policy exploration and more variance in prediction error which could aid accurate inference of agency. Independent of accuracy, we expect a positive correlation between agency-driven prediction error minimisation and judgements of agency, partly based on active inference theory and partly on prior literature on the role of prediction-expectation mismatch for agency reports.

It is instructive to consider how prediction error minimisation might differ in clinical or subclinical populations because such comparisons help reveal how the prediction error mechanisms work. We focus here on predictive processing accounts of autism, according to which autistic individuals have difficulty abstracting causal rules to higher statistical levels, and thus classify more uncertainty as irreducible. This has been theorised to be due to weaker priors, weightier prediction errors or hyper-flexible estimates of volatility, which all result in a higher learning rate in autism (for review and details, see Palmer, Lawson, & Hohwy, 2017). Manipulation of uncertainty in tasks that rely on perceptual inference has been shown to change performance in autistic populations (Lawson, Mathys, & Rees, 2017). Characteristic differences in action in autism, such as restricted and repetitive behaviours, may indicate differences in active inference in variable environments (Palmer, Lawson, & Hohwy, 2017). Previous research, not framed in terms of predictive processing, has used basic versions of the task we use here, and found no difference between groups of autistic and non-autistic participants (Grainger, Williams, & Lind, 2014; Russell & Hill, 2001; Williams & Happé, 2009), however, we predict the relationship between autism traits and agency attribution should be specific to interactions with uncertainty in the environment as the action-perception loop is dynamically closed (cf. Zalla et al. (2015)). This in turn speaks to underexplored topics in autism research relating to the sense of self and agency (Perrykkad & Hohwy, 2020b). Hence, here we additionally measured autistic traits in our sample and we predict that uncertainty will differently affect policies for movement and prediction error minimisation for participants along this scale.

1. Methods

1.1. Participants

Fifty neurotypical adult participants were recruited. Ten participants were excluded: five participants were removed for technical errors in recording, two for poor quality eye-tracking data (>35% lost trials) and three for poor accuracy (<45%). The final sample of 40 participants were primarily undergraduate students (55%, the remainder had completed tertiary degrees) with an overall mean age of 22.8 years (SD: 3.65, range: 18–34) and included 24 female participants. None of the participants reported neurological conditions, taking medications which affect cognition, nor a history of drug abuse. One participant reported a diagnosis of depression, and one of ADHD, removing these participants did not affect the primary results of interest (see supplementary materials). Two participants reported previously suffering a blow to the head that rendered them unconscious. All participants were fluent in English, had normal or corrected-to-normal vision and 95% were right handed. This study was approved by Monash University Human Research Ethics Committee (Project Number 11396). The experiment was conducted in accordance with the relevant guidelines and regulations, and all participants signed informed consent documents upon commencing the protocol.

1.2. Autism quotient

None of the participants were previously diagnosed with Autism Spectrum Disorder or its nominal variants. All participants completed the Autism Quotient questionnaire (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) to quantify autistic traits. The mean AQ score was 21.43 (SD: 5.89, range: 12–38).

1.3. Experimental task design and procedure

For a schematic diagram of the experimental set up, task and experimental manipulation, see Fig. 1 and a video of the task is available at https://figshare.com/s/fd2742b897e21d901dd0 (DOI: 10.26180/5eabbfb9a8aa4).

Testing was conducted in a quiet, darkened room. Participants were seated at a table with a chin rest set to a comfortable height, 84 cm from the screen, and approximately 55 cm from eye to eye-tracking camera. The task was completed using a computer mouse in the participant's dominant hand which was hidden in a curtained box (base dimensions: 32 cm wide x 30 cm deep). Their opposite hand gave judgement of

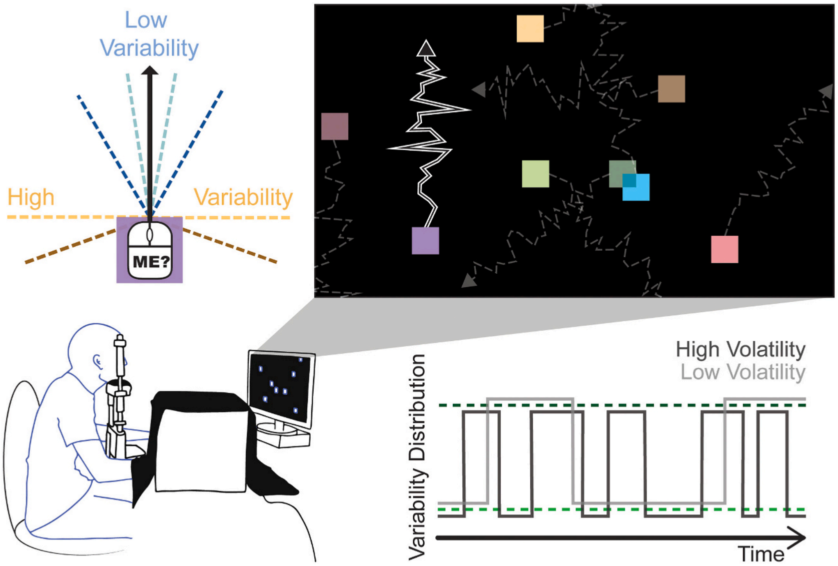


Fig. 1. Task and manipulation.
The participants' task involved using a hidden mouse to control eight squares on the screen. The mean of the target square's movement was the participants' movement, and distractor squares moved at a random angle offset from mouse movement. Jitter was added to the motion of all squares depending on the condition. In *low variability* blocks, the 95% confidence interval from which jitter was sampled switched between ±10° (light blue) and ± 30° (dark blue), and for *high variability,* it switched between ±90° (yellow) and ± 110° (brown) around the input movement or offset. In *low volatility* the distribution changed three random times per trial (light grey), and in *high volatility* it changed ten times (dark grey). See also https://figshare.com/s/fd2742b897e21d901dd0 (DOI: 10.26180/5eabbfb9a8aa4) for a video demonstration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

agency responses using the numbers on a keyboard. Participants had self-timed breaks between blocks.

We implemented a variant of *the Squares Task* (Grainger et al., 2014; Russell & Hill, 2001; Williams & Happé, 2009), presented using Psychtoolbox-3.0.14 version beta in Matlab 2017b (Mathworks, Natick, Massachusetts) on a 1920 × 1080 screen (60 Hz refresh rate). Eight randomly coloured squares (100px^2) appeared in an array at the beginning of each trial. All the squares moved when the mouse was moved and all the squares stopped when the mouse stopped, so participants had to move in order to accurately complete the task. Participants were given 15 s to identify the target square which they controlled. Distracter squares moved at a random angle offset from the vector of mouse movement, and this angle was also independently and randomly changed (and smoothly transitioned) five times in each trial. This means that each distracter square appeared to turn five times when the participant did not initiate a turn, breaking any illusion of control resulting from motor adaptation. Other than these turns, because the distracter squares took mouse input as part of determining their trajectory (mouse movement + angular offset + variability), the structural features of the motion of the target square (mouse movement + variability) and distracter squares were identical. There were also less frequent *no-control* trials in which all the eight squares were distracter squares. After the 15 s, all squares froze and were numbered, and prompted an unspeeded numerical response from participants indicating which square they controlled or '0' if they thought they controlled none of the squares.

There were four uncertainty conditions in a 2 × 2 design (variability × volatility). Some jitter was added to all squares (*variability*), such that depending on the condition, there was a range (95% CI) of random noise around the mean angle input by the mouse (or the mouse angle + distractor offset for distractors). This specified range also changed throughout the trial; the number of these changes was specified by the *volatility*. In the *low variability* condition, the distribution switched between a 10° and 30° 95% confidence interval on either side of the mean, and for *high variability,* it switched between 90° and 110°. The volatility manipulation decided how frequently the variability changed between these two distributions. In the *low volatility* condition, the variability changed three times, while in *high volatility*, there were 10 changes (pseudo-randomly timed with at least 50 frames between). Each trial's starting distribution was randomly selected. So, for example, in blocks of the *low variability, low volatility* condition, if the distribution started with 10°, the target square was initially jittering within a 95% confidence interval spanning 10° either side of the input mouse angle, and this distribution randomly changed three times during the trial (so widened to ±30°, then narrowed to ±10°, then back to ±30°) with a minimum of 833 ms between these changes in variability (see Fig. 1). There were two blocks of each condition (variability-volatility pair) with 18 trials per block (16 agentive trials, 2 no-control trials) and block order was randomized for each participant.

Prior to completing the task blocks, participants engaged in an interactive instruction demonstration. During the instruction period, participants were given control of a square and the features of the experiment were slowly introduced with text explanations. For instance, on the first page of the instructions, participants were shown one square with no variability added, and the text read: "This is you. Try moving around the screen…". On subsequent screens features such as wrapping around the screen edge, random colours between trials, variability ("Sometimes, the square will not perfectly match your movements, but they are pretty close! Try this…"), distracter squares, and the typical trial structure ("In each trial, you will have 15s to determine which square is you. The colours and starting positions are random, so do not rely on them! You only get information about which one is you by moving around the screen.") including how to respond were slowly introduced. Participants then completed a practice block containing sixteen total trials consisting of all trial types, which was excluded from all analyses. Participants received feedback following practice trials and summative feedback following the practice block. No feedback was given in the main task.

At the end of the experiment there was a short motor control task. In this task, participants were asked to move a perfectly controllable square along a white path as fast and as accurately as possible. There were 10 predesignated paths ranging in length and complexity. This task allowed us to quantify participants' ability to execute motor intentions.

1.4. Analysis

1.4.1. Behaviour

Performance on the motor control task was summarised by multiplying average area traversed outside the white path by average reaction time. This index accounts for the speed-accuracy trade-off, where low scores indicate better motor performance.

In the 'squares' task, *accurate* trials were those in which participants either correctly identified the target square, or correctly identified that there was no such square (*no-control* trials). Accuracy was the primary measure of overall task performance.

The *time spent moving* on each trial was calculated in seconds. This served as a proxy for environmental sampling, as participants were given freedom to start and stop moving as they pleased though only got task-relevant information by moving.

The *speed* of movement was calculated as the average pixels moved per frame, *acceleration* as change in speed per frame, and *jerk* as the change in acceleration per frame. Derivatives to the level of jerk were analysed to investigate the minimum jerk hypothesis of motor control (Wolpert, 1997) and for its possible relationship to movement trajectories in autism and its traits (Palmer, Paton, Hohwy, & Enticott, 2013; Palmer, Paton, Kirkovski, Enticott, & Hohwy, 2015).

On each frame, the participant's angle of motion was discretised into one of eight cardinal directions. These were plotted for visual inspection. Participants were found to primarily move in the cardinal directions (up, down, left, right), with smaller peaks at the diagonal midpoints. These plots, in combination with observation of trial replays, informed subsequent policy definition. A *turn* was defined as any change in direction which was preceded by at least three frames of one direction and sustained for at least three frames. More than simply sampling, which also occurred in straight movements, turning involves participant induced intervention on expected stimuli direction. These turns were further grouped into types, which were taken to indicate the participant's policy. These are pictorially and algorithmically defined in the Supplementary Materials. In brief, six policy types were identified: 1) Horizontal, 2) Vertical, 3) Perpendicular-Cardinal, 4) Non-Cardinal, 5) Hesitant-straight and 6) Circle. Note that rounded corners and circles were redefined for analysis as one turn each as they are taken as a unified intent of intervention by the participant.

While none of these policies has an a priori advantage over any other for task performance, we were interested in how flexible each participant was in switching between policies. For each policy, we created a mean percentage of turns that were of that type, across all conditions. We defined a participants' *dominant policy* as the policy which had the highest percentage of turns across the entire experiment. This allowed us to look at the number of turns in each trial which were of the participants' dominant policy as compared to alternative policies, as a proxy for exploratory behaviour (i.e., more dominant policy use as exploitative policy selection, less dominant policy use as exploratory). The number of turns on each trial which fell into a participants' dominant policy were used for this analysis.

1.4.2. *Eyetracking*

Binocular eyetracking data was collected using the SR Research Eyelink 1000 system. For each participant, binocular thirteen-point calibration was conducted; where calibration was unsuccessful using both eyes, one eye was used. The screen x and y coordinates were preprocessed for analysis. Preprocessing involved removing any values outside of the screen bounds, interpolating eyeblinks (as defined by pupil size outside of 1.5 standard deviations below to 2 standard deviations above participant average pupil size), applying a Hanning window of 15 samples (93% overlap) to smooth the eyetrace, and replacing temporarily lost values in one eye with valid data from the other eye (including for whole trials if one eye was excessively noisy). Data was then epoched into trials, and downsampled to match the stimuli framerate for alignment with behavioural data. Trials with poor signal were defined as those with more than 30% of the samples interpolated in both eyes, or whose recorded behavioural data was outside of two standard deviations above or below the participants mean recorded trial length (as the source of these outliers could not be identified). For the final sample of participants, there were a maximum of 65 poor-signal trials (mean = 33.4). Poor-signal trials were removed from all analyses (including behavioural only dependent variables above).

The square the participant was looking at was determined by a novel biased-nearest-object method (see Supplementary Materials), which assumes that at a given moment the participant is looking at one square. While this assumption is nearly always correct, observation of the replay of many trials during development of this method (see also task demonstration video: https://figshare.com/s/fd2742b897e21d901dd0 (DOI: 10.26180/5eabbfb9a8aa4)) showed occasionally participants did not fixate or smoothly pursue one square at a time, instead seemingly relying on peripheral vision. This was likely the participants initially and temporarily tracking multiple squares to narrow down their next hypothesis, consistent with multiple object tracking literature (Fehd & Seiffert, 2008). Times of *hypothesis switch* from one square to another were defined as any change in the looked-at square that lasts longer than one frame.

The Euclidean distance between the expected location (had the stimuli followed the mouse) and the actual location of the hypothesised square was calculated as a proxy for *prediction error*. This means that the prediction error is contingent on how quickly the participants move (the error is higher if they move faster). Due to the manipulation, low variability trials accrue less prediction error on average than high variability trials. The prediction error is also impacted by the magnitude of the distracter's angular offset when the hypothesised square is not the target, so the quality of the hypothesis will affect prediction error. The *average prediction error* for each participant was calculated across each trial. The *slope of prediction error*, representing the rate of prediction error minimisation, was the slope of the line of best fit of the average prediction error at each time point in each condition for each participant (see Fig. 5a and c). As such, negative values here represent prediction error minimisation, while positive values represent accumulating prediction error.

Finally, given the temporal resolution of our prediction error measure, we were interested in the pattern of prediction error around key temporal events – namely hypothesis switches and changes to the variability distribution (due to volatility). We call these analyses *event-related prediction errors* (ERPE). A one second epoch was centred on the event of interest (time zero) and prediction errors were averaged for each participant in each condition to create an average pattern of activity around the event. Means over five 200 ms time bins for each participant were taken for statistical analysis (bin number three is centered on the event onset, see Fig. 6a). There was no effect of time-bin in the volatility ERPE analysis, hence these are reported in the Supplementary Materials.

1.5. *Statistics*

All statistical analyses were conducted as Mixed Linear Models (MLM) using Jamovi version 1.1.4 and the GAMLj module (Gallucci, 2019; R Core Team, 2018; The Jamovi Project, 2019). Trial-wise data was used for all dependent variables except prediction error slopes and ERPEs for which condition-wise data was used. Variability and volatility were modelled as simple fixed effect factors, and AQ score was modelled as a continuous fixed effect. All interactions between fixed effects were included. By-participant random intercepts were included to address the non-independence of subject-level observations across trials and capture individual variability in task performance. Compared to traditional methods, this approach affords more sophisticated handling of missing and outlying data, thus improving the accuracy, precision, and generalisability of fixed effect estimates (Singmann & Kellen, 2020). See Table 1 for additional covariates for each model. Degrees of freedom are reported as estimated by the Satterthwaite method. *Post-hoc* analyses were conducted with Bonferroni correction for multiple comparisons and post-hoc *p*-values are reported with this correction. For ease of interpretation, post-hoc tests for interactions with AQ were simple effects contrasting participants with three levels of autism traits: low (<Mean-1SD = 16, *n* = 6), within one standard deviation from the mean, and high (>Mean + 1SD = 27, n = 6) scores.

1.6. *Data availability*

The dataset used for Results, Table 1, and Figs. 2–6 is freely available at https://figshare.com/s/77dececaa2b966db4cf7 (DOI:10.26180/5ed0708f103a2) (Perrykkad, Lawson, Jamadar, & Hohwy, 2020).

Table 1
Significant results summary.

	Dependent Variable	Additional Covariate	M.E. Variability	M.E. Volatility	M.E. AQ	Var*Vol	Var*AQ	Vol*AQ	Var*Vol*AQ
Task	Accuracy		*** low>high			** effect of var. is stronger in high vol			
Movement and Strategy	Time Spent Moving	Time to Movement	*** low<high						*** only for high AQ in high var.: low vol > high vol
	Speed		*** low>high			*** low var. high vol > both low vol & high high; low low>high high			
	Acceleration		*** low<high		* decreasing with AQ				
	Jerk								
	Turning Behaviour		*** low>high			* effect of var. is stronger in low vol			
	Dominant Policy Use	Number of Turns						*** in low AQ low vol < high vol. In high AQ, low vol > high vol	* the low AQ*vol diff. is lost in high variability
	Hypothesis Switches		*** low>high			** in low var. only low vol < high vol			
Prediction Error	Average Prediction Error	Accuracy (and all interactions)	*** low<high				*** var. difference is smaller in high AQ		
	Condition-wise Slope		*** low<high[1]	* low>high					
			M.E. Agency	M.E. Accuracy	M.E. AQ	Accuracy*Agency	Accuracy*AQ	Agency*AQ	Accuracy*Agency*AQ
	Agency-wise Slope		*** not>agent			*** in no agency judgement there is no diff. between correct and incorrect			* High AQ there is no diff. between accuracy; Low AQ when judge agency correct<incorrect, reverse when no-agency and no diff. between agency in incorrect
			M.E. Variability	M.E. Volatility	M.E. AQ	Var*Vol	Var*AQ	Vol*AQ	Var*Vol*AQ
	Volatility ERPE	Average Prediction Error	* low<high						
	Hypothesis Switch ERPE	Average Prediction Error; Hypothesis Switches	*** low>high	* low<high		** only in low var., low vol < high vol			
			…	M.E. Time Bin	AQ*Time Bin	Var*Time Bin	…		
				*** T2 and T3 (event) are>T1, T4, T5 which are equal	*** In T3, AQ is negatively associated with prediction error	*** Only in low var. T2 > T1, bigger var. diff. at T3			

For variables, M.E. = Main Effect, * = Interaction, For results, * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$ (Post-hoc values Bonferroni corrected for multiple comparisons), Var = Variability, Vol = Volatility, AQ = Autism Quotient, T1-5 = Time bins 1–5, diff. = difference, [1]See Supplementary Materials.

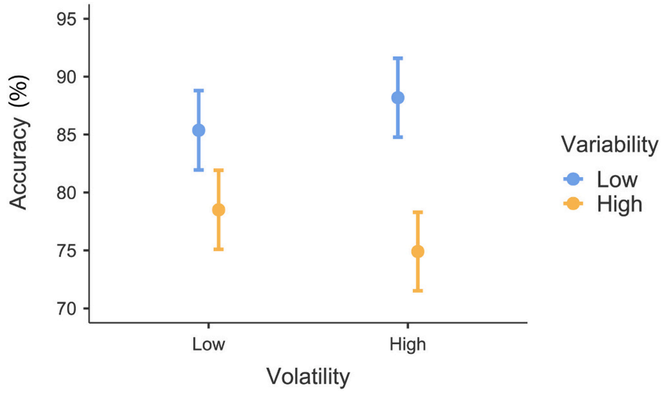


Fig. 2. Accuracy.
Proportion of trials where participants chose the correct square. Participants were more accurate in low variability (blue) than high variability (orange), and this difference was more pronounced under high volatility (right) than low (left). Error bars are 95% CI.

2. Results

In this section, we summarise all statistical models in three sections, first, covering overall task performance, second, the movement and strategy measures, and last, prediction error measures. For each section, we describe the effect of uncertainty on the dependent variables followed by AQ results (though all statistical models included all fixed factors as above). For brevity, we report only main effects of uncertainty in the movement and policy variables. Full statistical reporting is included in Supplementary Materials. See also Table 1 for a summary of all significant results. Performance on the motor control task did not significantly correlate with AQ ($r = 0.07$, $p = 0.65$) or overall accuracy ($r = -0.21$, $p = 0.20$), and so was not included as a random effect in any mixed model.

2.1. *Overall task performance: judgement of agency*

Average accuracy in the judgement of agency task (Fig. 2) was moderately high across conditions (μ = 81.0%, σ = 9.12%). MLM results show a significant main effect of variability (F(1,4664) = 85.07, $p <$ 0.001) such that accuracy was approximately 10% higher in the low variability condition than in the high variability condition. Additionally, there was a significant interaction between variability and volatility (F (1,4665) = 8.62, $p = 0.003$). Post-hoc analysis revealed significant differences in all comparisons between the four conditions (z = 4.41–8.66, $p < 0.001$) except between low and high volatility when variability remained constant (low/low vs low/high $p = 0.421$, high/low vs high/high $p = 0.115$). This result indicates that while volatility does not make a significant difference to accuracy on its own, the effect of variability on accuracy was stronger under high volatility. There was no significant effect of AQ on accuracy.

2.2. *Movement characteristics and policy selection*

Participants moved for an average of 13.7 s per trial (σ = 1.55, Range = 3.95–15.1). An MLM comparing the average duration of each trial spent moving across conditions (with the additional fixed effect of time to movement on each trial to account for possible confound) found a significant main effect of variability (Fig. 3a; F(1,4660) = 727.71, $p <$ 0.001). Participants moved for longer in high variability conditions compared to low variability conditions by an average of 801 ms.

An MLM analysis on average speed of movement revealed a significant main effect of variability (Fig. 3b; F(1,4661) = 36.42, $p < 0.001$) such that participants moved faster in the low variability condition compared to high (z = 6.03, p < 0.001). An MLM on acceleration showed a main effect of variability (Fig. 3c; F(1,4664) = 12.68, p < 0.001), with faster average acceleration in the high variability trials, compared to low (z = 3.56, p < 0.001). An MLM on jerk showed no significant results (Fig. 3d).

On average, each trial contained 35 turns (Fig. 3e; σ = 13.9, Range = 6–107). An MLM on turn count showed a significant main effect of

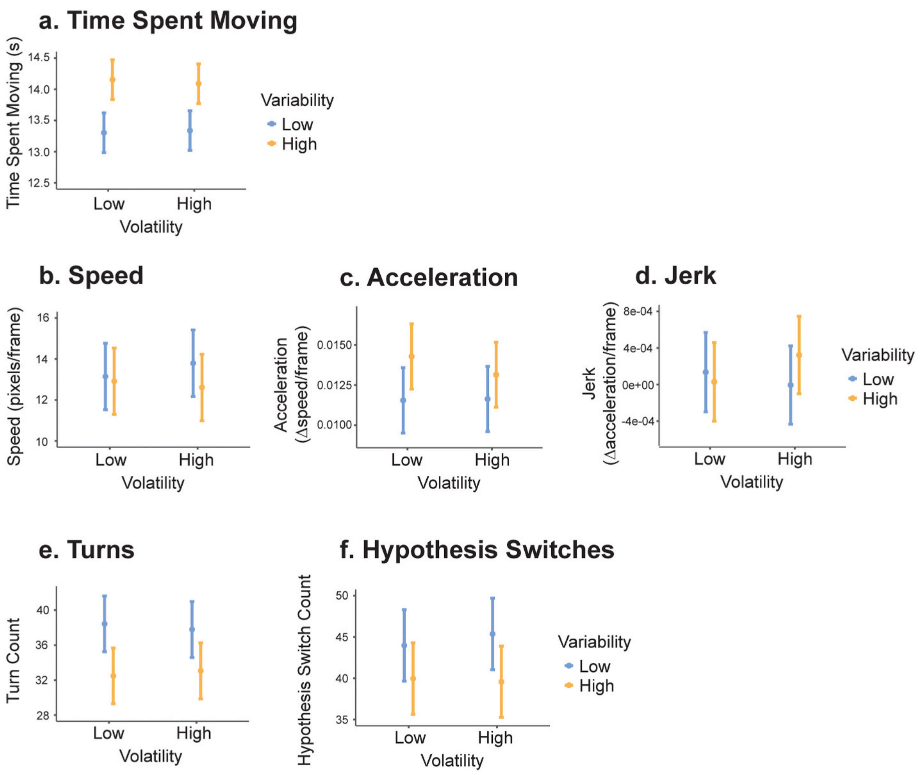


Fig. 3. Movement and strategy.
These graphs depict movement and strategy variables (except dominant policy use, see Fig. 4) across all participants. Volatility is along the x-axis for each graph. Orange bars represent high variability, blue bars represent low variability. Error bars are 95% CI. a) shows mean duration of each trial spent moving, controlling for time to movement onset on each trial; b) shows average speed of movement, c) average acceleration and d) average jerk; e) shows average turn count on each trial; f) shows the average number of hypothesis switches on each trial, when the participant moves their eyes from one square to another. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

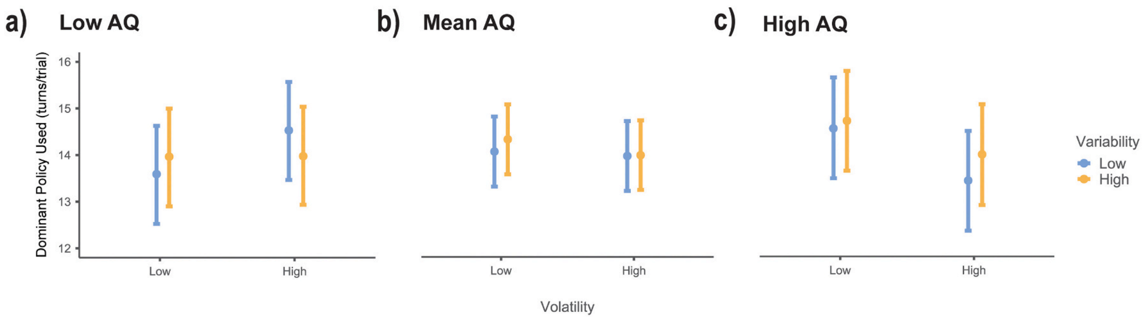


Fig. 4. Dominant policy use.
The turns participants made were categorised into types. This figure shows the number of turns in the participants' own dominant strategy, controlling for total number of turns. For participants with low AQ (<16, panel a), only for low variability trials (blue), participants used their dominant policy more in high volatility (right) than low (left). For participants with AQ scores within one standard deviation of the mean (panel b), there was no difference between volatility conditions (left/right). In both variability conditions (blue and orange), participants with high AQ (>27, panel c) used their dominant policy more in low volatility (left) than high (right). Error bars are 95% CI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

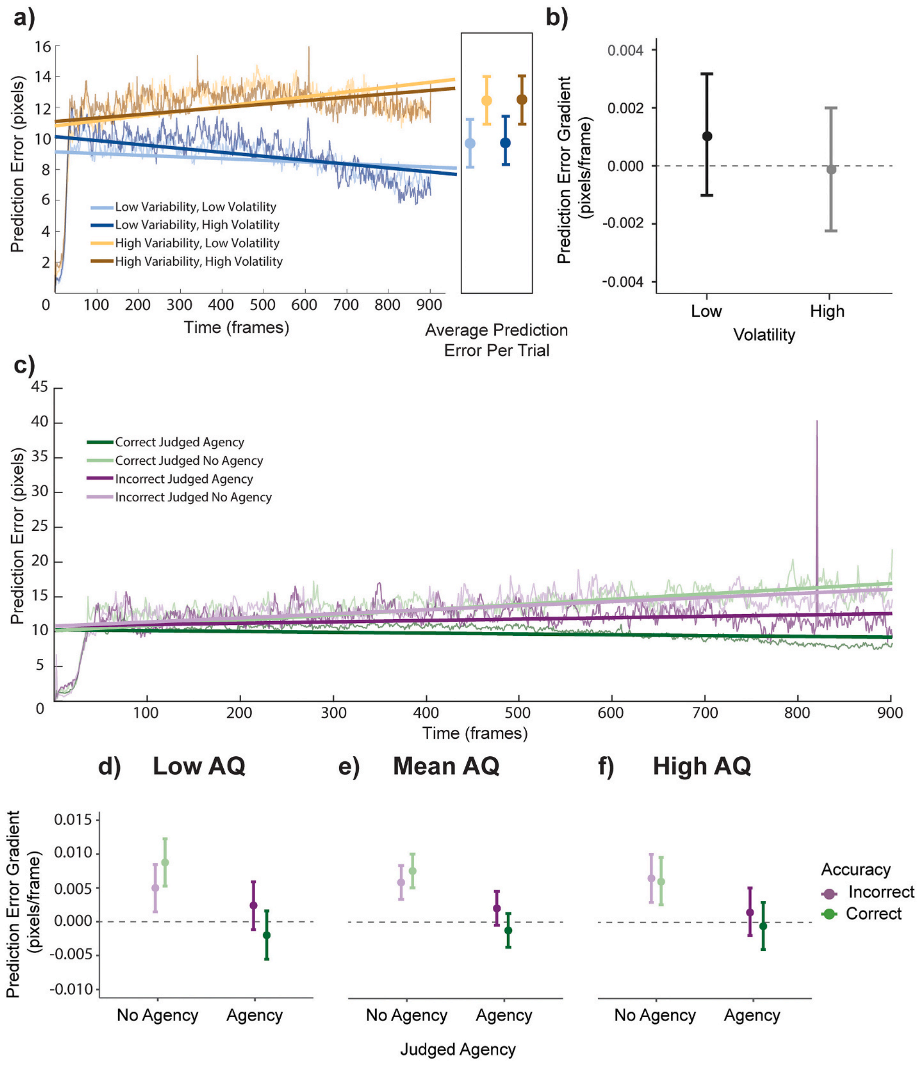


Fig. 5. Prediction error average and gradient.
Panel a) shows the grand average prediction error across the trial split by condition with lines of best fit for each. The box at the end of the graph shows the average prediction error across trials in each condition. Panel b) shows the mean gradient or slope for the lines of best fit for each participant under different levels of volatility. Data used for the box at the end of panel a) is adjusted to account for the influence of accuracy. Panel c) shows the grand average prediction error across the trial split by correct (green) and incorrect (purple) trials and whether the participants chose a square (Judged Agency, dark colours) or said that it was a no-control trial (Judged No Agency, light colours) with lines of best fit for each. Panels d-f show the mean gradient or slope for the lines of best fit for each participant in each combination, split by AQ score. Error bars are 95% CI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

variability (F(1,4661) = 346.22, p < 0.001) such that participants turned more frequently in low variability than high variability trials (z = 18.61, $p < 0.001$). The dominant turn-types (*policies*) in order of frequency across participants were Non-Cardinal (n = 22), Hesitant-straight (n = 14), Horizontal (n = 3) and Circle (n = 1). On average, in each trial, participants used their dominant policy 39.3% of the time (σ = 16.0, Range = 0–100), and within each participant, the average percent of turns on each trial that were of their dominant policy ranged from 30.1% to 51.9%. For the MLM on dominant policy turn count for each trial, the additional covariate of absolute number of turns on each trial was included to account for this confound. There were no significant main effects.

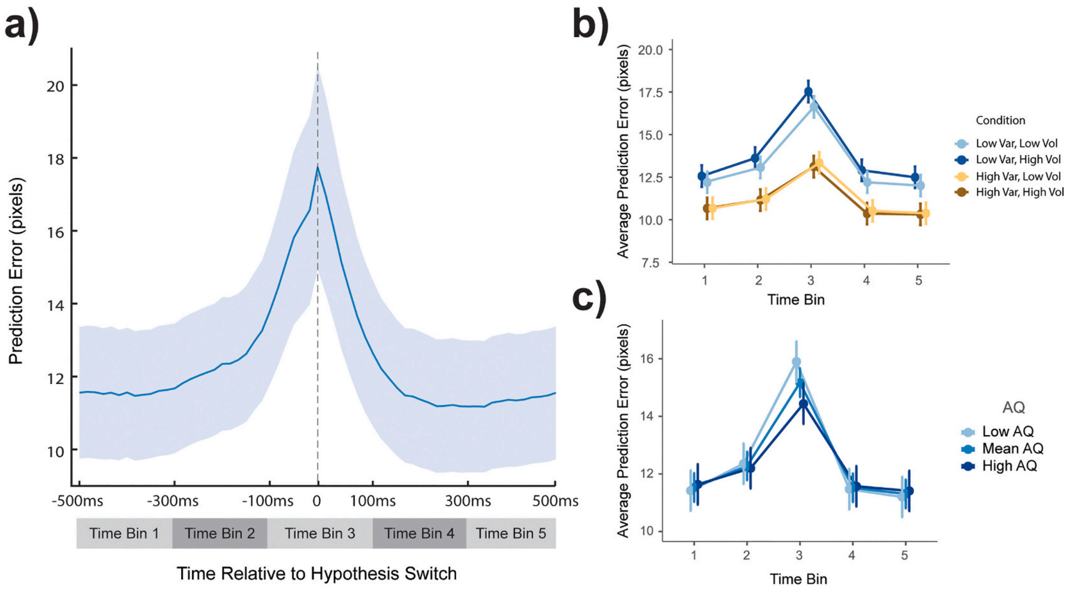


Fig. 6. Hypothesis switch event-related prediction error (ERPE).

Panel a) shows the grand average (blue line) prediction error across participants in a one second epoch centered on hypothesis switches. Time bins used for statistical models are represented in grey shaded bars below. Data used in statistical models, and therefore in panels b) and c) is adjusted for average prediction error differences between conditions and average number of hypothesis switches. Panel b) shows average prediction error in each time bin for each condition. There is more prediction error in this epoch in low variability (blue) conditions than high (orange). This difference is greatest in time bin three, at the time of the event. In low variability (blue), low volatility (light blue) conditions showed less prediction error in this epoch than high (dark blue). Time bin three has the greatest prediction error, followed by time bin two, and none of the others are significantly different from each other. The increase from time bin one to two is only significant in low variability (blue). Panel c) shows the data split by AQ score - lower AQ scores (lightest blue) are associated with greater prediction error at the time of the event (time bin three). Error bars and shading are 95% CI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

On average participants switched hypotheses 42.2 times per trial (Fig. 3f; σ = 13.48, Range = 6–134). An MLM on hypothesis switch counts in each trial showed a main effect of variability (F(1,4661) = 195.91, $p < 0.001$) such that participants switch hypotheses more when variability is low than when it is high (z = 14.00, p < 0.001).

These findings suggest that participants' movement was strongly affected by increased environmental variability, causing participants to move more, move slower but accelerate more quickly, and switch hypotheses less often.

2.3. Autism traits and movement and policy

For the dependent variable of time spent moving, there was a significant three-way interaction between AQ, variability and volatility (F (1,4660) = 11.37, $p < 0.001$). Post-hoc tests showed that for participants with high AQ only, under high variability only, participants moved for an average of 200 ms longer in low volatility than high volatility conditions. Additionally, the model considering acceleration showed a main effect of AQ (F(1,38) = 5.73, $p = 0.022$), such that mean acceleration decreased with increasing AQ (R^2 = -0.011, $p < 0.001$). There were no significant findings relating to autism traits across other movement characteristics.

Considering how participants across the AQ range changed their policies in response to uncertainty, the model for dominant policy use showed a significant interaction between AQ and volatility (F(1,4660) = 19.17, p < 0.001), and a significant three way interaction between AQ, variability and volatility (Fig. 4; F(1,4661) = 4.27, $p = 0.039$). Post-hoc analyses for the two-way interaction showed that for low AQ (Fig. 4a) participants used their dominant policy more in the high volatility condition (z = 2.14, $p = 0.032$), but only when variability was low (z = 2.92, $p = 0.004$), otherwise volatility made no difference (z = 0.10, $p = 0.918$). For high AQ (Fig. 4c) participants used their dominant policy more in the low volatility condition (z = 4.05, p < 0.001), regardless of the variability (high: z = 2.20, $p = 0.028$; low: z = 3.53, p < 0.001).

These findings suggest that different levels of autism traits were associated with differences in the quantity of sampling behaviour, differences in fine-grained movement qualities and differences in the flexibility of policy-selection itself.

2.4. Prediction error

Across all participants, the average calculated prediction error per trial was 10.5 pixels per frame (Fig. 5a; σ = 4.93, Range = 0.43–50.1). An MLM analysis with the addition of accuracy and all of its interactions with the other fixed factors revealed that, as expected, average prediction error across each trial was significantly associated with the variability condition (F(1,4653) = 284.05, $p < 0.001$).

Comparing the slope of prediction error in each condition, an MLM with the addition of accuracy as a random effect revealed a significant main effect of variability (F(1,114) = 58.15, p < 0.001), which indicated that there was more prediction error minimisation in the low variability condition (lower gradient) than the high (See Fig. 5a; $t = 7.63$, $p < 0.001$). However, this main effect may be explained by a confound of the effect of accuracy which could not be modelled as a fixed effect due to high correlation between the effect of variability and the effect of accuracy on this dependent variable (see Supplementary Materials for a model including accuracy as a random effect in which the significant main effect is removed). There was also a marginally significant main effect of volatility (F = 3.96, p = 0. 049), which showed steeper prediction error minimisation in high volatility compared to low volatility (See Fig. 5a and b, $t = 1.99$, $p = 0.049$).

To investigate the relationship between prediction error minimisation and the participant's judgements, we performed an MLM with a different structure. For each participant, a linear fit to prediction error across trials with the same accuracy and agency judgement served as the dependent variable. AQ score, accuracy and agency were included as fixed effects, and participant as a random intercept. This MLM showed a main effect of agency (F(1,113) = 82.89, p < 0.001) and an interaction between agency and accuracy (Fig. 5c; F(1,113) = 12.79, p < 0.001). Agency was associated with increased prediction error minimisation (t (113) = 9.10, p < 0.001). Post-hoc tests for the interaction showed that only when participants judge that they did not control any of the stimuli was there no difference in prediction error minimisation between correct and incorrect trials (t(113) = 1.74, p = 0.51). When participants judge that they did have agency, there is more prediction error minimisation when they are correct than incorrect (t(113) = 3.31, p = 0.007). However, when considering either only the correct or incorrect trials, prediction error minimisation was steeper when participants judged that they had agency (correct: t(113) = 9.00, p < 0.001; incorrect: t(113) = 3.89, p = 0.001;) which confirms that the judgement of agency was associated with steeper prediction error minimisation regardless of accuracy. Numerically, the mean slope of the prediction error was only negative (indicating successful prediction error minimisation) when participants were both accurate and judged that they had agency.

To look at the effect of uncertainty and AQ score on dynamics of prediction error and hypothesis testing, we performed an MLM on the ERPE centered on hypothesis switches. In addition to the standard MLM, we included time-bin as an additional fixed effect of interest and average prediction error and average number of hypothesis switches in each condition as fixed-effect covariates. Fig. 6a shows the timeseries for the average prediction error across conditions and participants in the analysed epoch. There were significant main effects of variability (F(1,512) = 125.10, p < 0.001), volatility (F(1,719) = 6.14, p = 0.013) and time-bin (F(4,719) = 252.29, p < 0.001) and two-way interactions between variability and volatility (Fig. 6b; F(1,729) = 10.76, p = 0.001) and variability and time-bin (F(4,719) = 17.94, p < 0.001). Time bins one, four and five were not significantly different from one another (t(720) = 0.06–1.55, p = 1.00) but the others were all significantly different from one another (t(720) = 5.24–26.79, p < 0.001), indicating a significant increase before the hypothesis switch starting at least 300 ms before, and a drop after back to the initial level of prediction error. Post-hoc analyses into the main effect of variability showed that low variability conditions had *greater* prediction error around the time of a hypothesis switch than did high variability conditions (t(519) = 11.09, p < 0.001), which is the inverse of the pattern for average prediction error across the whole trial. Post-hoc analysis of the interaction between time-bin and variability showed that the difference between variability conditions held across all time bins surrounding the hypothesis switch (t(697) = 6.15—13.81, p < 0.001), but that this difference was greater during time bin three (3.89 pixels, greater than other bin averages by at least 1.69 pixels). Further, only in low variability is there a significant increase from time bin one to two (t(720) = 6.15, p < 0.001), indicating the increase may occur closer to the event in high variability conditions. While overall, low volatility was associated with less prediction error than high volatility around the time of hypothesis switches (t(721) = 2.48, p = 0.013), post-hoc analysis of the interaction between variability and volatility showed that this only holds when variability was low (t(727) = 4.07, p < 0.001). The main effects of this analysis also hold when data is restricted to only incorrect trials, suggesting this result is not driven by an artefact of trial accuracy (see Supplementary Materials for full statistical model and figure).

These findings suggest that increased prediction error minimisation is associated with increased volatility and correctly and positively inferring agency. We have also shown that hypothesis switches function to reduce rising prediction error, and that the dynamics of minimising prediction error in this way is affected by environmental uncertainty at the levels of both variability and volatility.

2.5. *Autism traits and prediction error*

The model considering the effect of uncertainty and autism traits on average prediction error across a trial showed a significant interaction between variability and AQ (F(1,4653) = 10.58, p = 0.001). Post-hoc analyses of the variability × AQ interaction showed that the difference between variability conditions decreases as AQ increases (though they are still significantly different across all AQ scores; z = 9.01–14.44, p < 0.001).

Additionally, the MLM considering agency, accuracy and AQ showed a three-way interaction between these variables (F(1,113) = 5.69, p = 0.02). Post-hoc analyses showed no difference between agency judgements for incorrect trials for participants with a low AQ score (Fig. 5d; F (1,113) = 3.51, p = 0.064), but otherwise, when participants judged that they had agency over one of the stimuli, the slope of their prediction error was lower, indicating that they were more effective at minimising prediction error (t(113) = 9.10, p < 0.001) (both the mean and high AQ groups, and when correct in low AQ). Further, while low AQ participants' prediction error was maximally sensitive to accuracy (lower slopes when correctly judging agency than incorrectly doing so, F (1,113) = 10.06, p = 0.002; and lower slopes when incorrectly denying agency than when correctly doing so, F(1,113) = 7.75,p = 0.006); high AQ participants' prediction error was not sensitive to accuracy at all (Fig. 5f; F(1,113) = 0.11–2.29, p = 0.13–0.74). Participants with a mean AQ showed the appropriate difference only when they judged that they had agency (F(1,113) = 10.99, p = 0.001).

Looking at the prediction error dynamics limited to the epoch around hypothesis switches showed a significant interaction between AQ and time-bin (Fig. 6c; F(4,719) = 12.16, p < 0.001). Post-hoc analysis showed a significant difference only in time-bin three (the time of the event) depending on the AQ score (F(1,50) = 8.58, p = 0.005). A further Pearson's correlation test of AQ by prediction error in this time-bin showed that as AQ increased, the prediction error at the time of a hypothesis switch decreased (r = −0.21, p < 0.001).

These findings suggest that uncertainty in the environment differentially affects participants' prediction error depending on measured autism traits, including the relationship between prediction error minimisation and judgement of agency, and propensity to switch hypotheses in response to increasing prediction error.

3. Discussion

In this experiment, we closed the action-perception loop to investigate how uncertainty in self-caused sensations influences successive choices about which actions to perform to infer agency. Unlike many previous studies, these actions were freeform and temporally contiguous with ongoing sensory consequences. We showed that action selection changes depending on uncertainty in the mapping between actions and sensory outcomes. We also demonstrate that agency inferences reflect the temporal dynamics of prediction error.

One of the most significant advances of this study on previous designs is the ability to measure and interrogate the temporal dynamics of prediction error, and how this relates to participant behaviour. Using this proxy for prediction error there were particularly interesting findings in the behavioural pattern around hypothesis switches and prediction error minimisation for trials with different judgements of agency. We will now discuss each of these in turn.

Our eye-tracking analysis indicates a hypothesis switch when the participant moves from looking at one square to another and is indicative of a change in the moment to moment beliefs about agency with respect to the candidate square. For action to occur under the active inference account, prediction error comes first, and the action is performed to resolve it. This is consistent with the increasing prediction error leading to a hypothesis switch in our task, indicated by the significant peak in prediction error at the time of the hypothesis switch. The current agential hypothesis is abandoned when the prediction error is too high – there is decreasing evidence that one can achieve one's expected state with the available actions under the current hypothesis, which leads to a switch that alleviates prediction error. This finding is uniquely consistent with predictive processing (Friston, 2017).

Environmental uncertainty influences this pattern too; after removing trial-wise average prediction error, low variability conditions have a higher prediction error in the hypothesis switch epoch. Also, only in these low variability conditions is there a significant increase from time bin one to time bin two, preceding the switch. Both of these findings suggest that when variability is low, prediction error is allowed to increase for a comparatively longer period of time before the participant decides to switch. This may reflect more reliance on priors in such environments, which allow stable accumulation of evidence for a given hypothesis, and a reluctance to abandon hypotheses in the face of sensory evidence to the contrary.

By looking at the relationship between participants' agency reports and the trend in prediction error over time, our results suggest that participants could be using these trends to inform their judgement of agency. Agency judgements, whether correct or not, were associated with a more negative prediction error slope. Under the predictive processing account, a correct judgement of agency should be associated with a negative trend in prediction error, and a correct judgement of no-agency should not be associated with prediction error minimisation, as the participant cannot effectively control the stimuli to reduce prediction error. These hypotheses were fully borne out for participants with low AQ scores – when participants correctly judged that they had no agency, the slope of the prediction error was more positive (i.e. failed prediction error minimisation) than when they incorrectly said that they had no agency.

Traditionally, internal representations of agency have been explained using a comparator model. In this model, upon movement, the neural system creates an efference copy of motor commands, which predicts "future states of the motor system and the sensory consequences of movement" (Moore & Obhi, 2012)p. 549). This is then compared with incoming sensory information. In both the comparator and predictive processing accounts, agency is associated with small prediction error, or a match between expected and actual outcomes of actions. The comparator however focuses on net retrospective prediction error and cannot account for hypothesis switches in the face of accumulating prediction error or other changes in future action based on inferences of agency (see also Zaadnoordijk, Besold, and Hunnius (2019)). The predictive processing account positions agency in a broader theory of action and policy selection. So, if the projected reliability of policy-outcome mappings over time under a particular hypothesis (occurrent agency) changes, this account is consistent with a threshold in accumulating prediction error after which the agent switches hypotheses and is especially well equipped if this threshold is sensitive to environmental volatility. Our hypothesis switch ERPE suggests that hypothesis switching is sensitive to volatility when variability is low, with more prediction error around a hypothesis switch when volatility is high.

These results provide a reminder that agents' ability to discern, and make judgements about, agency arises as they *actively* close the action-perception loop, not just in passive perceptual processes. The results also offer an indication of how agents do this, namely through exploratory titration of prediction error, in a pattern that is sensitive to variability and volatility. It may be that affording agents the opportunity for exploration of the action-perception loop is critical for agency inference and judgement.

Comparing the two levels of uncertainty manipulated here, changes to variability caused the most broad-reaching effects. Under high variability, participants were less accurate but spent longer sampling the environment, moved slower but accelerated more quickly, switched hypotheses less frequently and turned less, compared to the low variability conditions. The finding that participants move more under increased variability is consistent with the findings by Wen and Haggard (2020) in a similar judgement of agency paradigm.

While volatility was expected to have effects independent from variability, most of the significant effects for volatility were interactions with variability; volatility only showed two main effects. The first main effect indicated that prediction error was reduced more quickly under high volatility. In our manipulation, the timing of volatile switches was unpredictable, so this effect is likely due to an increased vigilance or sensitivity to incoming information manifesting as an increased learning rate under high volatility (Mathys, Daunizeau, Friston, & Stephan, 2011). The second main effect of volatility indicated that higher volatility was related to higher prediction error in the epoch surrounding hypothesis switches, however this was only true when variability was low. In two further cases, the effect of volatility was only seen in low variability; specifically that participants move faster and switch hypotheses more in high volatility than low. This could reflect an attempt to garner more evidence about the current state of the world before it changes. Lower volatility also magnified the effect of variability on the number of turns made during each trial. Higher volatility, on the other hand, increased the effect of variability on task accuracy. Future studies should consider ways of highlighting changes in volatility to enhance the potential effect of higher order uncertainty, such as making them large enough to stand out more saliently to the participant.

It is important to keep in mind too that our analyses of prediction error were limited to a behavioural proxy (combining eye-tracking and mouse movement) for prediction error that does not directly reflect changing internal representations of environmental uncertainty. This also affects what conclusions we can draw about the relationship between certain kinds of uncertainty and prediction error where the uncertainty strongly affected accuracy. For example, in our statistical model of the rate of prediction error minimisation split by uncertainty conditions, the effect of accuracy on the slope of prediction error was nearly identical to the effect of decreased variability on the slope of prediction error. This means that our results cannot distinguish between steeper prediction error minimisation due to an easier task or lower variability (see Supplementary Materials for statistical models including and removing the random effect of accuracy). To address the question of the effect of variability (independent of accuracy) on the rate of prediction error minimisation in this kind of task, difficulty across conditions could be titrated for each participant by adjusting other features which may affect task difficulty (such as distracter similarity or number, as in Williams and Happé (2009) and Grainger et al. (2014)). Including more levels of variability (rather than simply high and low), may also help to statistically distinguish the effect of accuracy and variability for future experiments. Future research should also consider using neural estimates of prediction error or computational modelling that appropriately changes priors with uncertainty.

Here, we found no difference in accuracy of judgement of agency between healthy participants across a range of autism traits, consistent with previous research comparing autistic and healthy participants on similar measures (David et al., 2008; Grainger et al., 2014; Russell & Hill, 2001; Williams & Happé, 2009; Zalla et al., 2015). As previously noted by Perrykkad and Hohwy (2020b) and Zalla and Sperduti (2015), this is in contrast to sense of agency in autism being shown to be reduced under typical experimental paradigms (Sperduti, Pieron, Leboyer, & Zalla, 2014; van Laarhoven, Stekelenburg, Eussen, & Vroomen, 2019). Our study also shows no main effects of AQ on other outcomes, except for a negative association with acceleration.

To our knowledge, Zalla et al. (2015) is the only other case where variability of a similar kind (which they labelled '*turbulence*') was added in a judgement of agency task pertaining to autism, in their case contrasting participants with and without an autism spectrum diagnosis. Their results demonstrated that the accuracy of autistic participants' agency judgements was less sensitive to differences in variability than the neurotypical group. This study supports our hypothesis that the addition of uncertainty has a distinctive effect on judgement of agency related to autistic traits. While we do not show any significant interactions with AQ in accuracy, our results showed participants with high autism traits were less sensitive to differences in variability in their average prediction error. Since this measure is behavioural, this suggests that participants with high AQ were moving (that is, exploring the environment) in a way that did not reflect underlying differences in variability. Further, AQ was negatively associated with prediction error in the 200 ms window surrounding hypothesis switches. This suggests participants with high AQ are switching hypotheses earlier than participants with low AQ, or tolerating less uncertainty before abandoning their current hypothesis (see also Lawson et al. (2017)).

By additionally manipulating volatility, we could demonstrate further effects of uncertainty dependent on AQ. Participants with high AQ were more sensitive to differences in volatility such that only for this group was increased volatility associated with more time spent moving (if only in high variability) and more flexibility in policy selection. This might reflect less consistent or shallower internal models (Perrykkad & Hohwy, 2020b), which leads to less precision over all policies in high volatility, and so the selection of one over another fluctuates more frequently. This pattern is the opposite of the low AQ group, where high volatility was associated with more dominant policy use (but only in low variability). This is also consistent with Lawson et al. (2017), who showed that autistic participants update their learning in response to volatility more readily than neurotypical participants.

Of note, our findings with respect to autism are limited to scores on a trait-based measure, which may not generalise to diagnosed autistic populations. Our sample had a high average AQ score compared to what is expected in the general population (Baron-Cohen et al., 2001), so our results for “low” AQ may actually be more representative of “average” AQ individuals. While overall the sample size in post-hoc analyses is low, the omnibus interactions were based on modelled trends in the full dataset of continuous AQ scores. Nevertheless, environmental uncertainty might be particularly relevant to action selection for different levels of autistic traits and we do show interactions between uncertainty and AQ. These are worth following up in future studies in diagnosed samples.

In summary, this suggests autistic traits are related to 1) subtle differences in more abstract action policies, which are more sensitive to volatility, 2) smaller differences in prediction error between variability conditions, and 3) a greater propensity to switch hypotheses at a lower prediction error threshold when inferring agency. Notably, despite these differences, there was no significant effect of AQ on overall number of hypothesis switches or on accuracy.

4. Conclusion

This experiment shows that uncertainty in the mapping between actions and their outcomes changes not only how effectively participants can identify which stimuli they have control over, but also changes the actions they make and the overall strategies they employ. These changes have downstream impacts on the prediction error which can be used to inform their next action, and their overall response in each trial. In addition, our data illuminates subtle differences in this perception-action loop dependent on autism traits.

Acknowledgements

The authors appreciate the time and effort of their participants without whom this research would not be possible. The authors would also like to thank members of the Cognition and Philosophy Lab, particularly Andrew Corcoran, Jonathan Robinson, Stephen Gadsby, and Andrew McKilliam for their helpful comments on an earlier version of this paper. Thanks also to Chase Sherwell and Bryan Paton for analysis advice.

Perrykkad is supported by an Australian Government Research Training Program (RTP) Scholarship. Hohwy supported this project using an Australian Research Council Discovery Grant (DP160102770, DP190101805). Lawson is supported by a Wellcome Trust Henry Dale Fellowship (206691/Z/17/Z) and an Autistica Future Leaders Award (7265). Jamadar is supported by an Australian National Health and Medical Research Council Fellowship (APP1174164) and the Australian Research Council Centre of Excellence for Integrative Brain Function (CE1400100007).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.104598.

References

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, Malesand females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5–17. https://doi.org/10.1023/A:1005653411471.

Bednark, J. G., Poonian, S., Palghat, K., McFadyen, J., & Cunnington, R. (2015). Identity-specific predictions and implicit measures of agency. *Psychology of Consciousness: Theory, Research and Practice, 2*(3), 253.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind.* Oxford University Press.

David, N., Gawronski, A., Santos, N. S., Huff, W., Lehnhardt, F.-G., Newen, A., & Vogeley, K. (2008). Dissociation between key processes of social cognition in autism: Impaired Mentalizing but intact sense of agency. *Journal of Autism and Developmental Disorders, 38*(4), 593–605. https://doi.org/10.1007/s10803-007-0425-x.

Desantis, A., Hughes, G., & Waszak, F. (2012). Intentional binding is driven by the mere presence of an action and not by motor prediction. *PLoS One, 7*(1), e29557. https://doi.org/10.1371/journal.pone.0029557.

Engbert, K., & Wohlschlager, A. (2007). Intentions and expectations in temporal binding. *Consciousness and Cognition, 16*(2), 255–264. https://doi.org/10.1016/j.concog.2006.09.010.

Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition, 108*(1), 201–209. https://doi.org/10.1016/j.cognition.2007.11.008.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138.

Friston, K. (2017). The variational principles of action. In J.-P. Laumond, N. Mansard, & J.-B. Lasserre (Eds.), *Geometric and numerical foundations of movements* (pp. 207–235). Cham: Springer International Publishing.

Friston, K., Samothrakis, S., & Montague, R. (2012). Active inference and agency: Optimal control without cost functions. *Biological Cybernetics, 106*(8), 523–541. https://doi.org/10.1007/s00422-012-0512-8.

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. (2013). The anatomy of choice: Active inference and agency. *Frontiers in Human Neuroscience, 7*(598). https://doi.org/10.3389/fnhum.2013.00598.

Gallucci, M. (2019). GAMLj: General analyses for linear models [jamovi module]. Retrieved from https://gamlj.github.io/.

Grainger, C., Williams, D., & Lind, S. E. (2014). Online action monitoring and memory for self-performed actions in autism Spectrum disorder. *Journal of Autism and Developmental Disorders, 44*, 1193–1206.

Hohwy, J. (2013). *The predictive mind.* Oxford University Press.

Hohwy, J. (2015). Prediction, agency, and body ownership. In *The Pragmatic Turn:: Toward Action-Oriented Views in Cognitive Science* (pp. 109–120). The MIT Press.

Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin, 139*(1), 133.

Kuhn, S., Nenchev, I., Haggard, P., Brass, M., Gallinat, J., & Voss, M. (2011). Whodunnit? Electrophysiological correlates of agency judgements. *PLoS One, 6*(12), e28657. https://doi.org/10.1371/journal.pone.0028657.

van Laarhoven, T., Stekelenburg, J. J., Eussen, M. L., & Vroomen, J. (2019). Electrophysiological alterations in motor-auditory predictive coding in autism Spectrum disorder. In *Autism research*.

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*. https://doi.org/10.1038/nn.4615.

Majchrowicz, B., & Wierzchoń, M. (2018). Unexpected action outcomes produce enhanced temporal binding but diminished judgement of agency. *Consciousness and Cognition, 65*, 310–324. https://doi.org/10.1016/j.concog.2018.09.007.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience, 5*, 39. https://doi.org/10.3389/fnhum.2011.00039.

Moore, J., & Haggard, P. (2008). Awareness of action: Inference and prediction. *Consciousness and Cognition, 17*(1), 136–144. https://doi.org/10.1016/j.concog.2006.12.004.

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological Bulletin, 143*(5), 521–542. https://doi.org/10.1037/bul0000097.

Palmer, C. J., Paton, B., Hohwy, J., & Enticott, P. G. (2013). Movement under uncertainty: The effects of the rubber-hand illusion vary along the nonclinical autism spectrum. *Neuropsychologia, 51*(10), 1942–1951.

Palmer, C. J., Paton, B., Kirkovski, M., Enticott, P. G., & Hohwy, J. (2015). Context sensitivity in action decreases along the autism spectrum: A predictive processing perspective. *Proceedings of the Royal Society of London B: Biological Sciences, 282*(1802), Article 20141557.

Perrykkad, K., & Hohwy, J. (2020a). Fidgeting as self-evidencing: A predictive processing account of non-goal-directed action. *New Ideas in Psychology, 56*, Article 100750. https://doi.org/10.1016/j.newideapsych.2019.100750.

Perrykkad, K., & Hohwy, J. (2020b). Modelling me, modelling you: The autistic self. *Review Journal of Autism and Developmental Disorders, 7*, 1–31. https://doi.org/10.1007/s40489-019-00173-y.

Perrykkad, K., Lawson, R. P., Jamadar, S., & Hohwy, J. (2020). *Perrykkad, Lawson, Jamadar, Hohwy - Action-Perception Loop Under Uncertainty Dataset*. Figshare. https://doi.org/10.26180/5ed0708f103a2.

R Core Team. (2018). R: A Language and environment for statistical computing. Retrieved from https://cran.r-project.org/.

Russell, J., & Hill, E. L. (2001). Action-monitoring and intention reporting in children with autism. *Journal of Child Psychology and Psychiatry, 42*(3), 317–328. https://doi.org/10.1111/1469-7610.00725.

Singmann, H., & Kellen, D. (2020). An introduction to mixed models for experimental psychology. In D. H. S. E. Schumacher (Ed.), *New methods in neuroscience and cognitive psychology* (pp. 4–31). New York, NY: Routledge.

Sperduti, M., Pieron, M., Leboyer, M., & Zalla, T. (2014). Altered pre-reflective sense of agency in autism spectrum disorders as revealed by reduced intentional binding. *Journal of Autism and Developmental Disorders, 44*(2), 343–352.

The Jamovi Project. (2019). jamovi (Version 0.9). Retrieved from https://www.jamovi.org.

Wen, W., & Haggard, P. (2020). Prediction error and regularity detection underlie two dissociable mechanisms for computing the sense of agency. *Cognition, 195*, Article 104074. https://doi.org/10.1016/j.cognition.2019.104074.

Williams, D., & Happé, F. (2009). Pre-conceptual aspects of self-awareness in autism spectrum disorder: The case of action-monitoring. *Journal of Autism and Developmental Disorders, 39*(2), 251–259.

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences, 1*(6), 209–216.

Zaadnoordijk, L., Besold, T. R., & Hunnius, S. (2019). A match does not make a sense: On the sufficiency of the comparator model for explaining the sense of agency. *Neuroscience of Consciousness, 2019*(1), Article niz006.

Zalla, T., Miele, D., Leboyer, M., & Metcalfe, J. (2015). Metacognition of agency and theory of mind in adults with high functioning autism. *Consciousness and Cognition, 31*, 126–138. https://doi.org/10.1016/j.concog.2014.11.001.

Zalla, T., & Sperduti, M. (2015). The sense of agency in autism spectrum disorders: A dissociation between prospective and retrospective mechanisms? *Frontiers in Psychology, 6*, 1278. https://doi.org/10.3389/fpsyg.2015.01278.

Zama, T., Takahashi, Y., & Shimada, S. (2017). The effects of trajectory and endpoint errors in a reaching movement on the sense of agency. *Psychology, 8*(14), 2321.

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: a review. *Consciousness and Cognition, 21*(1), 546–561.

This study provided many avenues of interest for future research, both to further understand how prediction error and uncertainty inform neurotypical agency judgements, and to test the robustness of these findings in a diagnosed autistic population. The next chapter details my first steps in following up on the findings presented here.

Chapter 6. Judgements of Agency and Block-wise Volatility: A Pilot Study

Though the last chapter revealed many main effects of variability on all of the dependent measures, I also emphasised that it was surprising that main effects of volatility were lacking. Though both variability and volatility are kinds of uncertainty, volatility should have distinct effects on active inference and the mechanisms of predictive processing. It is represented separately because it captures higher order expectations about changes to variability, and these higher order expectations have important effects on how prediction errors are treated in the brain.

In this chapter, the main change was to the timescale over which I manipulated volatility. Here, the volatility determines the block-wise stability of the variability. That is, there are stable blocks where the variability doesn't change (is either low or high for the whole block) and therefore the volatility is low. There are also unstable blocks, where the block context does not give the participants valid expectations for the variability trial to trial (it is randomly low or high), and thereby the volatility is high. The thought is that the change to the timescale of the volatility manipulation will make its effect stronger.

As we saw in earlier chapters, volatility might be particularly relevant for the way autistic participants interact with the world and build a self-representation. Though restricted to visual perception, Lawson, Mathys, and Rees (2017) found that autistic participants overlearn volatility leading to a reduction in learning from surprising events. This is also reflected in the review by Palmer, Lawson, and Hohwy (2017) which highlights volatility expectations which modulate learning rate and active inference as the three key features from a predictive processing perspective that will prove illuminating for autism. I also include a range of psychiatric traits questionnaires in this design that allows for transdiagnostic comparison as in Chapter 3.

This study was planned to be run in a diagnosed autistic population. This is an important step in following up on the AQ results from the last chapter. However, in early

2020, after piloting this design, the COVID-19 pandemic began, halting experimental progress in person for labs across the world. This study was dramatically affected by the pandemic. As a result, in this chapter of the thesis, I report early results from the pilot, and plans for analysis involving quantifying learning rate in an autistic population. Since the manipulations of variability and volatility can be analysed using mixed models on trialwise data as in the last chapter, the pilot data can shed some light on whether there is a stronger effect of the new volatility manipulation on movement and strategy, hypothesis selection and prediction error. However, results of AQ and other condition-wise variables are underpowered, and so should be treated with caution. I plan to complete this study when the risks of in-person data collection have reduced with the widespread use of a vaccine in Australia.

Judgements of Agency and Block-wise Volatility: A pilot study

Kelsey Perrykkad[1], Rebecca P. Lawson[2], Sharna Jamadar[3], & Jakob Hohwy[1]

1. Cognition and Philosophy Lab, Philosophy Department, School of Philosophy, Historical and International Studies, Monash University, Clayton, Australia
2. Department of Psychology, University of Cambridge, United Kingdom
3. Turner Institute for Brain and Mental Health, Monash University, Clayton, Australia

Abstract

When encountering prediction error, the brain must decide whether to ignore it as inherent, meaningless noise in the sensory signal or treat it as useful for informing future expectations. The trade-off between these two extremes is captured by the learning rate, which weights the impact of prediction error on model updating and is modulated independently by expectations of variability and volatility (uncertainty). Further, expectations for volatility have a direct impact on policy inferences, and thus change how one acts and how one infers that the consequences of their actions are one's own. Previous work manipulating volatility as relatively short-term, within-trial changes to variability in a judgement of agency paradigm showed few independent effects of volatility. In this experiment, we therefore manipulate volatility over longer time scales (blocks) to draw out its effects on movement, policy use and prediction error. Dynamic updating of learning rate in the face of volatility is also thought to be the key quantity that drives characteristic behaviours in autism spectrum disorder. As such, this experiment is designed to be run in a clinically diagnosed population. Results from the pilot data reported here suggest that volatility has independent effects on speed, acceleration and jerk of movements, proportion of dominant policy use, average prediction error, and the pattern of prediction error around the act of switching hypothesis about what the participant controls. Planned analyses on the full dataset are discussed.

Keywords: uncertainty, variability, volatility, prediction error, autism spectrum

Introduction

It is increasingly thought that the primary function of the brain is to model the world to maximise the accurate prediction of sensory input by minimising the difference between what is predicted and what occurs (*prediction error*) (Clark, 2015; Friston, 2010; Hohwy, 2013). Since the real world is so complex and changeable and the human brain is limited, there is always some residual prediction error. The brain has two possible responses to the presence of prediction error that was unaccounted for in a previous model. The first is to dismiss it as part of the normal noise in the sensory signals; best to treat as irreducible noise. The second option is to take the prediction error as evidence that the model needs to be improved, and to change predictions for next time. The trade-off between these two is determined by the *learning rate*, which acts to weight prediction error in updating the model (Mathys, Daunizeau, Friston, & Stephan, 2011). A small learning rate means the prediction error is not trusted, where as a larger learning rate means the prediction error will inform larger changes to the model. Learning rates have primarily been studied in the context of reward learning and the exploration-exploitation distinction (Behrens, Woolrich, Walton, & Rushworth, 2007; Payzan-LeNestour & Bossaerts, 2011; Rushworth & Behrens, 2008), but their role in epistemic behaviour that is not explicitly based in pragmatic rewards is relatively neglected.

Optimal modulation of the learning rate is intimately related to expectations for environmental uncertainty. Environmental uncertainty can be represented at multiple hierarchical levels, and each level has a different effect on the interpretation of prediction error. These include expectations for the environmental *volatility*, which quantifies how often the *variability* in the environment changes. The variability denotes the inverse precision of sensory input from the world – it is the lowest level of uncertainty or noise that is represented by the system (Mathys et al., 2011). Expectations about variability and volatility are separate quantities that can be independently manipulated in an experimental setting – for instance, you could have a very precise sensory signal (low variability) that changes frequently in its mean value (high volatility). The learning rate is sensitive to expectations for both kinds of uncertainty. In a world where the volatility is high, learning rate should also be high because there is an expectation that the world is changing, so taking prediction error as informative is a rational response. Conversely, when the volatility is accurately expected to be low, it is efficient to rely on priors, since there is a low probability that they have changed. This is reflected by a low learning rate. Thus, in low volatility environments, more prediction error is required to invoke model updates. With regard to variability, on the other hand, when variability is high, learning rate should be lower, since there is expected to be greater prediction error inherent in the sensory signal but since it is not meaningful, it should not result in model updating. In this way, variability and volatility have independent effects on learning rate, such that learning rate should be greatest when variability is low and volatility is high.

Importantly, formally, volatility is the inverse precision of state transitions (Parr & Friston, 2017). That is, greater volatility is associated with less precision in the expectation that performing a particular action will garner the expected result. This is because in a highly volatile world, there is no guarantee that the action will have the same effect that it did previously. In this way, uncertainty directly affects the inference of *policies* for action.

Following Perrykkad, Lawson, Jamadar, and Hohwy (2021), in this experiment, we investigate the effect of uncertainty in the form of both variability and volatility in a judgement of agency task. To make a judgement of agency, participants must infer whether their expected state transitions were successful. In this paradigm, we could look at moment-to-moment prediction error and policy use in response to uncertainty. In the original study, Perrykkad et al. (2021) found many interactions between variability and volatility on movement and strategy and prediction error dynamics, but volatility did not have noteworthy independent effects. One reason for this might be that the brain interpreted the within-trial manipulation of volatility as random samples of variability at the wider distribution without a deeper hidden cause. As such, the primary change to the paradigm in this experiment was to manipulate volatility over blocks rather than trials. In low volatility blocks participants could successfully rely on prior expectations for the variability that would be experienced in the next trial, whereas in high volatility blocks, the variability trial to trial was equally probable between two options, and so unpredictable until evidence was garnered through acting in the trial.

Expectations for volatility and appropriate modulation of the learning rate in response to these expectations is hypothesised to be the source of characteristic behavioural differences in autism (Lawson, Mathys, & Rees, 2017; Palmer, Lawson, & Hohwy, 2017). Previous experiments have mostly tested volatility learning in autism in the context of pragmatic action for explicit rewards (Goris et al., 2019; Manning, Kilner, Neil, Karaminis, & Pellicano, 2017) but to our knowledge this has not been tested in autistic participants in the context of action for epistemic gain. As such, this experiment was designed to be run in a diagnosed autistic population, and additional measures of the social responsiveness scale (Constantino & Gruber, 2012) were taken to confirm this diagnosis. Additionally, measures of intelligence quotient sought to ensure matched intelligence between autistic and neurotypical groups.

The interest in volatility learning in autism is reflective of a broader computational approach to psychiatry more generally (Fineberg, Stahl, & Corlett, 2017; Friston, Stephan, Montague, & Dolan, 2014; Montague, Dolan, Friston, & Dayan, 2012). Agency, as a subcategory of both self-representation and sensorimotor systems, has also been specified as a possible transdiagnostic dimension of psychiatric conditions by the research domain criteria framework (Insel et al., 2010; Morris & Cuthbert, 2012). We thus further measure psychiatric traits for Borderline Personality Disorder, Schizophrenia, Depression and Anxiety in this experiment. In a similar vein, we measure confidence and sense of agency ratings, which may vary by psychiatric condition in this task. A general self-concept clarity measure was taken to account for other self-representative features that might be relevant to inferring agency. Since the experiment allows us to look at many aspects of inferring agency and self-representation generally, associating dependent variables with these traits may help us narrow down which aspects of agency are relevant to psychiatric traits more generally.

Lastly, in this version of the task, unlike in Perrykkad et al. (2021), participants were able to end the trial early if they felt they had determined agency before the end of the trial. This allows us to look at the clinically relevant jumping to conclusions phenomena (Sahuquillo-Leal et al., 2019; Speechley, Whitman, & Woodward, 2010) where participants make decisive judgements on relatively little data. In the context of agency, this may give us another insight into individual variability in self-representative behaviours.

In summary, this study aimed to look at the effect of variability and volatility on prediction error, action selection, policy selection and judgements of agency in a sample of participants with a diagnosis of autism as compared to neurotypical participants. We were also interested in how the process of inferring agency is related to traits for other psychiatric conditions. Unfortunately, due to the Covid-19 pandemic, data collection for this experiment has been postponed. Here, we present data from a pilot sample of twelve mostly neurotypical participants using analyses that mirror those presented in Perrykkad et al. (2021) primarily to test whether changes in the volatility manipulation successfully bring out independent effects of volatility on movement, policy use and prediction error variables.

Methods

Participants

For this pilot dataset, data was collected from fifteen participants. Three datasets were eliminated for technical issues during recording. Of the remaining twelve datasets analysed for this chapter, ten participants had no diagnosed mental conditions, one participant reported a diagnosis of depression and one participant had a diagnosis of Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS). PDD-NOS is a historical diagnosis which has since been collapsed into Autism Spectrum Conditions. All participants were were fluent in English, had normal or corrected-to-normal vision, had no history of head injury and were right handed. The final sample consisted of four males and eight females, with an average age of 25 (range: 20:41).

Procedure

This study was approved by Monash University Human Research Ethics Committee (Project Number 13211) and was conducted in accordance with the relevant guidelines and regulations.

Prior to attending the experimental session in person, participants completed a consent form, demographics survey and eight non-diagnostic psychometric surveys online: the Autism-Spectrum Quotient (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), Borderline Personality Questionnaire (Poreh et al., 2006), Schizotypal Personality Questionnaire (Raine, 1991), Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), Beck Anxiety Inventory (Beck, Epstein, Brown, & Steer, 1988), Illusory Beliefs Inventory (Kingdon, Egan, & Rees, 2012), Self Concept and Identity Measure (Kaufman, Cundiff, & Crowell, 2015), and the Self Concept Clarity Scale (Campbell et al., 1996).

On arriving for the in-person session, participants completed another consent form, and the Test of Premorbid Functioning (TOPF), which is an updated version of the Wechsler Test of Adult Reading and can be used as a short estimate of verbal intelligence quotient (VIQ) (Mathias, Bowden, & Barrett-Woodbridge, 2007). This measure is unavailable for one participant in this pilot dataset. Additional estimates of full scale intelligence quotient (FSIQ) were obtained through the demographic questionnaire following Crawford and Allan (1997). As the current sample is too small to reliably investigate relationships between the task and these demographic and psychometric measures, we report mean and spread in Table 1, below.

Table 1- Demographics Summary

Demographic	**Mean**	**Range**	**1st Qu.**	**3rd Qu.**
Age	25	20:41	22	25
Autism-Spectrum Quotient	19	10:30	14	26
Borderline Personality Questionnaire	15	1:32	8	21
Schizotypal Personality Questionnaire	15	2:29	8	26
Beck Depression Inventory	9	0:41	2	10
Beck Anxiety Inventory	7	1:15	4	12
Illusory Beliefs Inventory	57	30:84	36	71
Self-Concept Clarity	42	29:57	38	46
Self-Concept and Identity Measure	73	38:137	60	79
Test of Premorbid Functioning VIQ Estimate	114	87:126	113	122
Demographic-based FSIQ Estimate	104	90:117	98	112

Testing was conducted in a quiet, darkened room. Participants were seated at a table with a chin rest set to a comfortable height, 84cm from the screen, and approximately 55cm from eye to eye-tracking camera. The task was completed using a computer mouse in the participant’s right hand which was hidden in a curtained box (base dimensions: 32cm wide x 30cm deep). Their opposite hand gave end-of-trial responses using the numbers on a keyboard. Participants had self-timed breaks between blocks.

Experimental Task Design

The procedure and design of this experiment are similar to Perrykkad et al. (2021). The *Squares Task,* a variant of which is the primary task in this experiment, is the most common way of testing judgment of agency in an autistic population (Grainger, Williams, &

Lind, 2014; Perrykkad & Hohwy, 2020; Russell & Hill, 2001; Williams & Happé, 2009). Stimuli were presented using Psychtoolbox-3.0.14 version beta in Matlab 2017b (Mathworks, Natick, Massachusetts) on a 1920x1080 screen (60Hz refresh rate).

At the beginning of each trial, after an initial fixation cross, eight grey squares (100px^2) appeared in an array on a lighter grey background. Each square had an unique, luminance-matched (Willenbockel et al., 2010) pattern, making it easier for the participant to track individual squares as they moved. All the squares moved when the mouse was moved and all the squares stopped when the mouse stopped, so participants had to move in order to accurately complete the task. Participants were given 15sec to identify the target square which they controlled but could also press the return key or spacebar to end the trial any time after the first three seconds. Distracter squares moved at a random angle offset from the vector of mouse movement, and this angle was also independently and randomly changed (and smoothly transitioned) five times in each trial. This means that each distracter square appeared to turn five times when the participant did not initiate a turn, breaking any illusion of control resulting from motor adaptation. If the participant had not ended the trial after the 15sec, all squares froze and were numbered, and prompted an unspeeded numerical response from participants indicating which square they controlled or ‘0’ if they thought they controlled none of the squares. They also answered “How confident are you that your answer is correct?” on a continuous sliding scale from ‘Not at all’ to ‘Very’, and “How much agency/control did you feel in that trial?” on a sliding scale from ‘None’ to ‘A lot’. These responses were also recorded using the keyboard. Values ranged from -100 to 100 for sliding scale responses.

Some jitter was added to all squares (*variability*), such that depending on the condition, there was a range (95% CI) of random noise around the mean angle input by the mouse (or the distractor offset). While this experiment also primarily manipulates variability and volatility in the movement of the squares, unlike Perrykkad, Lawson, Jamadar, and Hohwy (2020), there were no changes to the variability distribution within trials. In *low variability* trials, jitter was sampled from a 20˚ 95% confidence interval around the angle of mouse movement, and for *high variability* trials*,* the sampled distribution was ±100˚. Volatility was manipulated block-wise. In *stable* blocks, the variability on every trial was either low or high and did not change. In *unstable* blocks, each trial had a 50:50 chance of being high or low variability and were randomly ordered. In the first four trials of every unstable block, there were two of each level of variability. Each block consisted of twelve normal *control* trials, and four *no-control* trials in which all the eight squares were distracter squares. Participants completed four unstable blocks, two low variability, and two high variability stable blocks (four stable blocks) in a random order. This gave a total of 128 task trials per participant.

Prior to completing the task blocks, participants engaged in an interactive instruction demonstration. Participants then completed a practice block containing sixteen total trials consisting of all trial types, which was excluded from all analyses. Mean accuracy in the practice block was 58.33% (SD=14.43).

At the end of the experiment there was a short motor control task. In this task, participants were asked to move a perfectly controllable square along a white path as fast and as accurately as possible. There were 10 predesignated paths ranging in length and complexity. This task allowed us to quantify participants' ability to execute motor intentions.

Analysis

Behaviour

Behavioural measures of *accuracy, time spent moving, speed, acceleration and jerk of movements, turn counts* and *dominant policy use* were calculated in the same way as described in Perrykkad et al. (2021). Additional behavioural measurements included *trial* *duration* (since participants could end the trial early), *confidence* and *sense of agency* (from continuous slider responses).

Eyetracking

Binocular eyetracking data was collected using the SR Research Eyelink 1000 system. For each participant, binocular thirteen-point calibration was conducted; where calibration was unsuccessful for both eyes, one eye was used. The screen x and y coordinates were preprocessed for analysis. Preprocessing involved removing any values outside of the screen bounds, interpolating eyeblinks (as defined by pupil size outside of 1.5 standard deviations below to 2 standard deviations above participant average pupil size), applying a Hanning window of 15 samples (93% overlap) to smooth the eyetrace, and replacing temporarily lost values in one eye with valid data from the other eye (including for whole trials if one eye was excessively noisy). Data was then epoched into trials, and downsampled to match the stimuli framerate for alignment with behavioural data. Trials with poor signal were defined as those with more than 30% of the samples interpolated in both eyes, or whose recorded behavioural data was outside of two standard deviations above or below the participants mean recorded trial length (as the source of these outliers could not be identified). For the final sample of participants, there were a maximum of 40 poor-signal trials (mean = 12.7). Poor-signal trials were removed from all analyses (including behavioural only dependent variables above).

The square the participant was looking at was determined by the biased-nearest-object method (Perrykkad et al., 2020). Times of *hypothesis switch* from one square to another were defined as any change in the looked-at square that lasts longer than one frame.

As in Perrykkad et al. (2020), the Euclidean distance between the expected location (had the stimuli followed the mouse) and the actual location of the hypothesised square was calculated as a behavioural proxy for *prediction error*. This means that the prediction error is contingent on movement speed and the quality of their hypothesis. Due to the manipulation, this means that low variability trials accrue less prediction error on average than high variability trials. As in previous work, prediction error measures included *average prediction error* across each trial and the *slope of prediction error* in each condition. To estimate the slope of prediction error across trials of inconsistent length the prediction error for all trials within each participant were downsampled to the minimum trial length for that participant, and mean slope in each condition was fit to averages over this downsampled data.

Statistics

Given that this dataset is a pilot dataset, we present primarily results which are directly comparable to our previous work. Statistical analysis presented here follows the same mixed model structures presented in (Perrykkad et al., 2021). All statistical analyses were conducted as Mixed Linear Models (MLM) using Jamovi version 1.6.15 and the GAMLj module (Gallucci, 2019; R Core Team, 2020; The Jamovi Project, 2019). Trial-wise data was used for all dependent variables except prediction error slopes and ERPEs for which condition-wise data was used. Variability and volatility were modelled as simple fixed effect factors, and AQ score was modelled as a continuous fixed effect. All interactions between fixed effects were included. By-participant random intercepts were included to address the non-independence of subject-level observations across trials and capture individual variability in task performance. See Table 1 for additional covariates for each model. Degrees of freedom are reported as estimated by the Satterthwaite method. *Post-hoc* analyses were conducted with Bonferroni correction for multiple comparisons and post-hoc p-values are reported with this correction. While we have included AQ as a fixed effect in the results reported here, readers should be very cautious to interpret statistics relating to AQ due to the very small sample size for participant level measures. This is similarly true for those effects run on condition-wise data as these are also likely underpowered. Other, trial-wise data is more reliable, but should still only be given weight appropriate to its status as pilot data.

We also report results for three new models which individually test the relationship between the new variables of *sense of agency*, *confidence*, and *trial duration*. For these models, each of these served as the dependent variable, and the mixed models included only fixed effects of variability and volatility and their interaction. A random intercept for participant was also included as above.

Results

Full statistical results from the mixed models which match previous analysis from Perrykkad et al. (2021) can be found in Table 2. Blue text indicates results that match with the previous study, and red results indicate differences from the previous study. Summaries of the more reliable trial-wise results will be further specified in the sections that follow (ie. not AQ or slope data).

Accuracy and Post-trial Responses

On average, accuracy for this pilot experiment was 66.4%. The mixed model for accuracy showed a main effect of variability (F(1,1367) = 8.97, p = 0.0028), such that low variability trials had a 7% increase in accuracy compared to high variability trials.

The sliding scale responses following the trials were scored on a ±100 point scale. On average, confidence was rated at +20.22, and sense of agency at +2.60 across all trials (including no control trials). The mixed models for confidence and sense of agency also each showed main effects of variability (confidence: F(1,1369) = 105.77, p = $6.0\text{x}10^{24}$; sense of agency: F(1,1370) = 99.52, p = $1.13\text{x}10^{-22}$). Low variability showed greater confidence by 18 points on average, and sense of agency was rated 23 points greater compared to high variability trials.

Movement and Strategy

On average, trials were ended after 12.05s (out of the possible 15s). The mixed model looking at the effect of uncertainty on participant controlled trial duration showed a main effect of variability, such that high variability trials lasted 1.11s longer on average than low variability trials (F(1,1369) = 74.43, p = 1.72x10^{-17}). The analysis on how much of this time participants spent moving, controlling for time to initial movement and total trial duration, showed a significant main effect of variability (F(1,1364) = 28.51, p = 1.09x10^{-7}) and a trending main effect of volatility, which is only noted here because of the pilot nature of this report (F(1,1364) = 3.56, p = 0.060). Participants moved for more of the trial when variability was high (by 187ms on average) and in trials in stable contexts (low volatility, by 64ms).

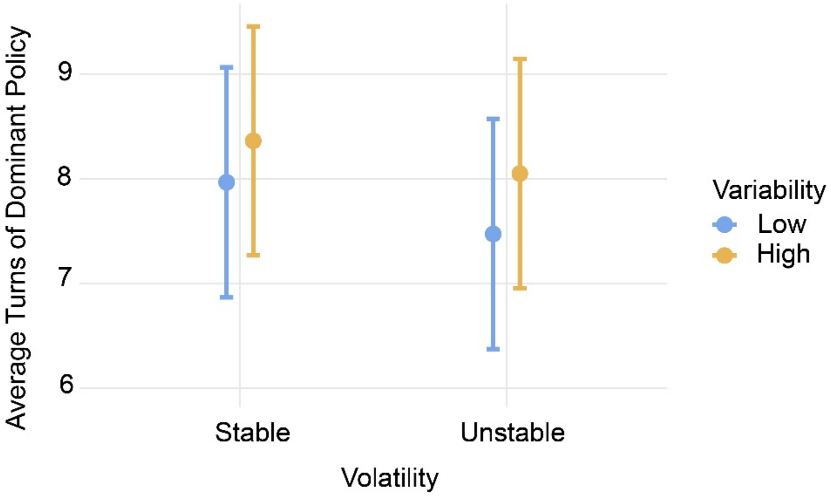
Regarding speed, the mixed model analyses showed a main effect of both variability (F(1,1366) = 52.22, p = 8.23x10^{-23}) and volatility (F(1,1366) = 5.12, p = 0.024), and a significant interaction between the two (F(1,1366) = 4.57, p = 0.033). Post-hoc analyses revealed that in general, participants moved faster in low variability trials and in high volatility blocks. But the interaction revealed that while there was no difference between low variability trials in stable and unstable blocks, participants did move faster in high variability trials in unstable (high volatility) blocks than high variability trials in stable blocks (t(1366) = -3.18, p = 0.0090). The mixed model for acceleration also showed a main effect of both variability (F(1,1366) = 47.66, p = 7.76x10^{-12}) and volatility (F(1,1366) = 5.73, p = 0.017), and a significant interaction between the two (F(1,1366) = 14.55, p = 1.42x10^{-4}). Posthoc analyses showed that in general, both high variability and high volatility were associated with greater acceleration. There was no difference in acceleration between volatility contexts when variability was high (t(1366) = 1.027, p = 1.00) but when variability was low, there was greater acceleration in unstable than stable blocks (t(1366) = -4.30, p = 1.12x10^{-4}). Similarly, when volatility was high (unstable blocks), there was no difference between variability conditions (t(1366) = -2.19, p = 0.17). But in stable blocks, acceleration was greater in high variability than low (t(1366) = -7.60, p = 3.20x10^{-13}). The mixed model analysis of jerk also showed main effects of variability (F(1,1367) = 7.09, p = 0.0079) and volatility (F(1,1367) =

4.30, p = 0.038) and a significant interaction between them (F(1,1366) = 5.16, p = 0.023). Both high variability trials and high volatility (unstable) blocks are associated with greater jerk. Post-hoc analyses of the interaction shows that low variability trials in stable blocks has lower jerk than any other condition combination, which have equal jerk (t(1367) = -3.01-3.50, p = 0.0049-0.016). Also noted is a trending interaction between volatility and AQ (F(1,1367) = 3.79, p = 0.052) which appears to be a result of a greater impact of high volatility on participants with increasingly greater AQ score (values look similar across AQ in stable blocks, but jerk increases with greater AQ in high volatility).

The mixed model on turning behaviour showed no reliable effects of uncertainty.

The mixed model investigating the effect of uncertainty on dominant policy use showed main effects of both variability (F(1366) = 5.72, p = 0.017) and volatility (F(1365) = 3.93, p = 0.048). Stable blocks (low volatility) were associated with more dominant policy use than unstable blocks. Conversely, low variability was associated with less dominant policy use. This pattern can be seen in Figure 1.

Figure 1 - Dominant Policy Use Across Conditions



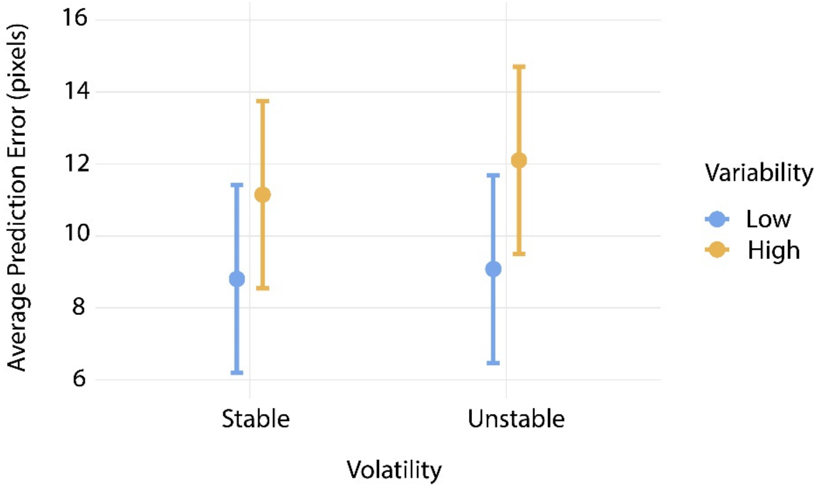
In each trial, participants switched hypotheses an average of 25.65 times per trial (SD = 13.83). The mixed model which looked at the relationship between number of hypothesis switches per trial and uncertainty showed a main effect of variability (F(1,1366) =15.25, p =

9.87×10^{-5}) such that greater variability was related to fewer hypothesis switches (by an average of 2 per trial). There was a trending main effect of volatility in the reverse direction (F(1,1366) = 2.94, p = 0.087) – there was one more hypothesis switch per trial on average in low volatility (stable) blocks than high.

Prediction Error

The mixed model investigating the effect of uncertainty on average prediction error over each trial showed main effects of both variability (F(1,1358) = 116.18, p = 4.75×10^{-26}) and volatility (F(1,1358) = 6.15, p = 0.013), while controlling for accuracy and all its interactions. In general, lower average prediction error was associated with low variability trials and low volatility (stable) blocks. See Figure 2.

Figure 2 - Average Prediction Error Across Conditions



Raw data from the hypothesis switch ERPE can be seen in Figure 3a, with the relevant time bins used for statistical analysis depicted beneath. The statistical analysis showed main effects of time bin (F(1,188) = 37.69, p = 3.80×10^{-23}) and both variability (F(1,197) = 15.60, p = 1.09×10^{-4}) and volatility (F(1,192) = 4.08, p = 0.045), and significant interactions between time bin and variability (F(1,188) = 4.06, p = 0.0035) and between variability and volatility (F(1,189) = 4.41, p = 0.037). Post hoc analyses showed that there was greater prediction error around hypothesis switches in unstable blocks (high volatility) and low variability trials. For the main effect of time bin, post hoc analyses showed no difference between time bins 1, 4 and 5. All other time bins were significantly increased from these, with maximum differences between the time of the event and all other time bins (t(188) = -3.96- -10.35, p < 0.001). This demonstrates a clear peaking pattern around hypothesis switches. For the interaction between variability and volatility, only when variability was low was there no difference between volatility conditions (t(190) = -0.001, p = 1.00) but when variability is high, unstable blocks show more prediction error in the epoch around hypothesis switches than stable blocks (high volatility>low volatility)(t(190) = -2.89, p = 0.026). The interaction between variability and time bin is due to a difference between variability conditions which is significant in time bin three (the time of the event). While in all other time bins the two variability conditions are equal in prediction error (p = 1.00 for all), at the time of the hypothesis switch, low variability trials have greater prediction error than high variability trials (t(192) = 5.41, p = 8.43x10^{-6}). Average prediction error data used for the mixed model is depicted in Figure 3b.

Figure 3 - Hypothesis Switch ERPE

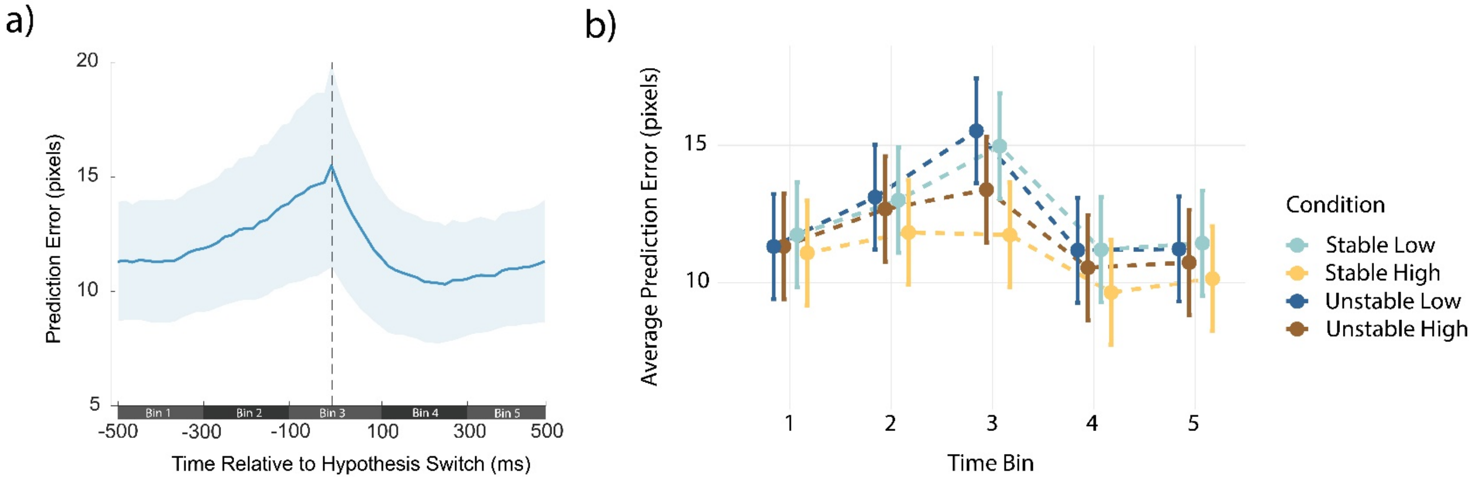


Table 2 - Significant Results Summary to Compare with Perrykkad et. al. (2021)

For variables, M.E.=Main Effect, *=Interaction, For results, *=p ≤ 0.05, **=p ≤ 0.01, ***=p ≤ 0.001 (Post-hoc values Bonferroni corrected for multiple comparisons), Var=Variability, Vol=Volatility, AQ=Autism Quotient, T1-5=Time bins 1-5, diff.=difference. Blue = same as previous experiment, Red = different from previous experiment, -- = no effect was present where one was in the previous experiment, The effect of AQ and slope models are likely underpowered (greyed)

	Dependent Variable	Additional Covariate	M.E. Variability	M.E. Volatility	M.E. AQ	Var*Vol	Var*AQ	Vol*AQ	Var*Vol*AQ
Task	Accuracy		** low>high			--			
Movement and Strategy	Time Spent Moving	Time to Movement, Trial Duration	*** low<high				***** *smaller var diff at larger AQ*		--
	Speed		*** low>high	* stable<unstable		* low var unstable>both stable & unstable high; low stable>high unstable and high var stable; high var stable < high unstable			
	Acceleration		*** low<high	* stable<unstable	--	*** low var stable < all others; high var stable>low var unstable			
	Jerk		** low<high	* stable<unstable		* low var stable < all others			
	Turning Behaviour		--			--	*** *Only at low AQ high>low*	**** *Only in low AQ stable>unstable Only in high AQ stable<unstable*	
	Dominant Policy Use	Number of Turns	* low<high	* stable>unstable			***** *At high AQ no difference between var*	--	--

	Hypothesis Switches		*** low>high			--	* *At low AQ no diff between var*	** *At high AQ only, stable<unstable*	
Prediction Error	Average Prediction Error	Accuracy (and all interactions)	*** low<high	* stable<unstable			--		
	Condition-wise Slope		--	--					
			M.E. Agency	**M.E. Accuracy**	**M.E. AQ**	**Accuracy *Agency**	**Accuracy*AQ**	**Agency*AQ**	**Accuracy *Agency*AQ**
	Agency-wise Slope		--			* in no agency judgement there is no diff. between correct and incorrect; no sig. pairwise differences			--
			M.E. Variability	**M.E. Volatility**	**M.E. AQ**	**Var*Vol**	***Var*AQ***	***Vol*AQ***	***Var*Vol*AQ***
	Hypothesis Switch ERPE	Average Prediction Error; Hypothesis Switches	*** low>high	* low<high		** When stable, low var>high var; in high var, stable<unstable; stable high var < unstable low var			

	M.E. Time Bin	AQ*Time Bin	Var*Time Bin	
...	*** T2 and T3 (event) are>T1, T4, T5 which are equal; T3>T2	--	** Only in low var T2>T1, bigger var diff. at T3	...

Discussion

In this pilot study, we investigated how variability in the mapping between actions and outcomes and the stability of the block-wise context (volatility) affected ongoing decisions regarding action selection and broader policy use (e.g. when to switch hypothesis) in a judgement of agency task. We were also able to measure a behavioral proxy for prediction error, which enabled us to look at relationships between these actions and the prediction error dynamics. The ultimate aim of the design is to test how these elements of the action-perception loop differ in clinically diagnosed autistic participants as compared to a neurotypical sample. However, the pilot results from mostly neurotypical adults presented here cannot answer this further question. The early results can speak to the interplay between prediction error and action in determining agency in a neurotypical population.

The design of this experiment was based on Perrykkad et al. (2021). In general, it is promising that many of the important effects replicate. For instance, in both experiments there is a clear peak in prediction error around hypothesis switch events modulated in the same way by both variability and volatility. Variability modulation of average prediction error over a trial regardless of whether the participant was correct also replicated but we see a new main effect of volatility.

Since the main change between these two experiments was to the volatility manipulation, it is also encouraging to see many new main effects of volatility, some in the opposite direction to the effect of variability (i.e. speed, dominant policy use, prediction error in the hypothesis switch ERPE epoch). These main effects demonstrate that greater volatility is associated with faster movement, greater acceleration and jerk, less dominant policy use and greater average prediction error. The magnitude of the differences in the main effects of volatility are still weaker than those of variability (while not a reliable measure, see, for example, the differences in average p-values in each of the columns in *Table 2*), suggesting that the context effects are more subtle than the ongoing variability in action-outcome mapping.

There were a few additional changes made to the paradigm to enable answering some new questions. Additional behavioural measures included participant ratings of confidence and sense of agency on each trial. Our results show that lower levels of uncertainty in the form of variability do affect both these explicit ratings of confidence and sense of agency. The negative effect of increased variability on accuracy of judgements of agency replicated Perrykkad et al. (2021). These additional measures show that not only are participants poorer at judging agency under high variability, they also feel less agency and are less confident in their judgement.

The results of the pilot study are promising. Changes to the design of the experiment also open up new avenues for future analysis that were not conducted on the small dataset presented here.

Allowing participants to end the trial at their own volition means that we can look at jumping to conclusions phenomena. That is, when do participants decide that they have garnered enough data to make a decision? In a clinical sample, we could also look at whether a diagnosis of autism is associated with earlier or later trial termination. Early results from the pilot did show that variability affected when participants decided to end the trial, but volatility did not. This measure may also prove interesting to correlate with other psychiatric trait measures taken in this experiment.

By looking at how accuracy or agency judgements are influenced by the condition of previous trials, we may be able to derive measures of learning rate (see Rushworth and Behrens (2008) Figure 4a for an example of this kind of analysis). This is because a greater influence of trials further back in time indicates a lower learning rate and greater reliance on priors. Large influence of recent trials but less influence of older trials would indicate a higher learning rate. Use of the hierarchical Gaussian filter to derive an estimate of learning rate may also be available in this design (Mathys et al., 2014). An individual's learning rate represents the amount their model is updated when they encounter prediction error, and is related to expectations for precision and volatility. It is also one of the most important quantities in predictive processing accounts of autism, so may prove particularly illuminating in a clinically diagnosed sample.

In this version of the experiment, the squares were luminance matched, on a grey, rather than black background and a fixation cross preceded each trial. This enables the valid use of pupillometry data from the eye tracking, since there is a period during which to obtain pupil baseline and there is no inherent difference in the brightness of the squares that would confound pupil size data. As a result, in this version of the task, we will be able to compare the use of pupil size as a measure of gain or attention as it relates to precision of prediction error (Lawson et al., 2017; Vincent, Parr, Benrimoh, & Friston, 2019), and our behavioural prediction error proxy. This would give us a better insight into the neural dynamics of prediction error that includes the updating of priors in different contexts (such as our blocks).

Conclusion

In conclusion, results from this pilot study suggest that greater volatility is associated with faster movement, greater acceleration and jerk, less dominant policy use and greater average prediction error, independently of the variability experienced. These results suggest that the manipulation of volatility over longer timescales was a successful alteration to bring its effects to the fore. In addition, this pilot experiment replicates many of the primary effects reported in Perrykkad et al. (2021). Further avenues using this design are similarly exciting.

References

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5-17. doi:10.1023/A:1005653411471

Beck, A., Epstein, N., Brown, G., & Steer, R. (1988). An Inventory for Measuring Clinical Anxiety: Psychometric Properties.

Beck, A., Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. (1961). Beck depression inventory (BDI). *Arch Gen Psychiatry, 4*(6), 561-571.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature neuroscience, 10*(9), 1214-1221.

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of personality and social psychology, 70*(1), 141.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*: Oxford University Press.

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale: SRS-2*: Western Psychological Services Torrance, CA.

Crawford, J. R., & Allan, K. M. (1997). estimating premorbid WAIS-R IQ with demographic variables: Regression equations derived from a UK sample. *The Clinical Neuropsychologist, 11*(2), 192-197. doi:10.1080/13854049708407050

Fineberg, S. K., Stahl, D. S., & Corlett, P. R. (2017). Computational Psychiatry in Borderline Personality Disorder. *Current Behavioral Neuroscience Reports, 4*(1), 31-40. doi:10.1007/s40473-017-0104-y

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138.

Friston, K., 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

Gallucci, M. (2019). GAMLj: General analyses for linear models [jamovi module]. Retrieved from https://gamlj.github.io/

Goris, J., Silvetti, M., Verguts, T., Wiersema, J. R., Brass, M., & Braem, S. (2019). Autistic traits are related to worse performance in a volatile reward learning task despite adaptive learning rates.

Grainger, C., Williams, D., & Lind, S. E. (2014). Online Action Monitoring and Memory for Self-Performed Actions in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 44*, 1193-1206.

Hohwy, J. (2013). *The predictive mind*: Oxford University Press.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry, 167*(7), 748-751. doi:10.1176/appi.ajp.2010.09091379

Kaufman, E. A., Cundiff, J. M., & Crowell, S. E. (2015). The Development, Factor Structure, and Validation of the Self-concept and Identity Measure (SCIM): A Self-Report Assessment of Clinical Identity Disturbance. *Journal of Psychopathology and Behavioral Assessment, 37*(1), 122-133. doi:10.1007/s10862-014-9441-2

Kingdon, B. L., Egan, S. J., & Rees, C. S. (2012). The Illusory Beliefs Inventory: A New Measure of Magical Thinking and its Relationship with Obsessive Compulsive Disorder. *Behavioural and Cognitive Psychotherapy, 40*(1), 39-53. doi:10.1017/S1352465811000245

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature neuroscience*. doi:10.1038/nn.4615

Manning, C., Kilner, J., Neil, L., Karaminis, T., & Pellicano, E. (2017). Children on the autism spectrum update their behaviour in response to a volatile environment. *Developmental Science, 20*(5).

Mathias, J. L., Bowden, S. C., & Barrett-Woodbridge, M. (2007). Accuracy of the Wechsler Test of Adult Reading (WTAR) and National Adult Reading Test (NART) when estimating IQ in a healthy Australian sample. *Australian Psychologist, 42*(1), 49-56. doi:10.1080/00050060600827599

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian Foundation for Individual Learning Under Uncertainty. *Frontiers in human neuroscience, 5*, 39. doi:10.3389/fnhum.2011.00039

Mathys, C., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in human neuroscience, 8*, 825.

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends Cogn Sci, 16*(1), 72-80. doi:10.1016/j.tics.2011.11.018

Morris, S. E., & Cuthbert, B. N. (2012). Research Domain Criteria: cognitive systems, neural circuits, and dimensions of behavior. *Dialogues in clinical neuroscience, 14*(1), 29.

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian Approaches to Autism: Towards Volatility, Action, and Behavior. *Psychological bulletin*.

Parr, T., & Friston, K. J. (2017). Uncertainty, epistemics and active inference. *Journal of The Royal Society Interface, 14*(136), 20170376.

Payzan-LeNestour, E., & Bossaerts, P. (2011). Risk, Unexpected Uncertainty, and Estimation Uncertainty: Bayesian Learning in Unstable Settings. *PLoS computational biology, 7*(1), e1001048. doi:10.1371/journal.pcbi.1001048

Perrykkad, K., & Hohwy, J. (2020). Modelling Me, Modelling You: the Autistic Self. *Review Journal of Autism and Developmental Disorders, 7*, 1-31. doi:10.1007/s40489-019-00173-y

Perrykkad, K., Lawson, R. P., Jamadar, S., & Hohwy, J. (2021). The effect of uncertainty on prediction error in the action perception loop. *Cognition, 210*, 104598. doi:https://doi.org/10.1016/j.cognition.2021.104598

Perrykkad, K., Lawson, R. P., Jamadar, S. D., & Hohwy, J. (2020). The Effect of Uncertainty on Prediction Error in the Action-Perception Loop. *bioRxiv*, 2020.2006.2022.166108. doi:10.1101/2020.06.22.166108

Poreh, A. M., Rawlings, D., Claridge, G., Freeman, J. L., Faulkner, C., & Shelton, C. (2006). The BPQ: a scale for the assessment of borderline personality based on DSM-IV criteria. *Journal of personality disorders, 20*(3), 247-260.

R Core Team. (2020). R: A Language and environment for statistical computing. Vienna, Austria: R foundation for statistical Computing. Retrieved from https://www.R-project.org/

Raine, A. (1991). The SPQ: A Scale for the Assessment of Schizotypal Personality Based on DSM-III-R Criteria. *Schizophrenia Bulletin, 17*(4), 555-564. doi:10.1093/schbul/17.4.555

Rushworth, M. F., & Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat Neurosci, 11*(4), 389-397. doi:10.1038/nn2066

Russell, J., & Hill, E. L. (2001). Action-monitoring and Intention Reporting in Children with Autism. *Journal of Child Psychology and Psychiatry, 42*(3), 317-328. doi:10.1111/1469-7610.00725

Sahuquillo-Leal, R., Ghosn, F., Moreno-Giménez, A., Almansa, B., Serrano-Lozano, E., Ferrín, M., . . . García-Blanco, A. (2019). Jumping to conclusions in autism: integration of contextual information and confidence in decision-making processes. *European Child & Adolescent Psychiatry*. doi:10.1007/s00787-019-01409-2

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 : JPN, 35*(1), 7-17. doi:10.1503/jpn.090025

The Jamovi Project. (2019). jamovi (Version 0.9). Retrieved from https://www.jamovi.org

Vincent, P., Parr, T., Benrimoh, D., & Friston, K. J. (2019). With an eye on uncertainty: Modelling pupillary responses to environmental volatility. *PLoS Comput Biol, 15*(7), e1007126. doi:10.1371/journal.pcbi.1007126

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior research methods, 42*(3), 671-684.

Williams, D., & Happé, F. (2009). Pre-conceptual aspects of self-awareness in autism spectrum disorder: The case of action-monitoring. *Journal of Autism and Developmental Disorders, 39*(2), 251-259.

Acknowledgements

The authors appreciate the time and effort of their participants without whom this research would not be possible. Additional thanks to eventual coauthors Rebecca Lawson, for extensive discussions about experimental design, to Beth Johnson, for future clinical support, and to Sharna Jamadar, for methodological assistance and resources, and to Jakob Hohwy, for discussions at all stages of the project. This pilot would have been impossible without the help from Disha Sasan and Alexandria Wulff, research assistants, who collected the data presented here. Special mention also to Dalia Maron, research assistant, who committed to ongoing data collection before COVID-19 changed all our plans.

Early results from the pilot study are in many places reassuringly similar to those from Chapter 5. Of particular note given what is to come, the hypothesis switch in Figure 3 has the same peaked pattern as in the previous experiment. There were also some of the expected differences given the change in manipulation.

The next chapter is the last of the squares task chapters. In the first two designs, I investigated how participants react to environmental uncertainty. However, in both designs, the statistical structure of the environment was imposed upon participants in a rather unnatural way. In the real world, organisms choose the environment they occupy, at least among available options. In the next chapter, the experimental design affords participants this kind of choice.

Chapter 7. The Beach Task: Environmental Niche Selection Under Uncertainty

"the construction of the self coincides with construction of a niche"

- Constant, Bervoets, Hens, and Cruys (2020, p. 5)

Throughout the thesis so far, I have emphasised the importance of action for the cognitive maintenance and construction of the self. I have also highlighted that the environment organisms occupy is uncertain and changing, and so this is an important consideration when investigating how selves are constructed and maintained.

So far, however, in these experiments about building a self in uncertain environments, I have ignored the fact that the organism also contributes to environmental changes. This happens both explicitly – e.g. choosing to sit in a noisy café to work instead of the quiet office, and implicitly – e.g. the formation of 'desire paths' when many people choose to walk through grass to get to the building from the bus stop quicker (Constant, Ramstead, Veissière, Campbell, & Friston, 2018). This leads some scholars to the extended mind hypothesis (Clark & Chalmers, 1998), however, it is not a necessary consequence of the view (Sterelny, 2010).

Recent work in active inference has turned to interpersonal environmental scaffolding and its role in the intergenerational propagation of cultural norms (cf. Fabry (2021)), including those that provide environmental cues that act as physical cognitive heuristics in a process termed 'uploading'(Constant et al., 2018). Interactions with, and changes to, other hidden causes in the environment can also be a policy used to minimise prediction error and respond to uncertainty.

Due to the COVID-19 pandemic, it is also the first time the squares task has been run online. This came with its own set of practical considerations that affected the design of the experiment. This study is a first step towards empirically testing the active use of environments as an uncertainty reducing policy in the context of inferences about the self.

The Beach Task: Environmental Niche Selection Under Uncertainty

Kelsey Perrykkad[1], Jonathan Robinson[1] and Jakob Hohwy[1]

1. Cognition and Philosophy Lab, Philosophy Department, School of Philosophical, Historical and International Studies, Monash University

Abstract

In order to act effectively, organisms must build and maintain mappings between possible policies for action and their expected outcomes across many environments. These environments are not independent of the organism's actions, but rather, as agents, organisms actively select, interact with, and switch between environments in an effort to maximise epistemic gain and utility. In many previous experiments, the participant does not have control over the environment they inhabit, even if they are allowed to change environments. In this online behavioural experiment, participants freely move between two environments to complete a judgement of agency task. Judgements of agency are made when the agents reflectively and consciously judge that they are able to effectively carry out such actions to produce expected outcomes. Here, we use a behavioural measure of prediction error to test how participants test their hypotheses of agency. In the 'sand' environment, there is greater irreducible uncertainty in the mapping between actions and outcomes. In the 'water' environment, the agent's model of the environment must be more complex, but successfully modelling this complexity affords less uncertainty in the action-outcome mapping. Results show that participants prefer the 'sand' environment, and judge agency more accurately and with more confidence with increased time spent in this environment. Further, we show that participants actively switch between the two environments (in either direction) as an additional policy, which facilitates management of the quality of action-outcome mappings. Finally, we show that the deployment of these policies is modulated by autism traits.

Keywords: environmental niche selection, environmental uncertainty, agency, autism, prediction error

Introduction

The classic dichotomy between exploration and exploitation is often motivated by appeals to maximizing pragmatic goods – for example, the choice between continuing to forage for berries in this bush or set out into an uncertain landscape to find a more fruitful bush (Hills et al., 2015). What is less often considered is that organisms also explore in order to optimise their learning environment, and minimise uncertainty within this environment. Under predictive processing accounts, exploration can be understood as a form of *epistemic action* (Friston et al., 2016; Friston et al., 2015), that reduces uncertainty about states of the world, including states of the agent themselves. Predictive processing is a popular theoretical framework that unifies cognition, including both perception and action, under the common imperative of minimising uncertainty, often understood as the long-term mismatch between expectations and sensory feedback from the world (or *prediction error*) (Clark, 2015; Friston, 2010; Hohwy, 2013). The role of epistemic action in ordinary behaviour is not well understood, and research in this area is particularly lacking for uncertainty reduction about the causal power of the agent itself.

Unlike some organisms (e.g. plants), humans exercise a great deal of agency over their environmental niche. We are able to place ourselves in any number of increasingly complex (or simple) environments as it suits our needs. In different environments, the mapping between actions and outcomes will change due to hidden environmental influences. Acting in some environments might be more informative than acting in others, and optimal actions (with a good or precise mapping between policies and outcomes) may change based on environmental features. Both *environment selection* and changing between environments (*environment switching*) are understood here as *policies.* Policies are defined as a set of actions that are inferred by the individual to reliably reduce prediction error. Expectations of state transitions based on policy selection depend on one’s model of the environment, and must update when the organism changes environments. We can also use the act of moving between environments itself as a policy for uncertainty reduction and prediction error minimisation. Testing a hypothesis across multiple environments can be *self-evidencing* in the sense that consistent data obtained from actions generated under that hypothesis across environments garners further evidence that the initial model is accurate (Hohwy, 2016).

Importantly, the optimal use of policies is inferred by the agent. To investigate how participants use environments to their epistemic advantage in an experimental setting, the participants must be able to vary use of the policies of interest freely. In many experiments about environmental exploration, the experimenter dictates the next environment, thereby limiting the active employment of switching policies based on prior beliefs about the nature of available environments (see Mehlhorn et al. (2015, p. 194) for discussion of exploration as choosing “any other option at random”). Often too in these paradigms, once a participant has chosen to move to a new environment, they are prohibited from returning to previous environments (as in experiments where each new ‘trial’ represents a new environment e.g. Hutchinson, Wilke, and Todd (2008)). Similarly, sometimes while participants can technically return to the previous environment, in the meantime it has changed so dramatically that returning to it is not equivalent to choosing the old statistical environment (as in many multi-armed bandit tasks see Cohen, McClure, and Yu (2007) for review). In order to allow free exploration and inference of a range of policies we allow the participants to both freely select and switch between two stable environments. In doing this, we can investigate how these policies are used to test hypotheses about action-outcome contingencies under uncertainty.

Prediction error arises in action when the outcomes of actions are not the expected ones. Agents need to determine whether the appropriate response to prediction error is to 1) rely on it to meaningfully update one's model or 2) to ignore it as part of the inherent noise in the environment. This prediction error processing depends on our estimations of environmental uncertainty (the inverse of which is *precision*). Due to a combination of the sheer complexity of the environments we inhabit and our limited cognitive resources, there is always some irreducible uncertainty in the sensory signals, which must be estimated as a lower bound on expected prediction error. Similarly, here we are interested in how environmental variability of different kinds (which impacts action-outcome contingencies) affects decisions to select environments, rather than how uncertainty about the quality of the next environment impacts switching decisions. Recent work suggests that uncertainty may affect visual exploration using simulated eye movements based on an active inference implementation (Parr & Friston, 2017) and that active inference, as captured by recorded and simulated pupillometry, is affected by uncertainty (Vincent, Parr, Benrimoh, & Friston, 2019). We aim to test this kind of inferential process by looking at human movements as captured by a mouse in an online space.

In this experiment, there are two environments: *sand* and *water*. When the participant acts, they are displaced some distance from their expected location (based on how they moved their computer mouse); a distance we can measure using a behavioural proxy for *prediction error*. In both sand and water environments this distance is equal by design. However, in the water environment, the direction of this displacement (to the left or right of the mouse heading) is fixed by periodic *waves*. This means the environmental variability in the water environment is effectively half of that in the sand environment, but to benefit from this reduced variability, one must correctly model the waves. Thus, the choice between the two environments is a trade-off between an environment with a less complex, flatter model but higher irreducible uncertainty (sand), and a model with more hidden causes but the benefit of greater reducible uncertainty (water). This is analogous to a classic trade-off in Bayesian model selection between model complexity and model fit (Myung, 2000). Importantly, other than counterbalancing which half of the screen these environments were on, the statistical structure of these two options were stable both within and across trials. We can thus infer from behavioural preference for one of these environments how participants weight minimising model complexity against minimising reducible uncertainty for optimal task performance.

While many previous exploration tasks are set in a pragmatic reward context (Mehlhorn et al., 2015), since the focus of the current study is on epistemic action, we ask participants to focus on making a *judgement of agency* as an appropriate task. A judgement of agency is the conscious, reflective decision that one's actions were the causal source of the specified sensory input. This is related to, but empirically and conceptually distinct from the *sense of agency*, which is based on feeling in control of sensory consequences during the execution of the relevant action (Majchrowicz & Wierzchoń, 2018; Saito, Takahata, Murai, & Takahashi, 2015; Synofzik, Vosgerau, & Newen, 2008). Under predictive processing accounts, participants are thought to judge agency when there is a belief that goals can be reached from the agent's current state (Friston, Samothrakis, & Montague, 2012; Friston et al., 2013; Hohwy, 2015; Perrykkad, Lawson, Jamadar, & Hohwy, 2021). As such, the participant's judgment of agency is directly related to the fidelity of their model of action-outcome mappings in a particular environment. In scenarios where the uncertainty in the environment directly impacts the quality of this model, environment selection is crucial for forming this judgement.

Inferring policies around exploration may be particularly relevant for understanding behaviour characteristics in particular psychiatric conditions (Addicott, Pearson, Sweitzer, Barack, & Platt, 2017). Here, we focus on autism spectrum conditions (autism). Constant, Bervoets, Hens, and Cruys (2020) theorise that the behavioural repertoire characteristic of autism is best understood as the consequence of environment-building based on distinctly autistic ways of modelling statistical regularities in the environment. As a consequence of these differences in modelling the world, one would expect that participants with higher levels of autistic traits would have a stronger preference environments that can be captured by flatter models (Constant, Bervoets, et al., 2020; Perrykkad & Hohwy, 2020), as in our sand environment here.

Thus, our aims with this experiment are threefold. First, we are interested the use of environments with different kinds of uncertainty by participants in service of their epistemic goals. These include epistemic gain about the states about the world, about themselves and about the fidelity of their mappings of action policies and outcomes. In the environments offered in this experiment, the choice involves the trade-off between selection of model complexity and minimisation of reducible uncertainty over the tolerance of greater irreducible uncertainty but a simpler model. It also involves the use of environment switches in and of themselves as a policy for prediction error minimisation. The second aim is to understand how these policies and other policies around how we move through environments function to aid us in accurately mapping action-outcome contingencies, and thus infer and judge agency (cf. Perrykkad et al. (2021)). Out of interest in the relationship between elements of self-cognition, we also take a general measure of self-concept clarity. Finally, the third aim is to uncover how differences in environment selection and agency attribution differ along the autism spectrum by associating these behaviours with autism traits.

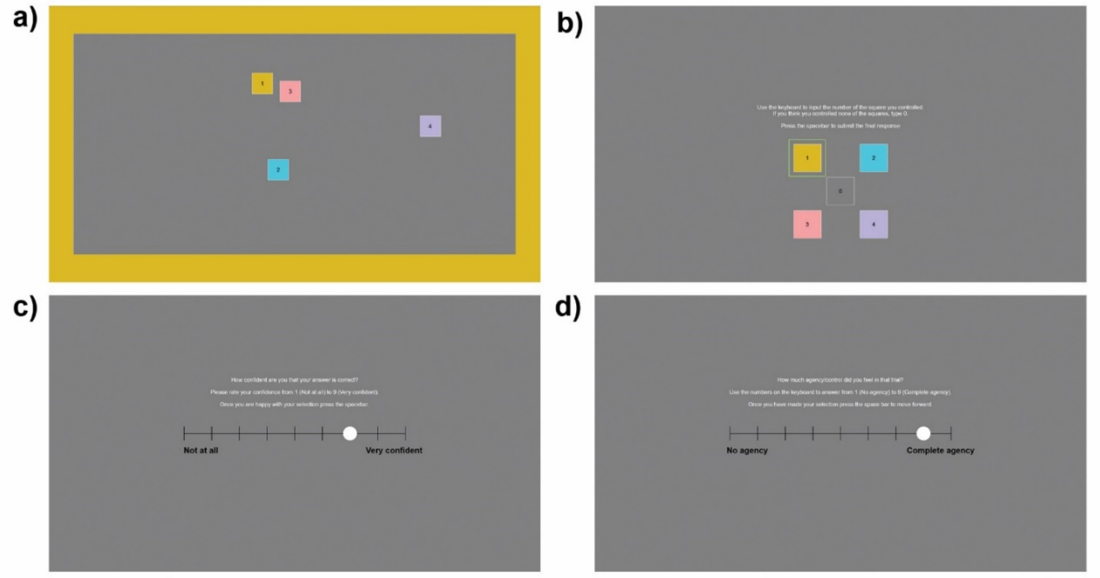
Methods

This study was approved by Monash University Human Research Ethics Committee (Project Number 26240) and was conducted in accordance with the relevant guidelines and regulations. All participants agreed to informed consent documents upon commencing the protocol.

Participants

A total of 229 participants were paid for completion of the study, 129 were recruited from Amazon Mechanical Turk using the Cloud Research platform (formerly TurkPrime (Litman, Robinson, & Abberbock, 2017)), and 100 were recruited from Prolific (http://www.prolific.co (Soler-Domínguez, de Juan, Contero, & Alcañiz, 2020)). The study had an overall completion rate of 67% (33% accepted but did not complete the posting). Data was collected in November, 2020. Participants were paid $4.50 USD (Amazon Mechanical Turk) or £4.10 GBP (Prolific) for completing the task, which took a median of 52 minutes to complete (including consent process and self-timed breaks, range: 26-146min total duration). Eligibility criteria included being fluent in English, aged 18-50, with no history of head injuries or neurological damage, normal or corrected-to-normal vision and no regular use of prescribed or unprescribed medication that may affect cognitive functioning.

Exclusion criteria were determined based on an initial sample of 21 participants recruited from Amazon Mechanical Turk and primarily focus on data quality due to unstable stimulus presentation online. Despite the usual timing and presentation accuracy of online tasks in PsychoJS (Anwyl-Irvine, Dalmaijer, Hodges, & Evershed, 2020; Bridges, Pitiot, MacAskill, & Peirce, 2020), we found that participants had large variability in actual presentation rates due to idiosyncrasies in computing set up and internet connection. To adapt to this reality, we created strict exclusion criteria for both trials and participants to ensure relatively stable presentation in the final dataset. Trials were defined as ‘bad’ if less than 50% of the frames were presented in the 15s, the frame rate during non-lagged periods was two or more times slower than expected, participants never pressed a square selection button, or participants moved the squares for less than 1% of the trial. In the final dataset, a total of 148 participants were excluded for one or more of the following reasons: incomplete dataset recorded due to technical difficulties (n=14), disqualification due to report of drug abuse in demographics survey (n=1), average accuracy less than 25% (chance = 20%, n = 53), more than 50% bad trials (n = 127), or if any one response occurred for 40 or more trials (n = 0). Of the final dataset, nine participants reported diagnosed mental conditions (ADHD: n = 2, Anxiety: n = 3, Depression: n = 2, PTSD: n = 1, Dyslexia: n = 1) but were not excluded from the study. 41 participants out of the final dataset of 84 were from Amazon Mechanical Turk, and 43 from Prolific. Participant demographic information for the final dataset is available in Table 1.



Use the keyboard to input the number of the square you controlled.
If you think you controlled none of the squares, type 0.

Press the spacebar to submit the final response

How confident are you that your answer is correct?

Please rate your confidence from 1 (Not at all) to 9 (Very confident).

Once you are happy with your selection press the spacebar.

How much agency/control did you feel in that trial?

Use the numbers on the keyboard to answer from 1 (No agency) to 9 (Complete agency).

Once you have made your selection press the space bar to move forward.

Figure 1 - Task Screenshots

Panel a) shows an example frame of the squares task, sand and water environments are hidden under the left and right halves of the central grey screen. In this example frame, square 1 is selected. Panel b) shows the judgement of agency response screen. Panel c) shows the confidence response screen and panel d) the sense of agency response screen.

Table 1 – General Demographic Information

Demographic	Category	N	% (n = 84)
Gender	Male	59	70.2%
	Female	24	28.6%
	Other	1	1.2%
Age	18-24	34	40.5%
	25-31	20	23.8%
	32-38	22	26.2%
	39-45	6	7.1%
	46-50	2	2.4%
Country of Residence	USA	41	48.8%
	Poland	16	19.0%
	United Kingdom	6	7.1%
	Portugal	6	7.1%
	Spain	3	3.6%
	Canada	2	2.4%
	France	2	2.4%
	Israel	2	2.4%
	Greece	1	1.2%
	Hungary	1	1.2%
	Czech Republic	1	1.2%
	Austria	1	1.2%
	Estonia	1	1.2%
	Netherlands	1	1.2%
First Language	English	53	63.1%
	Other – Fluent in English	31	36.9%
Highest Completed Education	Less than Highschool	2	2.4%
	Highschool or equivalent including Vocational Training	29	34.5%
	Bachelors, Honours, Associate, or Professional Degree	32	38.1%
	Masters or Doctorate	21	25.0%
Employment Status	Unemployed or Not Working	9	10.7%
	Student or Intern	23	27.4%
	Employed	52	61.9%

Procedure

Following the informed consent procedure, participants completed a general demographics survey, followed by the Autism-Spectrum Quotient (AQ) and Self Concept Clarity Scale (SCCS). Then they were forwarded to Pavlovia (http://pavlovia.org) to complete the agency task. Finally, they completed the Subthreshold Autism Trait Questionnaire (SATQ) and were compensated via a completion code or link. Participants were asked to complete the task using a chrome or firefox browser, using a laptop or desktop and with an external mouse (not laptop trackpad or touchscreen).

Autism-Spectrum Quotient (AQ)

The AQ is a 50-item questionnaire measuring autistic traits in the general population (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). The mean AQ score was 20 (SD:6.63, range: 8-37).

Subthreshold Autism Trait Questionnaire (SATQ)

The SATQ is a 24-item questionnaire which also measures subthreshold autism traits in the general population (Kanne, Wang, & Christ, 2012). The SATQ was designed to capture a broader range of autism symptoms as compared to the AQ. These symptoms include, "eye contact, being perceived as odd or strange, perception of facial expressions, being physically awkward, using gestures, and sharing enjoyment." Items on the SATQ are rated and scored on a 4-point likert scale indicating the extent to which the statement describes the participant on most days ("False, not true at all", "Slightly true, "Mainly True", "Very true"). The mean SATQ score was 26.98 (SD: 10.67, range: 5-51).

Self-Concept Clarity Scale (SCCS)

The SCCS is a 12-item questionnaire measuring structural properties of one's self-concept. Higher scores are related to increased clarity of self-concept, including temporal stability, certainty and perceived internal consistency of beliefs about oneself (Campbell et al., 1996). The mean SCCS score was 40.2 (SD: 11.02, range: 20-60).

Agency Task Design

This experiment was a variant of the *Squares Task* (Grainger, Williams, & Lind, 2014; Perrykkad et al., 2021; Russell & Hill, 2001; Williams & Happé, 2009), the most commonly used judgement of agency task with an autistic population (Perrykkad & Hohwy, 2020). Stimuli were presented online using PsychoJS (v2020.2)(Peirce et al., 2019). In this version of the task, there were four numbered squares on the screen in each trial. The squares were randomly coloured with perceived-luminance matched shades of blue, red, purple and yellow on a grey background. Participants pressed and held a number key to select a square, which coloured the border of the screen with the selected square's colour. If a square was selected, all the squares moved when the mouse was moved and all the squares stopped with a tiny amount of jitter (to ensure participants didn't think it had frozen) when the mouse stopped, so participants had to both select a square and move in order to accurately complete the task. The selected square moved at half the speed of the other squares. This square selection function allowed us to capture moment to moment hypothesis without using eye-tracking (cf. Perrykkad et al. (2021)), and as in foveal pursuit, the selected square would result in the most precise information, though more noisy information was available from other options. All squares moved faster than the actual mouse distance, to allow more screen wraps before the edge is hit, since repositioning the mouse using PsychoJS was not possible. To give participants more freedom of movement when this happens, they could also press ‘Q’ to reposition all squares to the left half of the screen or ‘R’ to reposition all to the right. All square positioning (initial and after these button presses) was random.

Square positions were only updated at 30hz, to limit computational performance variability across physical set ups. In what follows, we refer to a *frame* as one of these 30hz stimulus presentation data points. Regardless, refresh rate on monitors across participants was recorded as 60hz. Participants were given 15sec to identify the target square which they controlled. Distracter squares moved at a random angle offset from the vector of mouse movement, and this angle was also independently and randomly changed (and smoothly transitioned) five times in each trial. This means that each distracter square appeared to turn five times when the participant did not initiate a turn, breaking any illusion of control resulting from motor adaptation. Half of the trials were *no-control* trials in which all four squares were distracter squares. After the 15sec, all squares froze and were numbered, and prompted an unspeeded numerical response from participants indicating which square they controlled or ‘0’ if they thought they controlled none of the squares. Participants also responded on a 9-point likert scale to two additional questions asking for ratings of confidence (‘Not at all’, ‘Very confident’) and sense of agency (‘No agency’ to ‘Complete agency’). See Figure 1 for example screenshots of the main parts of the task. Participants completed a total of 48 non-practice trials in three blocks of eight agentive trials and eight no-control trials. Without breaks or computational delays, the judgement of agency task was expected to take 20-25min.

Within each trial, there were two hidden environments. On each trial, half the screen (left or right) was randomly assigned *water*, and the other half *sand*. In both environments on each frame, for each square, a sample was taken from a 95% confidence interval for ±70°. To save online processing costs, four random iterations of trial order and variability sampling for each frame across the whole experiment were pre-established, and randomly selected for each participant. For squares in the *sand* environment, the variability angle was added to the input mouse angle (and the offset angle if the square is a distracter) from the participant to generate the new square location for that frame (the distance moved depended on whether the square was currently selected). For squares located in the *water* environment, the sign of the sampled angle was forced to alternate between positive and negative values approximately every 500ms (depending on remaining presentation variability), creating regular *waves*, that pushed the participant consistently to the left or right of their heading. Since distribution and sampled magnitude of the variability was the same regardless of the environment, the displacement of the mouse on each frame across the two environments was equal. However, in water, there was a predictable structure to the variability that was not present in sand.

An interactive, self-timed block of instructions including six timed, full practice trials (three control, three no-control) explained the mechanics of the task and the presence of the beach environments at the beginning of the task. During the full practice trials, participants were given feedback about the accuracy of their agency judgements, however the main task had no feedback. With regard to the variability in general, participants were told “The square will not move as smoothly as you move the mouse, but it’s pretty close! This is the amount of turbulence you’ll experience in this experiment.” With respect to the environments, they were initially given screens filled with one environment at a time and differently coloured. On these screens, they were instructed, “In each trial, there will be two hidden environments on the screen. This screen is now filled with the *sand* environment. Try moving in the sand. As the sand shifts beneath you, your square will jitter randomly. The other environment is *water*. The screen is now filled with water. Try moving in the water. Notice how the waves in the water will push you from side to side." Then they were given a screen half coloured with the sand colour, and half with the water, "In this screen, like in the actual experiment, half the screen is *sand* and the other half *water.* You can quickly jump to one half of the screen by pressing 'Q' and 'R'. Try it!" It should be noted that this is the last screen on which either environment was coloured. The main task was presented with a grey background with the coloured border indicating the selected square.

We included additional measures to mitigate potential inattention in our online sample, including large warnings when participants failed to select squares or move the mouse during a trial, forced full screen at the end of every trial and instructional manipulation check questions during surveys (Oppenheimer, Meyvis, & Davidenko, 2009).

Analysis

Statistical Analysis

All statistical analyses were conducted using Jamovi version 1.1.9 and the GAMLj module (Gallucci, 2019; R Core Team, 2018; The Jamovi Project, 2019). As in Perrykkad et al. (2021) we used mixed models for the majority of our analyses. In addition to the fixed effects and covariates outlined in the model structures defined below, where a mixed model was used, by-participant random intercepts were included to address the non-independence of subject-level observations across trials and capture individual variability in task performance. Compared to traditional methods, this approach affords more sophisticated handling of missing and outlying data, thus improving the accuracy, precision, and generalisability of fixed effect estimates (Singmann & Kellen, 2020). Control covariates of number of frames and the standard deviation of wave duration in the water environment were also included as fixed effects (participant averages were used where participant-wise data was used) to account for variability in stimuli presentation quality.

In addition, the survey measures were included as continuous fixed effects along with their two-way interactions with other fixed effects in each mixed model below. Since including both SATQ and AQ measures of autism traits in the mixed models reported below would overlap greatly in variance accounted for, we decided to choose only the autism traits measure that was most orthogonal (least correlated) with our other survey measure (SCCS). For ease of interpretation, post-hoc tests for interactions with survey measures were simple effects contrasting participants mean scores to those above and below one standard deviation from the mean.

Across all statistical analyses, post-hocs are reported with a Bonferroni correction. For mixed models, the Satterthwaite method for estimating degrees of freedom is used.

Validation

We begin our statistical analyses by quantifying the average quality of stimulus presentation across participants devices and internet connections. We report the mean number of total frames in which stimuli was updated across each trial (for an intended 450 in 15s). We quantify the number of haemorrhaged lags which are periods of time between stimuli presentation frames longer than the expected duration of five frames (167ms). We also report the number of waves in an average trial in the water environment, their mode duration and their temporal variability given by the standard deviation of wave duration.

In order to validate the changes made due to practical considerations of putting the squares task online, we began with analyses which compare results from Perrykkad et al. (2021) with the experiment reported here. These comparisons primarily revolve around the use of button presses and reduced speed of selected object for moment-to-moment hypothesis selection rather than eye-position as in the previous version. To compare the studies directly, we compared how the percent of time participants spent with the correct square selected and the chosen square selected, split by correct and incorrect trials. The values for correct and chosen were equivalent for correct trials. In the current study, this percentage was of the total time spent with any square selected (using the button mechanic). In the previous study, these values are a percent of the total time in the trial, since a 'hypothesis' was always selected due to the eye-tracking methodology. To quantify any differences observed, a mixed model was used with the fixed factors of Study, Selected Square (chosen/correct) and Accuracy. We were primarily looking for interactions between study and the other factors as an indication of the magnitude of the impact of the change in mechanic to participant behaviour in completing the task. This mixed model did not control for stimulus presentation variability or include AQ as the other models did.

As in Perrykkad et al. (2021), a behavioural proxy for *prediction error* was calculated by taking the Euclidean distance between where the selected square would have ended up if it had followed the trajectory of the mouse input (with a constant speed multiplier as described above) and where the square actually went. This means that prediction error is largely under the control of the participants, and is influenced by three factors: 1) the speed of movement (determining the distance travelled, longer distances mean more prediction error) and the angular changes to the input determined by 2) the uncertainty in the environment and 3) the angular offset if the selected square is not the correct one.

Based on this prediction error measure, as part of the validation analyses, we computed a hypothesis switch centered ERPE for each participant using the same method as Perrykkad et al. (2021). To get roughly the same temporal epoch, we took 15 frames either side of the event (30hz presentation, ±500ms). This was uncorrected for temporal variability in stimulus presentation, so are more appropriately conceptualised as presented data points than time before and after the event of interest. In this version of the experiment, hypothesis switches require changing button presses, they are often surrounded by periods of inactivity. To account for this, in this experiment we separate all ERPEs into *lead-in* and *lead-out* epochs. In these epochs, if event is immediately preceded or followed by periods where nothing is selected, the ERPE epoch trigger is moved to the first (lead-out) or last (lead-in) time a hypothesis was validly selected, effectively removing periods of inactivity around the event of interest. Upon visual inspection of the hypothesis switch ERPE with the intended epoch, the lead-in was not at stable baseline levels at the start of a 15 frame epoch, so the lead-in epoch for this analysis was doubled (-30:15 frames around the event, centered at 0).

For statistical analysis, averages were taken in time bins of 5 frames during the epoch. For this analysis, there were nine time bins – six lead-in and three lead-out, with the time of the event occurring between or in time bins six and seven. The hypothesis switch ERPE was analysed using a mixed model with the fixed factor of time bin. The standard fixed effects of SCCS and autism traits; standard control fixed effects of number of frames and wave time variability; and standard random effect of participant were also included. All interactions between non-control fixed effects were included in this model. Main effects and interactions with time bin would indicate a pattern of prediction error around hypothesis switches, which is expected to peak at the time of the event, as in Perrykkad et al. (2021).

Prediction error slopes across various conditions could not be estimated as planned for most participants in this dataset due to the inconsistent nature of movement. Planned contrasts included slope by environment, and accuracy by agency as in Perrykkad et al. (2021). Slopes were successfully fitted to prediction error in both environments for only 21% of participants and across all four accuracy by agency conditions for only 5% of participants. Where they could be estimated at all, data on which the slopes were estimated was often very noisy. As such, no estimates of slope are reported.

Accuracy and Bias

Our primary manipulations in this study were the ground truth of agency or *trial type* (control vs no-control trials) and the environments included in each trial. Trials were classified according to the *dominant environment* based on the percent of time spent in each sand and water in each trial. So in a sand dominant trial, participants spent more than 50% of the trial in the sand half of the screen. To understand the influence of these features on accuracy, we performed a mixed model on trialwise data with the fixed factors of trial type and dominant environment. The standard fixed effects of SCCS and autism traits; standard control fixed effects of number of frames and wave time variability; and standard random effect of participant were also included. All interactions between non-control fixed effects were included in this model.

Since there were equal numbers of trials where the ground truth of agency was present or absent, we also computed signal detection theory measures for the agency signal for each participant. These included d’ and criterion. D’ is an unbiased measure of sensitivity to a signal and represents the separation between signal present and signal absent distributions in standard deviation units. Larger values indicate greater sensitivity to the presence of a signal and would indicate greater unbiased accuracy on the judgement of agency task. The criterion is a measure of the tendency for the participant to report present or absent in an ambiguous situation. A large positive criterion value implies that the participant requires strong evidence before reporting that they had agency in a trial. A smaller, negative criterion indicates that the participant is quite liberal with asserting agency, and when unsure, would be biased towards a positive response.

These two values were statistically tested against a value of zero, indicating no sensitivity (d') or no bias (criterion) as the null hypothesis.

Environmental Niche Selection

To quantify an overall preference for one environment over the other, we computed two measures in each trial – environment dominance (binary, described above) and percent of time spent in each environment (continuous). To test for overall environment preferences, one sample t-tests comparing these to 50% were used. Significant results would indicate a preference for one environment over another. To look at differences in accuracy or bias towards agency judgements depending on environmental preferences, the signal detection theory measures were recomputed for each participant, but split by trialwise environmental dominance. These were then compared using paired sample t-tests for each d' and criterion.

We then repeated two similar mixed models to the accuracy analyses, looking at factors within and following a trial that influenced the percent of time spent in each environment in each trial. The mixed model looking at how end of trial responses and trial type predict percent of time spent in each environment included fixed effects of sense of agency ratings, confidence, accuracy, judged agency and trial type. The standard fixed effects of SCCS and autism traits; standard control fixed effects of number of frames and wave time variability; and standard random effect of participant were also included. Only two-way interactions between autism traits, SCCS and other non-control fixed effects were included in this model.

The mixed model considering the association between in trial behaviours and percent of time spent in each environment included fixed effects of percent of frames spent moving, number of hypothesis switches, average prediction error, average speed, acceleration and jerk. The standard fixed effects of SCCS and autism traits; standard control fixed effects of number of frames and wave time variability; and standard random effect of participant were also included. Only two-way interactions between autism traits, SCCS and other non-control fixed effects were included in this model.

The final environment-based analysis looked at an ERPE using environment switches as the event. This analysis used the same method as the hypothesis switch erpe above, with the original matching 15 frame epoch for lead-in and lead-out periods. Time bins of 5 frames were used for statistical analysis – three of each lead-in and lead-out, with the event happening during or between time bins three and four. Since environment switches that occurred concurrently with hypothesis switches were included in the previous ERPE and could not be conceptually distinguished as a pure ‘environment switch’ policy rather than primarily a hypothesis switch policy, these switches were removed from this analysis. As such, environment switch events here were due either to traversing the center or outer edges of the screen or pressing ‘Q’ or ‘R’ to switch to the other side of the screen. A mixed model was used for analysis with the fixed factors of time bin and switch direction (to water or to sand). The standard fixed effects of SCCS and autism traits; standard control fixed effects of number of frames and wave time variability; and standard random effect of participant were also included. Only two-way interactions between autism traits, SCCS and other non-control fixed effects were included in this model. As with the hypothesis switch ERPE, main effects or interactions including time bin indicate a particular pattern of prediction error around the environment switch.

Judgement of Agency

To determine what features of a trial are associated with a judgement of agency for the participants, we performed two mixed models. The first looked at how other end of trial features were associated with the judgement of agency. Fixed effects in this model included the reported sense of agency, confidence, accuracy and trial type. The standard fixed effects of SCCS and autism traits; standard control fixed effects of number of frames and wave time variability; and standard random effect of participant were also included. Only two-way interactions between autism traits, SCCS and other non-control fixed effects were included in this model. This model was run on trialwise data.

The second mixed model instead focused on within trial behavioural features that could be associated with a judgement of agency. The fixed effects in this model included percent of frames spent moving, number of hypothesis switches, number of environment switches, percent of trial spent in water, average prediction error, average speed, acceleration and jerk. We also included the fixed effect of accuracy as a control, so that any results show a relationship with the judgement of agency regardless of the accuracy of that judgement. The standard fixed effects of SCCS and autism traits; standard control fixed effects of number of frames and wave time variability; and standard random effect of participant were also included. Only two-way interactions between autism traits, SCCS and other non-control fixed effects were included in this model. This model was run on trialwise data.

Results

To determine which autism trait measure to include in the models reported below, we performed Pearson's correlations between the survey measures. SATQ and AQ were significantly and strongly correlated ($r = 0.77$, $p = 1.9 \times 10^{-17}$) as expected. SATQ was significantly negatively correlated with SCCS ($r = -0.25$, $p = 0.02$) but AQ was not ($r = -0.12$, $p = 0.27$). As such, in order to maximise orthogonality of measures, AQ is used where autism traits are included in models below. Despite being included as continuous measures in the omnibus tests, for the below post-hoc contrasts the low autism traits group had AQ scores 13 or lower (n = 13), high autism traits group scored 28 or higher (n = 9), and the mean autism group scored between 14 and 27 (n = 62). The equivalent thresholds for SCCS scores put the low self-concept clarity group scoring 39 or lower on the SCCS (n = 37), the high group scoring 51 or greater (n = 15) and the mean group with scores between 40 and 50 (n = 32).

Validation

In terms of quality of data for our final dataset, on average, each participant had 12.58 trials of the total 48 removed for poor quality (std: 6.99, range: 0-24). In the remaining trials, an average of 317.20 frames (std: 64.97, range: 225-449) out of the programmed 450 were successfully presented per trial, putting actual stimulus presentation at an average of 21 hz. These trials had an average of 4.39 haemorrhaged lags (std:4.48, range:0-35) for an average lag duration of 2.98s (std: 2.93, range: 0-18.36). Since trials with less than half of frames presented were designated bad, when lags happen towards the end of the trial, after the majority of frames have been successfully presented, they may increase the overall time of the trial before the program recognises the trial time has elapsed on the frame following the lag. These variations to stimulus presentation due to variable computing and internet set ups mean that participants were presented with an average of 20.59 waves per trial (std: 4.28, range: 14-29) in the water environment with a mode wave time of 478.2ms per trial (std:112, range: 193-1432; compare to programmed 500ms) and a standard deviation of wave time per trial of 329.22 (std: 234.19, range: 7.16-2272).

The mixed model used to validate the hypothesis switch mechanic by relating accurate responses and dwell time spent on relevant squares showed significant main effects of Accuracy and Selected Square but no significant main effect of Study. Further, interactions between Accuracy and Selected Square, Accuracy and Study, Selected Square and Study and the three-way interaction were significant. See Figure 2. As the primary question of interest here concerned differences between studies, only the interactions with study were followed up with post-hoc analyses. Post-hoc analyses showed that despite the significant interaction, when Selected Square was held constant, there was no difference between the studies (Correct: t(481.2) = 2.447, p = 0.089; Chosen: t(481.2) = -0.70, p = 1.00). For the interaction between Accuracy and Study, for incorrect trials there is no difference between studies (t(481.6) = -1.66, p = 0.58) but for correct trials, participants in Perrykkad et al. (2021) looked at the square they chose on average 5.73% of the trial longer than in the current sample (t(481.2) = 3.41, p = 0.004). For the three way interaction, the only time the two studies were significantly different across the four combinations of accuracy and selected square was the percentage of time spent with the chosen square selected in incorrect trials, in which the current study showed an increase of 8.09% over

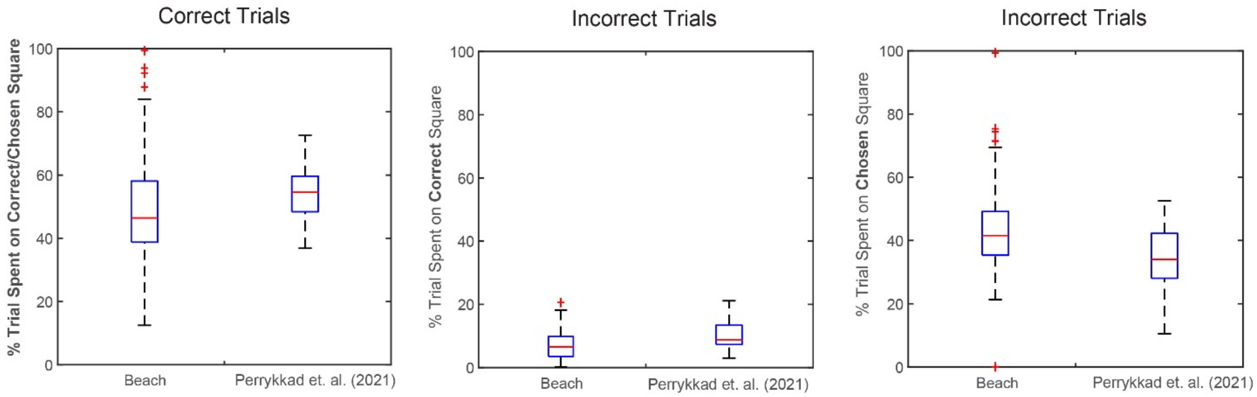


Figure 2 – Boxplot Comparison of Hypothesis Selection Mechanisms Current Study (Button Press) and Perrykkad et. al. (2021) (Eye-tracking). Y-axis indicates the percent of time spent on the relevant square. The left panel shows dwell time for correct trials, in which correct=chosen square, the right two panels are restricted to incorrect trials. The.middle panel represents correct square dwell time and right represents chosen square dwell time. Original finding from Perrykkad et al. (2021) is depicted in the right boxplot of each figure, and current study on the left for comparison.

the dataset from Perrykkad et al. (2021) (t(467.5) = -8.09, p = 0.006). Overall, even when there was a significant difference between the dwell times on relevant trials across these two studies it was quite small (<10%). We take this to indicate that the manual button press hypothesis selection method was comparable to the previous eye-tracking hypothesis selection method used in Perrykkad et al. (2021).

The next validation step was to look at the hypothesis switch ERPE, in which Perrykkad et al. (2021) showed a clear peak at the time of the event. In this analysis, time bins 1-6 relate to five frame windows preceding hypothesis switches, and time bins 7-9 represent the lead-out of the hypothesis switch. Time bins six and seven either include the hypothesis switch, or the hypothesis switch happens between the two (when movement is ceased while switching hypothesis). See Figure 3. Results of the mixed model analysis show only a main effect of Time Bin (F(8,620.10) = 4.60, p = 0.000018). Post-hoc comparisons showed the average prediction error in time bins three, four and five are significantly greater than time bin seven (three: t(620.1) = 3.24, p = 0.045; four: t(620.0) = 4.41, p = 0.00044; five: t(620.0) = 4.88, p = 0.000049) and

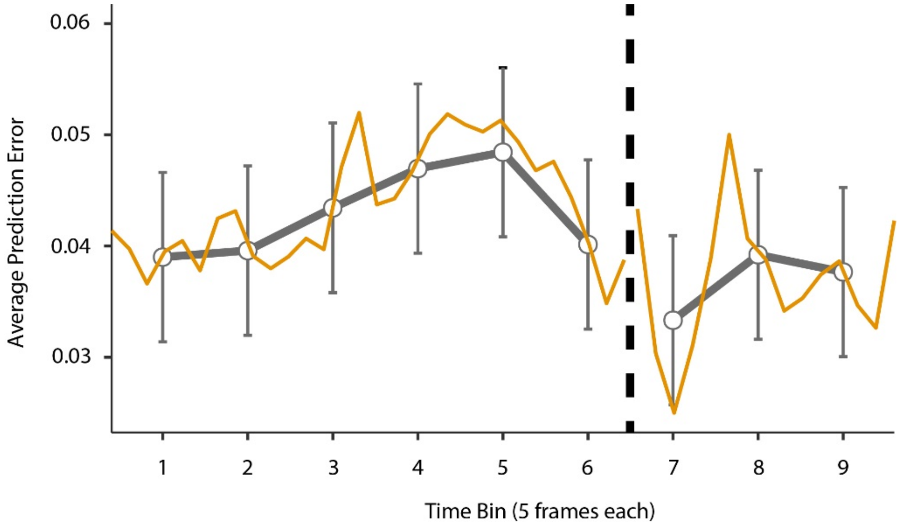


Figure 3 - Hypothesis Switch ERPE. Error bars are 95% CI. Y-axis shows average prediction error. The dotted line represents the time of a hypothesis switch from one square to another. The orange line depicts the grand average prediction error during the epoch, grey bars represent the average in each time bin which are used for statistical analysis.

that time bin five is also greater than time bin nine (t(620.0) = 3.48, p = 0.019). This suggests that while there is some increase before the time of a hypothesis switch (time bins 3-5), and potentially a decrease after (time bin seven), there is not such a clear cut peak around the hypothesis switch as in Perrykkad et al. (2021). This can likely be attributed to slowing down in anticipation of manually changing a button press (eg. time bin six), and a generally noisier signal due to many fewer hypothesis switches per trial (mean = 3.0, SD = 2.31; in contrast to mean = 42.2, SD = 13.48).

Accuracy and Bias

The average accuracy on the task was 54.07% (std: 15.7, range: 26.32-87.18), with chance responses falling at 20% (four squares or no-control). The mixed model for accuracy showed only a significant main effect of trial type (F(1,2886.3) = 8.42, p = 0.0038), such that no-control trials had poorer accuracy (by 5%) than trials in which the participant actually controlled one of the squares.

The signal detection theory analysis showed an average d’ of 0.55 (std: 0.89) and an average criterion of -0.46 (std: 0.41). See Figure 4. One sample t-tests showed that both d’ (t(83) = 5.68, p = 2.0x10^{-7}) and criterion (t(83) = -10.285, p = 1.8x10^{-16}) were significantly different

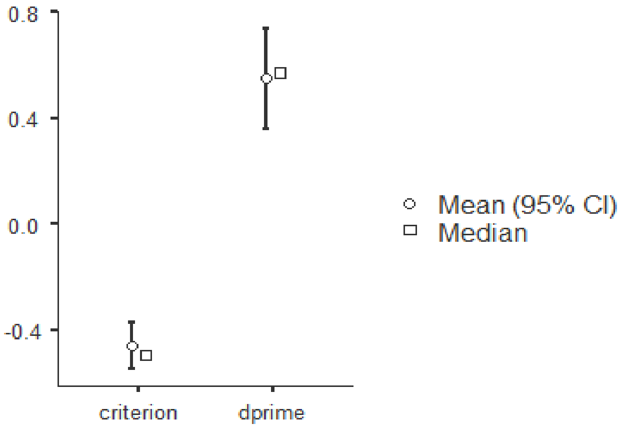


Figure 4 - Criterion and d' for Agency Signal. Error bars are 95% CI. Y-axis represents the value of signal detection measures in standard deviation units. For criterion (left), negative values reveal a liberal bias towards judging agency. Positive values of dprime indicate sensitivity to agency (unbiased measure of accuracy).

from zero. This shows that participants were sensitive to agency and had a liberal bias for attributing agency.

Environmental Niche Selection

Participants spent an average of 49.00% (std: 3.54) of each trial in the water environment, and a one sample t-test showed that though small, the bias towards the sand environment was significant (t(83) = -2.62, p = 0.010). Additionally, 47.50% of trials were classified as water dominant, and a one sample t-test showed that this too was significantly different from 50% (t(83) = -2.65, p = 0.010). Paired sample t-tests comparing d' and criterion for water dominant and sand dominant trials showed that environmental dominance did not affect either of these measures (d': t(83) = -1.55, p = 0.13; criterion: t(83) = 0.74, p = 0.46).

The mixed model used to investigate how time spent in each environment was predicted bv responses at the end of the trial and trial type revealed main effects of confidence (F(1,2955) = 7.58, p = 0.0059) and accuracy (F(1,2955) = 4.44, p = 0.035). In correct trials, participants spent 1.79% longer in the sand environment. Further, as confidence increase, time spent in sand increases (mean diff of 2.70% between highest and lowest confidence trials).

The mixed model used to investigate how time spent in each environment was predicted by other behaviours during the trial had no significant main effects or interactions.

On average, participants switched environments 7.38 times per trial (std:6.11, range: 0-67). Of these, an average of 87.88% (std: 18.18) were traverses of the center line or outer edges, 11.86% coincided with a hypothesis switch (std: 17.94) and 0.26% were performed using the 'Q' and 'R' key based mechanism (std: 2.95).

To investigate the pattern of prediction error around environment switches that were not due to hypothesis switches (which were captured as part of the analysis for Figure 3), we limited analysis here to only traverse and ‘Q/R’ based environment switches. The raw ERPE can be seen in Figure 5a. The planned mixed model showed a main effect of time bin (F(5,891.0) = 126.36, p = 3.83x10^{101}). Post-hoc analyses revealed that time bins three and four, at the time of the environment switch, were significantly greater in prediction error than all other time bins, and equal to each other (1-3: t(891.0) = -17.42, p = 1.53x10^{-57}; 2-3: t(891.0) = -14.25, p = 1.83 x10^{-40}; 5-3: t(891.0) = -14.69, p = 1.04 x10^{-42}; 6-3: t(891.0) = -16.30, p = 2.63 x10^{-51}; 1-4: t(891.0) = -16.54, p = 1.24 x10^{-52}; 2-4: t(891.0) = -13.37, p = 4.17 x10^{-36}; 5-4: t(891.0) = -13.81, p = 2.90 x10^{-38}; 6-4: t(891.0) = -15.42, p = 1.43 x10^{-46}; 3-4: t(891.0) = 0.88, p =1.0). Further, there was a significant increase from time bin one to two (t(891.0) = -3.17, p = 0.023). There was also a significant interaction between AQ and timebin (F(5,891.0) = 3.09, p = 0.0090). Post-hoc analyses showed that within each time bin there was no significant difference between the AQ

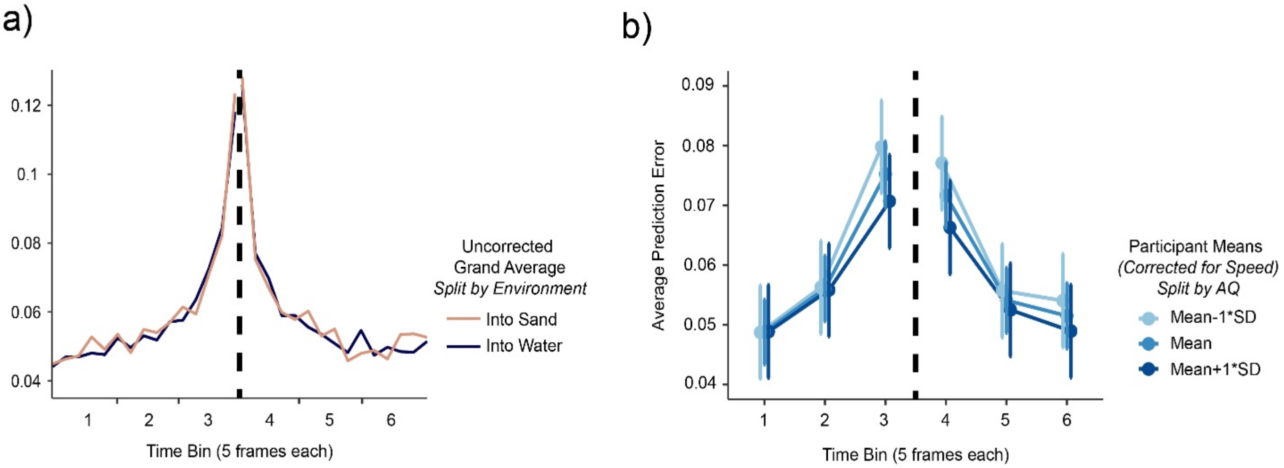


Figure 5 - Environment Switch ERPEs. Error bars in b) are 95% CI. Panel a) depicts the grand average prediction error (y-axis) across participants during the analysed epoch. The dotted line in both panels represents the time of the environment switch. The tan line is restricted to trials where the environment switch goes from water into sand, and the navy line the opposite direction. Panel b) shows averages in each time bin used for analysis (x-axis) as they are in the model corrected for speed (i.e. remaining error due to square offset or environmental variability), split by autism traits score.

groups, however, in the high AQ group time bin five is not significantly different to time bin one (t(891.0)=1.69, p = 0.091), whereas for both the mean group and the low group, time bin five is still elevated in prediction error compared to time bin one (mean AQ: t(891.0) = 2.74, p = 0.0064; low AQ: t(891.0) = 2.16, p = 0.031). There were no significant effects related to the direction of the switch.

Additional Analyses

We performed two additional analyses to determine the source of this peak in prediction error around the environment switches. As stated above, there are three sources of changes in prediction error: quality of hypothesis (which determines the presence and magnitude of an offset to the angle of movement), speed of mouse movement, and environmental variability. The previous analysis was restricted to environment switches that did not coincide with hypothesis switches, so while hypothesis switches can and do occur in the epoch surrounding the switch, this pattern is not merely a consequence of a hypothesis switch at the time of the event. As such, one uninteresting explanation of the pattern seen in Figure 5a is that the peak in prediction error occurs because participants decide to change environments, speed up to reach the environment boundary and then slow down to observe the stimuli move through the new environment. The following analyses sought to determine whether this wholly explains the pattern of prediction error around environment switches.

To begin, we performed the same mixed model as for the original environment ERPE, but replaced the dependent variable with mouse speed. As such, what follows is an analysis of event-related speed. The mixed model showed a main effect of time bin (F(5,891.0) = 24.38, p = 4.76 x10^{-23}) and an interaction between SCCS and time bin (F(5,891.0) = 2.41, p = 0.035). Post hoc analyses showed that our suspicions were somewhat verified – as in the analysis above, time bins three and four, at the time of the environment switch, were significantly greater in speed than all other time bins, and equal to each other (1-3: t(891.0) = -4.94, p = 1.40x10^{-5}; 2-3: t(891.0) = -5.93, p = 6.55x10^{-8}; 5-3: t(891.0) = -4.69, p = 4.67x10^{-5}; 6-3: t(891.0) = -5.73, p = 2.03x10^{-7}; 1-4: t(891.0) = -7.36, p = 6.21x10^{-12}; 2-4: t(891.0) = -8.35, p = 3.89x10^{-15}; 5-4: t(891.0) = -7.11, p = 3.46x10^{-11}; 6-4: t(891.0) = -8.15, p = 1.78x10^{-14}; 3-4: t(891.0) = -2.42, p =0.24). Post hoc analyses of the interaction showed a greater difference between the peak time bins and the first time bin with lower self-concept clarity scores (larger peak amplitude with poorer self-concept driven by nonsignificant differences in earlier time bins).

To determine the extent of the influence of this similar pattern of speed on the original environment switch ERPE, we performed a mixed model with the same structure as the original ERPE model, with the additional fixed covariates of mouse speed and the interaction between mouse speed and time bin. Even when removing variance associated with changes in speed across the time bins, this mixed model showed a significant main effect of time bin and a significant interaction between time bin and AQ. Again, time bins three and four showed significantly greater prediction error than any of the other time bins and were not significantly different from one another (1-3: t(891.0) = -16.74, p = 9.53x10^{-54}; 2-3: t(891.0) = -12.05, p = 6.23x10^{-30}; 5-3: t(891.0) = -13.37, p = 3.88x10^{-36}; 6-3: t(891.0) = -14.90, p = 7.28x10^{-44}; 1-4: t(891.0) = -14.27, p = 1.19x10^{-40}; 2-4: t(891.0) = -9.66, p = 6.85x10^{-20}; 5-4: t(891.0) = -10.95, p = 3.95x10^{-25}; 6-4: t(891.0) = -12.46, p = 7.92x10^{-32}; 3-4: t(891.0) = 2.25, p =0.37). This shows that the prediction error related to selected square offset angle and fluctuations in environmental variance also peak around the time of an environment switch, independently of the speed at which the participant is moving. While our initial suspicions about this effect being driven partly by changes in speed was confirmed, this shows that a significant portion of the original effect is also driven by the combination of environmental variability and hypothesis quality.

The interaction with AQ is slightly different in this speed-corrected model (see Figure 5b). While it is still true that for the high AQ group time bin five is not significantly different to time bin one (t(874.7)=1.64, p = 0.10), whereas for both the mean group and the low group, time bin five is still elevated in prediction error compared to time bin one (mean AQ: t(874.5) = 3.40, p = 7.0x10^{-4}; low AQ: t(874.9) = 3.11, p = 0.0019). For only the low AQ group, prediction error is still elevated in time bin six compared to time bin one (t(875.4) = 2.37, p = 0.018). The amplitude of the peak is greater for participants with low AQ (time bin 1-3: t(883.1) = 13.88, p = 9.25x10^{-40}) compared to high AQ (time bin 1-3: t(877.7) = 9.79, p = 1.55x10^{-21}) and mean AQ falling between. This difference in amplitude appears to be due primarily to differences in the peak itself, which is trending in the fourth time bin (t(104.9) = -1.98, p = 0.051). This linear amplitude difference across AQ was not present in the original model, which was uncorrected for speed contributions to prediction error.

Judgement of Agency

The mixed model used to investigate how end of trial responses and trial type influence the participants' final judgement of agency showed main effects of sense of agency rating (F(1,2726.9) = 731.61, p = 6.11x10^{-143}), confidence (F(1,1860.6) =46.84, p = 1.04x10^{-11}), accuracy (F(1,2957.9) =381.69, p = 4.81x10^{-80}) and trial type (F(1,2918.1) =284.96, p = 4.50x10^{-61}). Greater sense of agency was associated with greater judgement of agency; 90.6% of trials with a sense of agency more than one standard deviation above the mean were judged as agentive, compared to 40.9% of trials with an equivalently low sense of agency. Confidence was inversely related to judgment of agency, such that lower confidence trials were more likely to be judged as agentive than higher confidence trials (reflecting the negative criterion value). Further, incorrect trials were more likely to be judged as agentive (also reflecting the negative criterion). Finally, no-control trials were less likely to be judged as agentive than trials where the participant actually had control over a square (reflecting the positive d'). All three of these main effects are consistent with the signal detection analysis.

In this model, there were also significant interactions between AQ and trial type (F(1,2922.0) =5.06, p = 0.024) and AQ and sense of agency (F(1,2746.7) = 16.48, p = 5.05x10^{-5}). The difference in judgement of agency between control and no-control trials (no-control trials being less likely to be judged as agentive) increased with AQ (but was highly significant at all levels of AQ: p = 4.66x10^{-24}-4.50x10^{-61}). This suggests that autistic traits may be related to a weaker agentive bias, though AQ score did not significantly interact with trial type in the accuracy mixed model, so the effect is likely not very strong. The sensitivity of judgement of agency to sense of agency also increased with AQ (though was also highly significant at all levels of AQ: p = 1.54x10^{-52}-1.15x10^{-142}), suggesting a tighter relationship between reported sense and judgement of agency with more autism traits.

There were also significant interactions between SCCS scores and accuracy (F(1,2958.0) =5.43, p = 0.020), trial type (F(1,2898.2) = 11.66, p = 6.49x10^{-4}), sense of agency rating (F(1,2829.67) = 39.13, p = 4.57x10^{-10}) and confidence (F(1,2478.6) = 18.01, p = 2.27x10^{-5}). Higher self-concept clarity scores were associated with smaller differences between accuracy and judgment of agency and between trial type and judgement of agency (though was also highly significant at all levels of SCCS: p = 3.07x10^{-21}-4.81x10^{-80}). This suggests that a higher quality overall self-concept may temper agentive biases. Only when sense of agency is particularly high was SCCS positively associated with judgement of agency (sense of agency > mean + 1SD:

t(150.6) = 4.61, p = 8.34x10^{-6}), and when it was particularly low SCCS was negatively associated with judgment of agency (sense of agency > mean - 1SD: t(151.7) = -2.12, p = 0.036). When sense of agency scores fell around the mean, SCCS score did not significantly relate to judgement of agency scores (sense of agency within 1SD of mean: t(80.3) = 1.47, p = 0.14). In other words, when participants did not have a strong feeling of agency, people with a high quality self-concept were more likely to judge that they were not the agent, and when they had a high sense of agency, these participants were more likely to attribute agency than those with a low quality self-concept. Lastly, only when confidence was less than one standard deviation below the mean was greater SCCS associated with a positive judgement of agency (t(156) = 3.55, p = 5.13x10^{-4}). When confidence was higher, there was no significant relationship between SCCS and judgement of agency. This suggests that it is only when not confident that the quality of one's self-concept drives one's tendency to judge that one has agency.

The next mixed model was used to investigate how judgement of agency was predicted by behaviours during the trial. This analysis showed a main effect of percent of frames spent moving (F(1,1356.2) = 23.05, p = 1.75x10^{-6}) such that the more of a trial that is spent moving, the more often the trial is judged as agentive. There was also a main effect of number of hypothesis switches (F(1,990.2) =31.62, p = 2.44x10^{-8}) such the fewer hypothesis switches in a trial, the more likely the participant will judge that they had agency. A main effect of average prediction error (F(1,2234.7) = 35.20, p = 3.44x10^{-9}) revealed that lower prediction error is associated with judgements of agency. Finally, the model showed a significant main effect of average speed (F(1,2896.8) = 5.25, p = 0.022) such that greater average speed is associated with more frequent judgements of agency. Since accuracy was controlled for in this model, all these results are independent of the accuracy of the agency judgement.

In this model, there were also significant interactions between AQ and number of environment switches (F(1,2913.6) = 6.46, p = 0.011) and average speed (F(1,2898.3) = 3.85, p = 0.050). Post hoc analyses revealed that only in high AQ was the number of environment switches positively associated with a judgment of agency (t(2849.3) = 7.55, p = 0.0060), at mean and low AQ scores there was no significant relationship. Only in the low and mean AQ groups was there a significant positive relationship between speed and judgement of agency (low: t(2923.3) = 7.57, p = 0.0060, mean: t(2911.8) = 5.24, p = 0.022), which was not significant in participants with high AQ scores.

Discussion

One of the greatest uses of our agency is to place ourselves in environments that are best suited to our epistemic needs. In the experiment reported here, participants were able to freely move between two environments as they tried to determine which of four squares on screen they controlled, if any. Participants knew that half the screen would be filled with a *sand* environment, where their movements would be impacted by random variability, and a *water* environment, in which the same magnitude of variability was periodically limited to pushing the square to the left or right of its heading (*waves*). In this way, we could look at how participants used selecting and switching between these environments as part of their repertoire of policies to complete the agency task. We also measured self-concept clarity using the SCCS and autism traits, using AQ and SATQ (though results focus only on AQ). After every trial, participants selected the square they thought they controlled (or indicated no-control) and rated their sense of agency and confidence. The validation analysis suggested that this online version of the squares task can be used to measure similar constructs to lab versions of the task (Perrykkad et al., 2021), despite some broad differences in design.

Environmental Niche Selection

Of the two environments offered to participants in this experiment, there are substantial a priori benefits of each. They were matched for prediction error at the most basic level – that is, the Euclidean distance between where the objects would have ended up had they followed the mouse and where they did end up in each environment was the same. This is the same basic measure of prediction error reported as a behavioural proxy in our results. However, in the sand environment, there was no further structure underlying the variability in the environment. Successfully modelling the variability at this most basic level is the most one can reduce their uncertainty in this environment. However, in the water environment, given a good model of the environment one could predict which direction the variability would be confined to, essentially halving the range of expected locations. In this sense, the two environments pitted model complexity (modelling both the waves and the variability in the water environment) against irreducible uncertainty (the full range of variability experienced in the sand environment).

Our results suggest that while very slight, participants did have a significant preference for the sand environment, preferring reduced model complexity to the increased reducible uncertainty of the water environment. Again, this was very small, at a mean difference of only one percent across trials, but it was a reliable bias across different ways of measuring preference, with an average of 2.5% more trials being classified on the whole as sand dominant. This also did seem to garner a quantifiable advantage, with participants spending longer in the sand environment when they were correct, and participants reporting greater confidence with more time spent in the sand environment. Future work could titrate the variability to push participants towards greater model complexity or greater irreducible uncertainty over multiple trials to establish individual participant thresholds that may be associated with participant qualities such as autism traits or self-concept clarity. While we expect that higher autism traits would predict stronger preference for environments that require lower model complexity, the preference for sand was not dependant on AQ score. Both environments were very simple, and future experiments should consider naturalistic environments with more hidden causes to uncover possible effects of AQ on environment preference.

Further, the pattern of both prediction error and speed around environment switches suggests that participants were using a change in environment (in either direction) as an intentional policy. In line with our initial deflationary account of the first ERPE analysis, participants did speed up immediately before crossing the environmental boundaries and slowed down following. This behaviour suggests the action was somewhat intentional and that feedback from the new environment was worth paying attention to. The second ERPE mixed model showed that even when controlling for speed, the other factors contributing to prediction error magnitude also increased before participants switched environments, and decreased afterward – namely, environmental variability and square offset (when the selected square was not the correct one). This is compelling evidence that participants are using the act of switching environments as a policy in response to increasing prediction error.

In previous work, we have shown that this pattern of prediction error increase around an enacted policy does not occur around all events that might be of interest (eg. see the volatility ERPE analysis in the Supplementary Materials for Perrykkad et al. (2021)) but does occur around other participant-initiated policies that may reduce uncertainty (as for hypothesis switches in Perrykkad et al. (2021)). It is important to highlight, however, that the function of the policy in the reduction of prediction error in these two policies is different. In the hypothesis switch ERPE in Perrykkad et al. (2021), reductions following the switch can be reliably attributed to the act of switching hypotheses itself – there is usually a sizeable difference in the prediction error for different squares due to differences in offset. In our environment switch analysis, we remove environment switches that coincide with hypothesis switches (ie. hypothesis switches to a square in the other environment), and so offset changes occurring as a result of the environment switch policy itself are eliminated as a possible explanation for reduction in prediction error following the policy. As a reminder, prediction error magnitude, by this behavioural proxy measure, is identical in the two environments by design, so the reduction in prediction error between bins four and five in the lead-out ERPE cannot be attributed directly to the environment switch itself. It is of course important to keep in mind that the behavioural proxy for prediction error used here does not account for expectations about waves, which would in fact alter the prior and reduce neural/cognitive prediction error. The most reasonable explanation of the decrease in the lead-out ERPE is the presence of hypothesis switches in the period following the environment switch, which were not eliminated from the studied epochs. In this way, participants seem to be first changing environments, and then when their actions are still not self-evidencing in the new environment, they quickly change their hypothesis. These subsequent hypothesis switches in combination with random fluctuations in offset and variability are likely what actually cause the successful reduction in prediction error following environment switches. It could also represent a return to baseline levels of prediction error with a variable duration of the peak around the environment switch appearing as a steady decrease in the averaged data.

The way the environment switching policy is used in response to prediction error was modulated by participants' autism traits. Participants with low autism traits have a greater difference between the height of their peak at the time of switching environments and the baseline established in the first time bin than those with high autism traits. This difference is only present when the model is corrected for the influence of speed, suggesting that it is not a result of differences in motor execution, but rather a response to the increasing prediction error driven by environmental (and object based) factors. Participants with high AQ appear to tolerate less prediction error before enacting a policy in response to it, a finding that is consistent with AQ modulation of the hypothesis switch ERPE reported in Perrykkad et al. (2021).

This study is of course limited in conclusions about the relationship between environment selection and autism in its use of autism traits as opposed to a clinically diagnosed population. While overall the sample size in post-hoc analyses used to compare interactions with autism traits is low, the omnibus interactions were based on modelled trends in the full dataset of continuous AQ scores. Nevertheless, environmental uncertainty might be particularly relevant to action selection for different levels of autistic traits and we do show interactions between policies around environment selection and AQ. These are worth following up in future studies in diagnosed populations.

The experiment presented here is a first step towards testing notions of *niche construction* in human behaviour as a corollary of active inference (Bruineberg, Rietveld, Parr, van Maanen, & Friston, 2018; Constant, Bervoets, et al., 2020; Constant, Clark, Kirchhoff, & Friston, 2020; Constant, Ramstead, Veissière, Campbell, & Friston, 2018; Veissière, Constant, Ramstead, Friston, & Kirmayer, 2020). Cognitive niche construction, or organism-niche coordination dynamics, is the process by which individual organisms reciprocally shape and are shaped by the environments they inhabit (Fabry, 2021). Niche construction of this kind at the individual level often acts as a kind of informational epistemic engineering, where humans physically transform their environment in ways that make important features more salient (Sterelny, 2010), for example, to reduce the descriptive complexity of the environment (Clark, 2008). While mere environment *selection*, as we explored in this experiment, is frequently included as a part of formal definitions of niche construction (Laland, Odling-Smee, & Feldman, 2000), the obvious strength of this concept comes with cases where the individual is given the power to iteratively and physically (semi-permanently) alter the statistics of the environment in aid of prediction error minimisation and self-evidencing. As such, future experiments should aim to incorporate these more substantial features of organism-niche coordination dynamics into the experimental design – not only providing environment *selection* as an available uncertainty reduction policy for participants, but also environment *modification* or *construction*.

Judgement of Agency

Perrykkad et al. (2021) showed that a steeper reduction in prediction error over a trial in a similar task was associated with a judgement of agency, regardless of the accuracy of that judgement. While we could not reliably estimate slopes of prediction error in this study, we do show that lower average prediction error over a trial is associated with a judgment of agency, also regardless of the accuracy of that judgement. This shows that participants were using a quantity associated with the behavioural prediction error proxy to complete the task.

We have discussed above that a stronger preference for the sand environment in a trial was associated with higher accuracy and confidence on the judgement of agency task. Given that participants appeared to successfully use switching between environments as a policy to aid prediction error minimisation, it is conceivable that there would be a stronger relationship between number of environment switches and overall judgement of agency. Our findings show that only in participants with a particularly high AQ did more environment switches predict a judgement of agency (regardless of its accuracy). In combination with a smaller peak amplitude in the environment switch ERPE, this may suggest that participants with higher autism traits are using changes in environment as more of a self-evidencing policy than other participants. Other participants may be using the policy more frequently as a pragmatic action to optimise use of the the environment ('I suspect the noise is in the environment, switching will help') than an epistemic action used to confirm expectations of themselves ('I suspect the noise might be my square, let me test it for consistency in the other environment')(Friston et al., 2015). When the square continues to meet expectations in multiple switches, this acts in a self-evidencing way to reinforce the judgement of agency, thus leading to the association between number of environment switches and judgment of agency for individuals with higher AQ.

Participants overall showed both sensitivity in detecting the underlying agency signal and a bias towards judging agency when there was ambiguous evidence. These findings from d' and criterion were supported by other results. Low confidence trials and incorrect trials were both associated with a judgement of agency, which supports the bias towards agency. Supporting the sensitivity finding, no-control trials were more likely to be correctly judged as non-agentive than trials where the participant did actually have control. This difference in judgement of agency between control and no-control trials increased with AQ, even though there was no main effect of AQ on accuracy. Participants with high AQ had a stronger relationship between judgement of agency and sense of agency. While the final judgement of agency and the feeling of agency are usually thought to go hand in hand, there is a conceptual distinction and increasing empirical evidence that this is not always the case (Saito et al., 2015)(though much empirical literature also conflates explicit and implicit measures of agency, where in the current study both measures are explicit). For example, you may get a feeling of agency when you flick a light switch and immediately see the light go on, but judge that you do not have agency when you realise that the light switch was broken all along. Our results suggest that these two aspects of agency may be more tightly tied in people with higher autism traits, but the causal relationship is unclear.

Our results also showed interactions between self-concept clarity scores and performance on the judgement of agency task. Participants with a higher quality self-concept had smaller differences between accurate and inaccurate trials on their judgment of agency, suggesting a weaker bias towards agency. However, they also showed a weaker relationship between trial type and judgement of agency, suggesting also weaker sensitivity to agency. These findings may be illuminated by our results showing that the quality of self-concept also affected the relationship between sense and judgement of agency. When participants did not have a strong feeling of agency, people with a high quality self-concept were more likely to judge that they were not the agent than that they were. When they had a high sense of agency, these participants were more likely to attribute agency than those with a low quality self-concept. This suggests that people with a high quality self-concept trusted their sense of agency more in making judgements of agency. Self-concept clarity may temper general biases in favour of agency in favour of trust in their feeling of agency, whether positive or negative. These findings also suggest that there is a relationship between action-oriented attributions to the self as in the judgement of agency and the overall quality of ones' explicit self-concept across broader domains.

Conclusion

This experiment uncovered the use of environment-oriented policies by participants in the context of structured and unstructured variability. The two environments in each trial differed in their underlying complexity and irreducible uncertainty, but at the most basic level, afforded the same amount of uncertainty in action-outcome mapping. Results show that participants effectively employ both environment selection and switching policies in service of prediction error minimisation. There was a slight but significant preference for reduced model complexity over reduced irreducible uncertainty. We also show that prediction error minimisation informs participants' judgments of agency, regardless of the accuracy of their judgment. Finally, participants with more autism traits appear to tolerate less prediction error before enacting a policy in response to it and use environment switching to inform their judgement of agency more so than those with fewer autism traits.

References

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., & Platt, M. L. (2017). A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology, 42*(10), 1931-1939.

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2020). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior research methods*. doi:10.3758/s13428-020-01501-5

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5-17. doi:10.1023/A:1005653411471

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PeerJ, 8*, e9414.

Bruineberg, J., Rietveld, E., Parr, T., van Maanen, L., & Friston, K. J. (2018). Free-energy minimization in joint agent-environment systems: A niche construction perspective. *Journal of Theoretical Biology, 455*, 161-178.

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of personality and social psychology, 70*(1), 141.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*: OUP USA.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*: Oxford University Press.

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 933-942.

Constant, A., Bervoets, J., Hens, K., & Cruys, S. V. d. (2020). Precise Worlds for Certain Minds: An ecological perspective on the relational self in autism. *Topoi, 39*(3), 611-622. doi:10.1007/s11245-018-9546-4

Constant, A., Clark, A., Kirchhoff, M., & Friston, K. J. (2020). Extended active inference: Constructing predictive cognition beyond skulls. *Mind & Language, n/a*(n/a). doi:https://doi.org/10.1111/mila.12330

Constant, A., Ramstead, M. J. D., Veissière, S. P. L., Campbell, J. O., & Friston, K. J. (2018). A variational approach to niche construction. *Journal of The Royal Society Interface, 15*(141), 20170685. doi:10.1098/rsif.2017.0685

Fabry, R. E. (2021). Limiting the explanatory scope of extended active inference: the implications of a causal pattern analysis of selective niche construction, developmental niche construction, and organism-niche coordination dynamics. *Biology & Philosophy, 36*(1), 6. doi:10.1007/s10539-021-09782-6

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., J, O. D., & Pezzulo, G. (2016). Active inference and learning. *Neurosci Biobehav Rev, 68*, 862-879. doi:10.1016/j.neubiorev.2016.06.022

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive neuroscience, 6*(4), 187-214.

Friston, K., Samothrakis, S., & Montague, R. (2012). Active inference and agency: optimal control without cost functions. *Biological Cybernetics, 106*(8), 523-541. doi:10.1007/s00422-012-0512-8

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. (2013). The anatomy of choice: active inference and agency. *Frontiers in human neuroscience, 7*(598). doi:10.3389/fnhum.2013.00598

Grainger, C., Williams, D., & Lind, S. E. (2014). Online Action Monitoring and Memory for Self-Performed Actions in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 44*, 1193-1206.

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., & Cognitive Search Research, G. (2015). Exploration versus exploitation in space, mind, and society. *Trends Cogn Sci, 19*(1), 46-54. doi:10.1016/j.tics.2014.10.004

Hohwy, J. (2013). *The predictive mind*: Oxford University Press.

Hohwy, J. (2015). Prediction, agency, and body ownership. In *The Pragmatic Turn:: Toward Action-Oriented Views in Cognitive Science* (pp. 109-120): The MIT Press.

Hohwy, J. (2016). The Self-Evidencing Brain. *Noûs, 50*(2), 259-285.

Hutchinson, J. M., Wilke, A., & Todd, P. M. (2008). Patch leaving in humans: can a generalist adapt its rules to dispersal of items across patches? *Animal Behaviour, 75*(4), 1331-1349.

Kanne, S. M., Wang, J., & Christ, S. E. (2012). The Subthreshold Autism Trait Questionnaire (SATQ): development of a brief self-report measure of subthreshold autism traits. *J Autism Dev Disord, 42*(5), 769-780. doi:10.1007/s10803-011-1308-8

Laland, K. N., Odling-Smee, J., & Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *Behav Brain Sci, 23*(1), 131-146; discussion 146-175. doi:10.1017/s0140525x00002417

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior research methods, 49*(2), 433-442.

Majchrowicz, B., & Wierzchoń, M. (2018). Unexpected action outcomes produce enhanced temporal binding but diminished judgement of agency. *Consciousness and Cognition, 65*, 310-324. doi:https://doi.org/10.1016/j.concog.2018.09.007

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., . . . Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision, 2*(3), 191-215. doi:10.1037/dec0000033

Myung, I. J. (2000). The Importance of Complexity in Model Selection. *Journal of Mathematical Psychology, 44*(1), 190-204. doi:https://doi.org/10.1006/jmps.1999.1283

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of experimental social psychology, 45*(4), 867-872.

Parr, T., & Friston, K. J. (2017). Uncertainty, epistemics and active inference. *Journal of The Royal Society Interface, 14*(136), 20170376.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., . . . Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior research methods, 51*(1), 195-203.

Perrykkad, K., & Hohwy, J. (2020). Modelling Me, Modelling You: the Autistic Self. *Review Journal of Autism and Developmental Disorders, 7*, 1-31. doi:10.1007/s40489-019-00173-y

Perrykkad, K., Lawson, R. P., Jamadar, S., & Hohwy, J. (2021). The effect of uncertainty on prediction error in the action perception loop. *Cognition, 210*, 104598. doi:https://doi.org/10.1016/j.cognition.2021.104598

Russell, J., & Hill, E. L. (2001). Action-monitoring and Intention Reporting in Children with Autism. *Journal of Child Psychology and Psychiatry, 42*(3), 317-328. doi:10.1111/1469-7610.00725

Saito, N., Takahata, K., Murai, T., & Takahashi, H. (2015). Discrepancy between explicit judgement of agency and implicit feeling of agency: Implications for sense of agency and its disorders. *Consciousness and Cognition, 37*, 1-7. doi:http://doi.org/10.1016/j.concog.2015.07.011

Schwartz, S. J., Klimstra, T. A., Luyckx, K., Hale III, W. W., Frijns, T., Oosterwegel, A., . . . Meeus, W. H. (2011). Daily dynamics of personal identity and self-concept clarity. *European Journal of Personality, 25*(5), 373-385.

Soler-Domínguez, J. L., de Juan, C., Contero, M., & Alcañiz, M. (2020). I walk, therefore I am: a multidimensional study on the influence of the locomotion method upon presence in virtual reality. *Journal of Computational Design and Engineering*.

Sterelny, K. (2010). Minds: extended or scaffolded? *Phenomenology and the Cognitive Sciences, 9*(4), 465-481. doi:10.1007/s11097-010-9174-y

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition, 17*(1), 219-239. doi:http://doi.org/10.1016/j.concog.2007.03.010

Veissière, S. P. L., Constant, A., Ramstead, M. J. D., Friston, K. J., & Kirmayer, L. J. (2020). Thinking through other minds: A variational approach to cognition and culture. *Behavioral and Brain Sciences, 43*, e90. doi:10.1017/S0140525X19001213

Vincent, P., Parr, T., Benrimoh, D., & Friston, K. J. (2019). With an eye on uncertainty: Modelling pupillary responses to environmental volatility. *PLoS Comput Biol, 15*(7), e1007126. doi:10.1371/journal.pcbi.1007126

Williams, D., & Happé, F. (2009). Pre-conceptual aspects of self-awareness in autism spectrum disorder: The case of action-monitoring. *Journal of Autism and Developmental Disorders, 39*(2), 251-259.

Acknowledgements

The authors appreciate the time and effort of their participants without whom this research would not be possible. Additional thanks to eventual co-authors Jonathan Robinson and Jakob Hohwy. Jonathan Robinson translated task software developed by Kelsey Perrykkad for Perrykkad et al. (2021) to the online platform and provided useful discussion at all stages of this project. This work is supported by the Australian Research Council (DP190101805) to Jakob Hohwy.

Just as I argued in Chapter 4 that fidgeting, or autistic stimming, is best understood as performing actions to induce highly precise and reliable sensory feedback across contexts, Constant et al. (2020) argue that autistic individuals construct environments in order to create simplified predictable niches. This sets up the environment to provide reliable feedback in just the same way as fidgeting is a ‘portable’ strategy for prediction error minimisation.

Results from this experiment do suggest differences in environmental policy use for groups with different levels of autism traits. Findings from this chapter show a distinct relationship between autism traits and the pattern of prediction error around environment switches when correcting for speed of movement. This pattern is strikingly similar to differences in AQ around hypothesis switches in Chapter 5 (replicated in Figure 2, below). In both cases, low AQ participants enact the relevant policy when prediction error is greater, and high AQ participants enact the policy when there is less prediction error.

In many ways, these converging results support the hypothesis put forward in Chapter 4, that autistic stimming might be understood as a more ready response to rising uncertainty, leading to the behaviour being pathologised. One very plausible explanation of the pattern of prediction error seen in these two experiments is that participants with

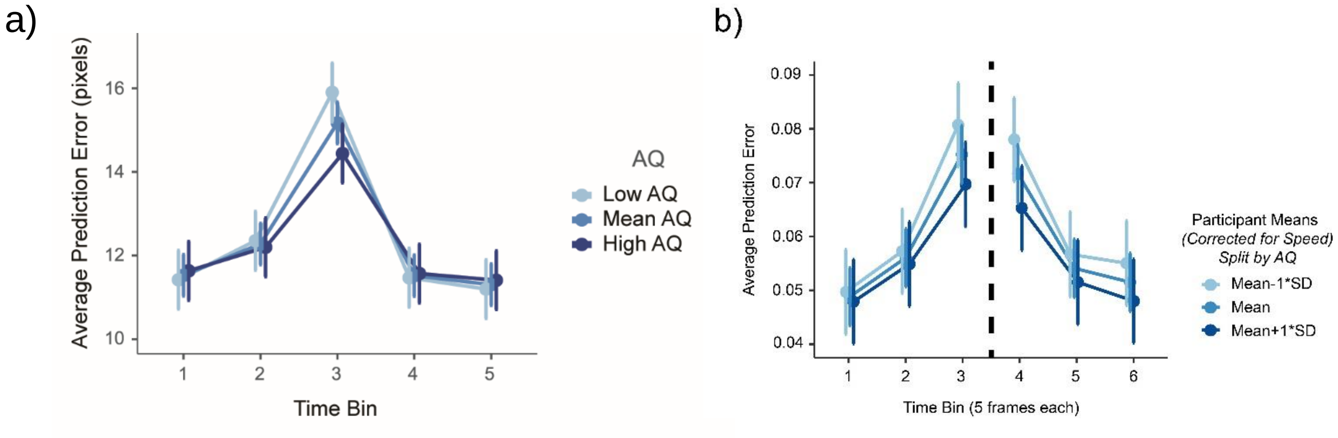


Figure 2 - Comparison of a) Chapter 5 Figure 6c ERPE around hypothesis switches split by AQ and b) Chapter 7 Figure 5b ERPE around environment switches, corrected for speed. In both figures, light blue represents Low AQ as defined by Mean-1*SD, mid-blue represents the mean group and dark blue represents High AQ. Y-axis values differ dramatically due to differences in measurement of XY position in psychtoolbox vs. psychoJS. Time bins are of equal length in each case, and total epoch in both represents ±500ms.

more autistic traits enact policies earlier in response to rising uncertainty. This implies that participants with more autistic traits are treating the prediction errors they receive with more precision. This in turn suggests a higher learning rate with respect to models of the self is associated with autistic traits, supporting predictive processing accounts of autism.

However, one major limitation of the evidence presented in the thesis is that I have no evidence from diagnosed autistic participants. As such, any conclusions that purport to tell us something revealing about autism as a condition fail to stand on the right kind of evidentiary ground. These results give us fascinating suggestions, but can ultimately do no more than that without further research using diagnosed populations. Even if we adopt a more dimensional approach to autism as was discussed in Chapter 3, characteristics at the extreme will be important for deciding the nature of the relevant dimensions, so more participants with a current diagnosis must be involved.

The next chapter will take a critical look at the conclusions that can be drawn from trait based measures of psychiatric conditions using a case study that claimed to confirm the extreme male brain theory of autism using data from over half a million people (Greenberg et al., 2018).

Chapter 8. When Big Data Aren't The Answer

Throughout the experimental chapters of this thesis, I have employed the Autism-Spectrum Quotient (Baron-Cohen et al., 2001) as a measure of autistic traits. Chapter 2 highlighted that the defining features of autism are continually evolving. Further, in Chapter 3, I found that some aspects of cognition that are highly related to autism traits, such as quality of self-concept, are also common to many other psychiatric conditions. Given how quickly we are learning about the core features of autism, it is increasingly problematic that I, along with much of the autism research community who rely on trait-based measures in the general population, primarily use a measure of autistic traits from twenty years ago. This chapter represents a critical analysis and scholarly discussion about some of the limits of such a measure.

The chapter focuses on a commentary I wrote in response to Greenberg et al. (2018). In this paper, authors collected data from 671,606 participants (plus a replication study of 14,354 participants) on four short-form trait questionnaires. Their sample included a large number of diagnosed autistic individuals (36,000+). The traits measured were from the Autism Spectrum Quotient, Empathy Quotient, Systemizing Quotient-Revised and the Sensory Perception Quotient. On the basis of these questionnaires and their combinations alone, the authors claim to provide robust evidence in support of both the Empathising-Systematizing theory of sex differences and the Extreme Male Brain theory of autism. The first of these theories that says that the difference between sexes is best understood as a dichotomy between the ability to empathise (more female) and systematise (more male). The second theory says that autism is best understood as an extreme male presentation on this axis. This theory is also supported in other literature by biological differences, for example prenatal testosterone (which is highlighted in the authors' reply to my commentary, see Greenberg, Warrier, Allison, and Baron-Cohen (2019)).

The original paper, Greenberg et al. (2018), attracted a lot of media attention. As of March 2021, the article is reported by Altmetrics to have been picked up by 55 news outlets, appeared in 11 blogs, was referenced by one policy source, referenced in a Wikipedia page, and appears 531 times in social media including Facebook, Reddit and Twitter. While some of the media articles are critical – e.g. “The ‘female’ brain, why damaging myths about women and science keep coming back in new forms” (Headline, Yahoo! News, 03 Aug 2020) – others applaud the progress made by the discovery – e.g. “Extreme male brain theory of autism confirmed in major new study” (Headline, Newsweek, 13 Nov 2018) or “’Male brain’ autism study backed by biggest-ever study” (Headline, Yahoo! News, 12 Nov 2018). Of course, how the media portrays scientific findings is not straightforward, and there are perverse incentives for all involved parties to overstate findings (Rens & Palghat, 2016). Nevertheless, the study was clearly influential and the impact of its conclusions had the potential to be profound.

However, using the traits questionnaires to support this conclusion was inappropriate. In the following commentary, I argue that the method used in the Greenberg et al. (2018) study to support the Extreme Male Brain theory of autism was problematically circular.

LETTER

When big data aren't the answer

Kelsey Perrykkad[a,1] and Jakob Hohwy[a]

In PNAS, Greenberg, et al. (1) use data collected using 4 surveys from over half a million people to support the Extreme Male Brain (EMB) theory of autism and the Empathizing–Systematizing (E-S) theory of sex differences. Large sample sizes are—all other things being equal—better than small sample sizes. However, the most serious criticisms of these 2 theories (see ref. 2) are not addressed by increasing the sample size.

The questionnaires used by this study were all developed with reference to autism, and are measuring not independent, but interrelated, constructs (3–6). Historically, it has been taken as a given that there is increased prevalence of autism in males. Autism has also been defined based largely on characteristic social difficulties (read: differences in empathizing) and restricted interests in highly patterned stimuli (read: systematized thinking). The Autism Spectrum Quotient (AQ) was developed in the context of these assumptions, and the original paper on AQ took it as reassuring that both high autistic traits, as measured by the AQ, and clinical diagnoses of autism were found to have the same gender trends (5). However, evidence suggests that females have been systematically underdiagnosed and may present with a different clinical profile to their male counterparts (7). This is understandably not reflected by the AQ, given that it was calibrated to fit with the male-biased symptomatology at the time of its conception. So, it is by virtue of its design that male groups have disproportionately high AQ scores.

The 3 other measures [Sensory Perception Quotient (SPQ), Empathy Quotient (EQ), and Systemizing Quotient (SQ)] were all developed and validated with reference to their expected relationship with the AQ in diagnosed autistic populations and in the general population, and thus inherit the AQ's foundational design properties. In the supplemental information of ref. 1, Greenberg et al. state that the short versions of the measures were developed "independently of autism." However, they are a subset of the longer questionnaires, so taking a representative subset of questions cannot justify the claimed independence from autism and the AQ. By design, SPQ correlates with AQ, EQ is anticorrelated with AQ, and SQ is correlated with AQ.

While Greenberg et al. (1) acknowledge concerns about the "risks of convergence across measures," they also claim that "these limitations are offset by. . .big data, an independent replication cohort, and. . .using multiple measures in the same cohorts." Here, we have argued that the associations between scores on these questionnaires (and the participants' sex) should not come as a surprise—in big or small cohorts. Their correlation should also not lead us to believe that autism should be defined by its maleness, or that maleness should be defined by its high systematizing and low empathizing scores. The underlying construct measured by each questionnaire is either the same or very highly correlated, and more prevalent in males by design. Thus, these measures beg the question (in the philosophical sense), and big data don't get us out of this trap. Because researchers can now run large studies online with relative ease, we should be mindful that bigger sample sizes are no substitute for better measures.

1 D. M. Greenberg, V. Warrier, C. Allison, S. Baron-Cohen, Testing the Empathizing–Systemizing theory of sex differences and the Extreme Male Brain theory of autism in half a million people. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12152–12157 (2018).

2 R. Ridley, Some difficulties behind the concept of the 'Extreme male brain' in autism research. A theoretical review. *Res. Autism Spectr. Disord.* **57**, 19–27 (2019).

3 S. Baron-Cohen, J. Richler, D. Bisarya, N. Gurunathan, S. Wheelwright, The systemizing quotient: An investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 361–374 (2003).

4 S. Baron-Cohen, S. Wheelwright, The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* **34**, 163–175 (2004).

[a]Cognition and Philosophy Lab, Philosophy Department, School of Philosophical, Historical, and International Studies, Monash University, Clayton, VIC 3800, Australia

Author contributions: K.P. and J.H. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: Kelsey.perrykkad@monash.edu.

Published online July 3, 2019.

www.pnas.org/cgi/doi/10.1073/pnas.1902050116

5 S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, E. Clubley, The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* **31**, 5–17 (2001).

6 T. Tavassoli, R. A. Hoekstra, S. Baron-Cohen, The Sensory Perception Quotient (SPQ): Development and validation of a new sensory questionnaire for adults with and without autism. *Mol. Autism* **5**, 29 (2014).

7 M. Kirkovski, P. G. Enticott, P. B. Fitzgerald, A review of the role of female gender in autism spectrum disorders. *J. Autism Dev. Disord.* **43**, 2584–2603 (2013).

Following the publication of this commentary, the original authors gave their own reply (Greenberg et al., 2019). In response to this, I posted the following text to the cog-phil-lab.org website on July 18, 2019:

Comments on Greenberg et. al. (2019): When big data are the answer

Kelsey Perrykkad and Jakob Hohwy

In their reply to our letter in PNAS, Greenberg et. al. argue that their original conclusion in favour of the Extreme Male Brain theory of autism was in fact justified based on results from their big data study. They argue that the autism quotient (AQ) does accurately capture [diagnosed cases of] autism in both sexes, contrary to our claims that it may be biased against female symptomatology. Also that the AQ was not designed to distinguish the sexes in a typical population, but show a significant correlation. They provide biological evidence for the Extreme Male Brain theory in addition to the data provided by the original study.

The most important part of their reply, we think, is that we focused on the development of the full original versions of the questionnaires, and their creation with reference to the full AQ and/or a diagnosed autistic sample. We acknowledge that we may have dismissed the importance of the adjustments made to the questionnaires for this study too quickly; all except those used in the replication cohort were shortened versions. The replication cohort did use the full versions of the EQ and SQ to confirm the predictions of the Empathizing-Systematizing theory of sex differences.

While these short versions were not tested against AQ or in autistic populations as part of their validation, they were validated by showing that they highly correlate (*r*=.82-.96) with previous versions of the questionnaires, including the full versions we cited. In other words, they indirectly relate to AQ in virtue of their design, even if they “were not developed to have an expected relationship with the AQ”. Greenberg et al.’s reply further states that using these different versions of the questionnaires means that “the results are to some extent independent of which items are included and rather an indication of effects in the underlying domains”. The specific items in each short version was a subset of the full original questionnaires (except for two items [32 & 33] on the newly developed SQ-R-Short that came from the SQ-R*), and were chosen based on their discrimination index – a measure of how much a specific answer on that question distinguishes between a particularly high or low result on the original questionnaire. In developing a short questionnaire (or a revised questionnaire), we believe a tension arises when claiming both that it is 1) sufficiently similar to the previous version (in both specific item content and correlation in scores) to warrant their use in measuring the same underlying constructs, and 2) different enough from the predecessors to establish conceptual and statistical independence.

We thank the authors for their thoughtful reply, and think this discussion has inspired many interesting thoughts about questionnaires and correlational research. We acknowledge too that development of psychological questionnaires is a difficult and often thankless task, and that the AQ has been hugely influential and remains a cornerstone of autism research. We commend Greenberg et. al. on the work that we know goes into analysing such a large dataset.

*The shortened form of the systematizing quotient (SQ) developed for this study was based on its revised form, which, as Greenburg et al. highlight, attempted to alleviate potential male bias in the content of the questions (by focusing on “mechanical and abstract systems”) by including more traditionally female domains such as “social systems and domestic systems”(Wheelwright et al., 2006, p. 54). Setting aside the question of whether this is a good way to remove gender bias in responses, only two out of the ten questions in the SQ-R-10 were not in the original SQ, so a small proportion of the questionnaire used was tapping into these added “feminine” domains.

One of the problems highlighted by this discussion that is relevant to the claims made in other chapters of this thesis is that a particular measure of autism traits will be validated in reference to the diagnostic criteria in operation at a particular time. Therefore, each measure inherits the scope of autism that that version of the diagnostic criteria defines. If there is an adjustment of the criteria, new measures will be needed to capture the new category. One obvious solution is to restrict ourselves to diagnosed participants, though the process of diagnosis could similarly be seen as a threshold ‘measure’ of the relevant features, and thus fall prey to the same issues. Another solution to this problem in a practical sense is to use multiple measures of autistic traits and hope that they converge. This is also more amenable to the dimensional approach to defining conditions like autism (see Chapter 3).

Over the course of completing this thesis, while perhaps not very evident in the final versions of the manuscripts, I have endeavoured to diversify how I measure autistic traits. Chapter 6 was designed to be completed in a clinical population, but progress was halted due to COVID-19. As part of that study, I have approval to run the Social Responsiveness Scale 2 (John N Constantino & Gruber, 2012) as well as AQ to confirm autism trait levels (John N. Constantino et al., 2003). In both Chapters 5 and 7, I used alternative self-report measures aimed at capturing some of the traits missing from AQ. In choosing additional measures for Chapter 5, I was focused on capturing sensory features of autism. As such, I included the Adult Sensory Questionnaire (Pfeiffer & Kinnealey, 2003) and the Sensory Perception Quotient (Tavassoli, Hoekstra, & Baron-Cohen, 2014) in addition to AQ. Though of course, as we saw here, the Sensory Perception Quotient is not independent of the AQ. However, neither were reported in the final study as I found that they did not clearly correlate with AQ and thus I could not easily interpret their contribution as expected (though both are included in the public dataset associated with the paper, see:10.26180/5ed0708f103a2). In Chapter 7, I collected scores from the Subthreshold Autism Trait Questionnaire (Kanne, Wang, & Christ, 2012) in addition to the AQ.

I do not mean to imply by this discussion that trait based measurements (and especially those measured by the AQ questionnaire) cannot be informative about autism at all, only that we must be careful about the nature of the conclusions we draw from them. I take it that the experimental chapters of this thesis that rely heavily on AQ do begin to tell part of the cognitive story of autism. However, it is also vital that these be replicated using other measures of autistic traits and in diagnosed autistic samples before their conclusions are taken as absolute truth about the autistic self. This is especially true when we keep in mind the instability of the definition of autism (see Chapters 2-3).

Discussion and Conclusion

Summary of thesis findings

In this thesis, I aimed to use the tools of predictive processing to better understand the self in autism. I began, in Chapter 1, with a review of the literature on self-cognition in autism. Across all domains of cognitive processing, previous literature showed differences in self-cognition in autism. For the most part, findings from the experimental work in the rest of the thesis also support this hypothesis.

In Chapter 3, we saw that low level self-prioritisation based on temporary self-associations was not related to autistic traits. I did, however, find that measures of explicit self-concept significantly predicted autism traits score. Chapter 1 also proposed that the cognitive hierarchy of the self-model in autism is flatter – with more representation at fine-grained sensory-oriented lower levels, and fewer resources dedicated to highly integrated and abstracted timescales. This is consistent with findings from Chapter 3 – the low-level implicit measure showed no difference in autism (if anything, I would expect stronger self-advantage with high traits), but the higher order reflective task did. Interestingly, the measures from the higher-order task here did not require 'learning' at all, but rather responses to introspective reflection about self-concept. As such, representations at the highest level didn't need to be updated, but simply reported (assuming fidelity in the introspective report is equivalent across participants). Here, there was a difference found along the spectrum of autism traits. The low-level task used in Chapter 3 required learning and rapidly applying arbitrary associations with the self. Since this is the part of the self that is most flexible in autism, participants with higher autistic traits could likely adapt to temporary associations that affected merely this lowest-level.

This is of course somewhat speculative, and is worth following up in diagnosed populations or with other measures of autism traits (as suggested in Chapter 8). The findings from Chapter 3 also highlight the lack of understanding about the relations between aspects of self-cognition in different domains and across the cognitive hierarchy in the neurotypical case, let alone when and how they differ in psychiatric conditions.

The themes identified as important for distinct features of autistic self-cognition in Chapter 1 also provided interesting avenues for investigation in the squares task experiments. These themes included levels of the neural hierarchy, changing environments, regularities at multiple timescales, self as cause, accumulating model evidence, and active inference. Consistent with previous studies, I never found a main effect of AQ on accuracy in the squares tasks (Chapters 5-7). I did, however, find that patterns of behaviour in completing the task differed with autistic traits scores. None of the previous paradigms incorporated uncertainty in the link between the participants' movements and the movements of the correct square on each trial. Further, the dependent variables used in previous versions of the squares task did not afford questions about dynamic employment of policies in completing the task. Based on the results presented in this thesis, both of these factors proved important in understanding inferences of self as cause in autism.

Across conceptual and experimental chapters of the thesis, I find that autism traits are associated with earlier deployment of prediction error minimising hypotheses under uncertainty. This was first suggested in Chapter 4, where I argued that autistic stimming could be understood as a form of self-evidencing behaviour in response to unexpected uncertainty. I also argued in Chapter 4 that if Chapter 1 were correct, and the self-model in autism had less hierarchical depth and if predictive processing accounts that said autistic participants expected greater environmental volatility were also correct, then autistic individuals would accumulate uncertainty faster than neurotypical. That is, the same amount of environmental uncertainty would be represented as greater in the cognitive system of an autistic person compared to a neurotypical person.

In Chapter 5, participants with more autistic traits were shown to switch hypotheses about which square they controlled at a lower value than participants with fewer autistic traits. The measure of prediction error I used in Chapters 5-7 did not account for changes in prior expectations or based on contextual information. These findings are consistent with the idea that participants with higher autistic traits are internally representing the experienced prediction error as greater than other participants, and act at an earlier threshold in response to it. Remember too, that this data was temporally centered on the hypothesis switch event as part of the data processing, so if this explanation were correct, this is the kind of pattern you would see, rather than a peak in an earlier time bin for one group compared to the others. This pattern was also mirrored in the pattern of prediction error around environment switches (corrected for speed) in Chapter 7. Across these studies, in using these policies, participants with higher autistic trait scores act earlier in response to rising prediction error. In summary, from Chapters 4, 5 and 7, participants with a more autistic phenotype employ prediction error reducing policies – switch hypothesis about what they control, change environments and choose to fidget – more readily in uncertain environments.

As was highlighted at the end of Chapter 7, this finding is consistent with predictive processing accounts of autism. Most predictive processing accounts of autism have a higher learning rate in common. This is thought to happen either through weak priors, high inflexible precision of prediction errors, or high expectations for volatility. In any case then, rising prediction error would be taken as a more precise signal which acts as an imperative to take action to reduce it. Exploring self-cognition in autism using the tools of predictive processing has thus proved to be a successful avenue for research so far.

Throughout the thesis, I continually returned to the question of what the core features of autism should be. This was at the fore in Chapters 1, 2, 3 and 8, but is an underlying current throughout the other chapters too. In Chapter 2, we took a historical look at the changing definition of autism over time, which raised questions of what counts as autistic traits, a theme that was carried forward particularly in the critical look at the AQ in Chapter 8. In the review presented in Chapter 1 and experiment presented in Chapter 3, I addressed the more specific question of whether differences in self-cognition should be one of those core features. The answer? Maybe – it looks promising, but it is complicated. How we should understand the self in autism of course rests heavily on what we take autism to be. Choosing the right conception has consequences not only for the direction and potential success of research moving forward, but also for self-understanding for autistic people, and cultivating optimal social responses to autism.

Looking forward

The ground remains fertile for future research on the self in autism. There are obvious ways of extending the paradigms used in the thesis, but also so many more directions the research could take. As a starting point, the experiments presented here need to be replicated in a diagnosed autistic population. Despite the conceptual difficulties around optimising the diagnostic criteria discussed throughout the thesis, this is an important step in understanding whether policies are actually enacted more readily in autism in the face of rising prediction error, stemming from differences in inferential processes of self-representation. There is also more to understand about how autistic participants build and shape their environments to their epistemic advantage, and whether this differs from neurotypical niche construction.

Further, Chapter 2 opened the door to a broader discussion of the relationship between self-cognitive constructs and psychiatric conditions more broadly. Self-representation has many facets, and different individuals or groups of individuals may have differences across these domains. Understanding which aspects of self-cognition differentiate autism, other psychiatric conditions and the neurotypical phenotype would pave the way for self-representation functioning as a dimensional diagnostic tool.

Along this line, it is also clear that our understanding of self-cognition in the neurotypical case is not very detailed. We have little to no understanding of how these different self-domains interact within one individual. There is broad scope for research on the hierarchical and multimodal structure of self-representation. That is, are there intermediate downstream consequences of differences in perceptual self-prioritisation even though I did not find relationships with self-concept or psychiatric traits? Do high-level predictions, stemming from self-concept clarity, impact on self-representation at lower levels? For this latter question, I have some evidence that they do impact on judgements of agency from the Beach task (Chapter 7), but as this was not the focus of the study there is much more work to be done to understand these dynamics.

I am especially interested to see what the future holds for understanding how we use and build the sense of self as cause of environmental change and ensuing sensory input. One of the biggest novel elements of the squares task experiments presented here is the ability to look at ongoing behavioural dynamics and the use of policies in the active and ongoing process of self-representation. There is still much to learn about policy inference in a naturalistic setting. This could be pushed, for example, in a more social direction, in line with most research in autism. Understanding policy selection in the context of social interactions could speak to the success or failure of interpersonal relationships, even if setting aside autism and looking at personality compatibility for friends or partners. However, there are many ways to make the statistics of the sensory consequences (or the 'environment') more realistic too. Experimental paradigms that allow the participant to actively and permanently alter the environment to their cognitive advantage may teach us a lot about individual differences in cognitive processing. The environmental models used in paradigms like the ones presented here could also be much more complex (and thereby uncertain), more successfully mirroring the real world. The relationship between experienced uncertainty and anxiety also has the potential to inform how we understand the autistic experience. It would be efficient to study action dynamics in gamified experiments, or data-mined games. Many digital games involve survival through complex world dynamics and allow for causal interaction with other hidden causes in the game (for example – picking up objects to use later, building paths that make you move faster etc.). If cognitive scientists could tap into data generated by games to ask controlled questions it might prove a gold mine for understanding policy use.

The final shape of the thesis also emphasises the important role theory has in driving interesting and successful research. It is tempting to go down effect rabbit holes – tweaking paradigms and comparing results to understand idiosyncratic effects. While there is merit to this kind of research, and it is increasingly clear that replication is a very important element in good research practice, we must also remember to be inspired by the big questions that brought us into the research area in the first place. In this thesis, the predictive processing framework provided a theoretical framework which inspired me to look at the intersection of the self and autism. For now, the predictive processing framework appears to be a promising avenue to pursue. For a theory to function in this beneficial way, however, it need not be so far reaching and detailed as predictive processing.

Interdisciplinarity can foster just this kind of theory driven research. In crossing disciplinary boundaries, agreed-upon theories can help keep everyone on the same page, and answering the same questions. In this thesis, philosophy played a vital role in spurring the direction for experimental research, and experimental evidence provided ample fodder for interesting philosophical arguments.

Another important boundary academics should continue to cross lays between academia and the ‘real’ world. For example, in researching autism, it is vital to pay attention to the autistic community and being responsive to their interests and needs. At the very least, this involves being receptive to voiced autistic experiences. The autobiographies I have referenced throughout the thesis reveal that even from the first-person perspective, appropriately representing oneself, forming dependable models, and acting in volatile environments can be difficult for autistic people. In her autobiography, Liane Holliday Wiley says,

> The memories I easily recall are all based on facts I am interested in or situational events that happened in my past. For some reason, I cannot seem to recall how to act as easily as I can recall how I did act. It is as if when I look backwards I see a photo album filled with vivid images and shapes, but when I try to look forward I cannot call to mind one reliable picture to guide me along.
>
> *(Willey, 2014, p. 90)*

She also describes how this kind of experience can dramatically affect one’s life trajectory,

> Most people who come from supportive families learn to jump from their childhood to their young adulthood as if they are on a trampoline. They have the neurological balance to be buoyant and carefree, so that as they move through their experiences they can bounce here and there, making mistakes along the way with the certain confidence that they will be given an opportunity to land on their trampoline and bounce right back up to begin again. People struggling with Asperger's often find there is no trampoline to catch them as they fall, no soft and pliable cushion to propel them back to the beginning for a new and improved, better prepared jump. [Asperger’s Syndrome] makes it difficult to learn from where you have been. It makes it difficult to generalize and problem solve.
>
> *(Willey, 2014, p. 54)*

Personally, and as a research community, I hope to engage in more frequent and successful dialogues with the autistic community about their experiences and needs to guide research in the future.

The self remains an enigma. With this thesis, we have made progress towards better understanding how uncertainty affects self-inferential processes, and how this relates to autism spectrum traits. This has the potential to help us better understand autism, and what kinds of environments are best suited to different kinds of selves. The thesis thus helps guide new ways to study of the self. There is still a long road ahead beckoning us to learn more about ourselves, which I plan to travel.

References in Thesis Text

Note: *This reference list does not include all references listed in publications, but only references in the body of the thesis including linking text, section prefaces and the discussion. References for each chapter can be found following it.*

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*: American Psychiatric Pub.

Andreou, M., & Skrimpa, V. (2020). Theory of Mind Deficits and Neurophysiological Operations in Autism Spectrum Disorders: A Review. *Brain Sciences, 10*(6). doi:10.3390/brainsci10060393

Arslan, M., Warreyn, P., Dewaele, N., Wiersema, J. R., Demurie, E., & Roeyers, H. (2020). Development of neural responses to hearing their own name in infants at low and high risk for autism spectrum disorder. *Developmental Cognitive Neuroscience, 41*, 100739. doi:https://doi.org/10.1016/j.dcn.2019.100739

Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen year review. *Understanding other minds: Perspectives from developmental cognitive neuroscience, 2*, 3-20.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences, 6*(6), 248-254. doi:https://doi.org/10.1016/S1364-6613(02)01904-6

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition, 21*(1), 37-46. doi:https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5-17. doi:10.1023/A:1005653411471

Bednark, J. G., Poonian, S., Palghat, K., McFadyen, J., & Cunnington, R. (2015). Identity-specific predictions and implicit measures of agency. *Psychology of Consciousness: Theory, Research, and Practice, 2*(3), 253.

Berna, F., Göritz, A. S., Schröder, J., Coutelle, R., Danion, J.-M., Cuervo-Lombard, C. V., & Moritz, S. (2016). Self-Disorders in Individuals with Autistic Traits: Contribution of Reduced Autobiographical Reasoning Capacities. *Journal of Autism and Developmental Disorders, 46*(8), 2587-2598. doi:10.1007/s10803-016-2797-2

Bury, S. M., Jellett, R., Spoor, J. R., & Hedley, D. (2020). "It Defines Who I Am" or "It's Something I Have": What Language Do [Autistic] Australian Adults [on the Autism Spectrum] Prefer? *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-020-04425-3

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of personality and social psychology, 70*(1), 141.

Cannon, J., O'Brien, A. M., Bungert, L., & Sinha, P. (2021). Prediction in Autism Spectrum Disorder: A Systematic Review of Empirical Evidence. *Autism Research: Official Journal of the International Society for Autism Research*.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences, 16*(4), 231-239. doi:https://doi.org/10.1016/j.tics.2012.02.007

Clark, A., & Chalmers, D. (1998). The extended mind. *analysis, 58*(1), 7-19.

Constant, A., Bervoets, J., Hens, K., & Cruys, S. V. d. (2020). Precise Worlds for Certain Minds: An ecological perspective on the relational self in autism. *TOPOI, 39*(3), 611-622. doi:10.1007/s11245-018-9546-4

Constant, A., Ramstead, M. J. D., Veissière, S. P. L., Campbell, J. O., & Friston, K. J. (2018). A variational approach to niche construction. *Journal of The Royal Society Interface, 15*(141), 20170685. doi:10.1098/rsif.2017.0685

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., . . . Reich, W. (2003). Validation of a Brief Quantitative Measure of Autistic Traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders, 33*(4), 427-433. doi:10.1023/A:1025014929212

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale: SRS-2*: Western Psychological Services Torrance, CA.

Corlett, P. R. (2017). I Predict, Therefore I Am: Perturbed Predictive Coding Under Ketamine and in Schizophrenia. *Biol Psychiatry, 81*(6), 465-466. doi:10.1016/j.biopsych.2016.12.007

Dinulescu, S., Alvi, T., Rosenfield, D., Sunahara, C. S., Lee, J., & Tabak, B. A. (2020). Self-Referential Processing Predicts Social Cognitive Ability. *Social Psychological and Personality Science,* 1948550620902281.

Dziuk, M., Larson, J., Apostu, A., Mahone, E., Denckla, M., & Mostofsky, S. (2007). Dyspraxia in autism: association with motor, social, and communicative deficits. *Developmental Medicine & Child Neurology, 49*(10), 734-739.

Fabry, R. E. (2021). Limiting the explanatory scope of extended active inference: the implications of a causal pattern analysis of selective niche construction, developmental niche construction, and organism-niche coordination dynamics. *Biology & Philosophy, 36*(1), 6. doi:10.1007/s10539-021-09782-6

Fineberg, S. K., Stahl, D. S., & Corlett, P. R. (2017). Computational Psychiatry in Borderline Personality Disorder. *Current Behavioral Neuroscience Reports, 4*(1), 31-40. doi:10.1007/s40473-017-0104-y

Fineberg, S. K., Steinfeld, M., Brewer, J. A., & Corlett, P. R. (2014). A Computational Account of Borderline Personality Disorder: Impaired Predictive Learning about Self and Others Through Bodily Simulation. *Frontiers in psychiatry, 5*(111). doi:10.3389/fpsyt.2014.00111

Fletcher-Watson, S., & Happé, F. (2019). *Autism: a new introduction to psychological theory and current debate*: Routledge.

Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., & Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: a synthesis and meta-analysis. *Journal of Autism and Developmental Disorders, 40*(10), 1227-1240.

Frith, U., & Happé, F. (1994). Autism: Beyond “theory of mind”. *Cognition, 50*(1), 115-132.

Frith, U., & Happé, F. (1999). Theory of mind and self-consciousness: What is it like to be autistic? *Mind & Language, 14*(1), 82-89.

Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological bulletin, 131*(1), 76.

Gowen, E., & Hamilton, A. (2013). Motor Abilities in Autism: A Review Using a Computational Context. *Journal of Autism and Developmental Disorders, 43*(2), 323-344. doi:10.1007/s10803-012-1574-0

Grainger, C., Williams, D., & Lind, S. E. (2014). Online Action Monitoring and Memory for Self-Performed Actions in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 44*, 1193-1206.

Greenberg, D. M., Warrier, V., Allison, C., & Baron-Cohen, S. (2018). Testing the Empathizing–Systemizing theory of sex differences and the Extreme Male Brain theory of autism in half a million people. *Proceedings of the National Academy of Sciences, 115*(48), 12152-12157. doi:10.1073/pnas.1811032115

Greenberg, D. M., Warrier, V., Allison, C., & Baron-Cohen, S. (2019). Reply to Perrykkad and Hohwy: When big data are the answer. *Proceedings of the National Academy of Sciences, 116*(28), 13740. doi:10.1073/pnas.1903773116

Happé, F. (1999). Autism: cognitive deficit or cognitive style? *Trends in Cognitive Sciences, 3*(6), 216-222.

Happe, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J Autism Dev Disord, 36*(1), 5-25. doi:10.1007/s10803-005-0039-0

Hempel, C. G. (1965). Aspects of scientific explanation.

Hohwy, J. (2016). The Self-Evidencing Brain. *Noûs, 50*(2), 259-285.

Huggins, C., Donnan, G., Cameron, I., & Williams, J. (2020). A systematic review of how emotional self-awareness is defined and measured when comparing autistic and non-autistic groups. *Research in Autism Spectrum Disorders, 77*, 101612.

Hume, D. (1741). A Treatise of Human Nature.

James, W. (1890). *The Principles of Psychology* (Vol. 1). New York: Henry Holt & Co.

Jaswal, V. K., & Akhtar, N. (2019). Being versus appearing socially uninterested: Challenging assumptions about social motivation in autism. *Behavioral and Brain Sciences, 42*.

Kanne, S. M., Wang, J., & Christ, S. E. (2012). The Subthreshold Autism Trait Questionnaire (SATQ): development of a brief self-report measure of subthreshold autism traits. *J Autism Dev Disord, 42*(5), 769-780. doi:10.1007/s10803-011-1308-8

Kanner, L. (1943). Autistic disturbances of affective contact.

Kohls, G., Chevallier, C., Troiani, V., & Schultz, R. T. (2012). Social 'wanting'dysfunction in autism: neurobiological underpinnings and treatment implications. *Journal of Neurodevelopmental Disorders, 4*(1), 1-20.

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature neuroscience*. doi:10.1038/nn.4615

Lecheler, M., Lasser, J., Vaughan, P. W., Leal, J., Ordetx, K., & Bischofberger, M. (2020). A Matter of Perspective: An Exploratory Study of a Theory of Mind Autism Intervention for Adolescents. *Psychological Reports, 124*(1), 39-53. doi:10.1177/0033294119898120

Letheby, C., & Gerrans, P. (2017). Self unbound: ego dissolution in psychedelic experience. *Neuroscience of Consciousness, 3*(1), nix016-nix016. doi:10.1093/nc/nix016

Lind, S. E., Williams, D. M., Nicholson, T., Grainger, C., & Carruthers, P. (2019). The Self-reference Effect on Memory is Not Diminished in Autism: Three Studies of Incidental and Explicit Self-referential Recognition Memory in Autistic and Neurotypical Adults and Adolescents. *Journal of Abnormal Psychology*.

Lipton, P. (2004). *Inference to the best explanation*: Routledge.

Lombardo, M. V., & Baron-Cohen, S. (2010). Unraveling the paradox of the autistic self. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(3), 393-403.

May, T., Sciberras, E., Brignell, A., & Williams, K. (2017). Autism spectrum disorder: updated prevalence and comparison of two birth cohorts in a nationally representative Australian sample. *BMJ Open*, *7*(5). doi:10.1136/bmjopen-2016-015549

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11*(1), 56-60. doi:10.1080/17470215908416289

Nicholson, T., Williams, D., Carpenter, K., & Kallitsounaki, A. (2019). Interoception is Impaired in Children, But Not Adults, with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-019-04079-w

Nijhof, A., Bird, G., Catmur, C., & Shapiro, K. (2020). No evidence for a common self-bias across cognitive domains. *Cognition, 197*.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive Function Deficits in High-Functioning Autistic Individuals: Relationship to Theory of Mind. *Journal of Child Psychology and Psychiatry, 32*(7), 1081-1105. doi:https://doi.org/10.1111/j.1469-7610.1991.tb00351.x

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian Approaches to Autism: Towards Volatility, Action, and Behavior. *Psychological bulletin*.

Palmer, C. J., Paton, B., Kirkovski, M., Enticott, P. G., & Hohwy, J. (2015). Context sensitivity in action decreases along the autism spectrum: a predictive processing perspective. *Proceedings of the Royal Society of London B: Biological Sciences, 282*(1802), 20141557.

Perrykkad, K. (2019). Adaptive behaviour and predictive processing accounts of autism. *Behavioral and Brain Sciences, 42*, e108. doi:10.1017/S0140525X18002248

Perrykkad, K., & Hohwy, J. (2019). When big data aren't the answer. *Proceedings of the National Academy of Sciences, 116*(28), 13738. doi:10.1073/pnas.1902050116

Perrykkad, K., & Hohwy, J. (2020a). Fidgeting as self-evidencing: A predictive processing account of non-goal-directed action. *New Ideas in Psychology, 56*, 100750. doi:https://doi.org/10.1016/j.newideapsych.2019.100750

Perrykkad, K., & Hohwy, J. (2020b). Modelling Me, Modelling You: the Autistic Self. *Review Journal of Autism and Developmental Disorders, 7*, 1-31. doi:10.1007/s40489-019-00173-y

Perrykkad, K., Lawson, R. P., Jamadar, S., & Hohwy, J. (2021). The effect of uncertainty on prediction error in the action perception loop. *Cognition, 210*, 104598. doi:https://doi.org/10.1016/j.cognition.2021.104598

Pfeifer, J. H., Merchant, J. S., Colich, N. L., Hernandez, L. M., Rudie, J. D., & Dapretto, M. (2013). Neural and behavioral responses during self-evaluative processes differ in youth with and without autism. *Journal of Autism and Developmental Disorders, 43*(2), 272-285. doi:10.1007/s10803-012-1563-3

Pfeiffer, B., & Kinnealey, M. (2003). Treatment of sensory defensiveness in adults. *Occupational Therapy International, 10*(3), 175-184.

Poonian, S. K., McFadyen, J., Ogden, J., & Cunnington, R. (2015). Implicit Agency in Observed Actions: Evidence for N1 Suppression of Tones Caused by Self-made and Observed Actions. *Journal of cognitive neuroscience, 27*(4), 752-764. doi:10.1162/jocn_a_00745

Rens, N., & Palghat, K. (2016, 12 May 2016). The danger of overselling science. *The Conversation*.

Rosen, N. E., Lord, C., & Volkmar, F. R. (2021). The Diagnosis of Autism: From Kanner to DSM-III to DSM-5 and Beyond. *Journal of Autism and Developmental Disorders*, 1-18.

Russell, J., & Hill, E. L. (2001). Action-monitoring and Intention Reporting in Children with Autism. *Journal of Child Psychology and Psychiatry, 42*(3), 317-328. doi:10.1111/1469-7610.00725

Sterelny, K. (2010). Minds: extended or scaffolded? *Phenomenology and the Cognitive Sciences, 9*(4), 465-481. doi:10.1007/s11097-010-9174-y

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1105.

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: a meta-analysis. *Psychological bulletin, 121*(3), 371.

Tavassoli, T., Hoekstra, R. A., & Baron-Cohen, S. (2014). The Sensory Perception Quotient (SPQ): development and validation of a new sensory questionnaire for adults with and without autism. *Molecular Autism, 5*(1), 29.

Thomas, R. P., Wang, L. A. L., Guthrie, W., Cola, M., McCleery, J. P., Pandey, J., . . . Miller, J. S. (2019). What's in a name? A preliminary event-related potential study of response to name in preschool children with and without autism spectrum disorder. *PloS one, 14*(5), e0216051. doi:10.1371/journal.pone.0216051

Torres, E., Brincker, M., Isenhower, R., Yanovich, P., Stigler, K., Nurnberger, J. I., . . . Jose, J. (2013). Autism: the micro-movement perspective. *Frontiers in Integrative Neuroscience, 7*(32). doi:10.3389/fnint.2013.00032

Torres, E., & Donnellan, A. M. (2015). *Autism: The movement perspective*: Frontiers Media SA.

Trevisan, D. A., Mehling, W. E., & McPartland, J. C. (2020). Adaptive and Maladaptive Bodily Awareness: Distinguishing Interoceptive Sensibility and Interoceptive Attention from Anxiety-Induced Somatization in Autism and Alexithymia. *Autism Res*. doi:10.1002/aur.2458

Trevisan, D. A., Parker, T., & McPartland, J. C. (2021). First-Hand Accounts of Interoceptive Difficulties in Autistic Adults. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-020-04811-x

Uddin, L. Q. (2011). The self in autism: An emerging view from neuroimaging. *Neurocase, 17*(3), 201-208. doi:10.1080/13554794.2010.509320

Verma, P., & Lahiri, U. (2021). Deficits in Handwriting of Individuals with Autism: a Review on Identification and Intervention Approaches. *Review Journal of Autism and Developmental Disorders*, 1-21.

Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., . . . Wakabayashi, A. (2006). Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain Research, 1079*(1), 47-56. doi:https://doi.org/10.1016/j.brainres.2006.01.012

Willey, L. H. (2014). *Pretending to be Normal: Living with Asperger's Syndrome (Autism Spectrum Disorder) Expanded Edition*: Jessica Kingsley Publishers.

Williams, D. (2010). Theory of own mind in autism: Evidence of a specific deficit in self-awareness? *autism, 14*(5), 474-494. doi:10.1177/1362361310366314

Williams, D., & Happé, F. (2009). Pre-conceptual aspects of self-awareness in autism spectrum disorder: The case of action-monitoring. *Journal of Autism and Developmental Disorders, 39*(2), 251-259.

Wuyun, G., Wang, J., Zhang, L., Wang, K., Yi, L., & Wu, Y. (2020). Actions Speak Louder Than Words: The Role of Action in Self-Referential Advantage in Children With Autism. *Autism Research, n/a*(n/a). doi:10.1002/aur.2274

Addendum 1. Chapter 5 Supplementary Materials

Supplementary Material

Table of Contents

POLICY DEFINITIONS: 218

Horizontal 218
Vertical 218
Perpendicular-Cardinal 218
Non-Cardinal 218
Hesitant-Straight 218
Circle 219

BIASED-NEAREST-OBJECT METHOD FOR HYPOTHESIS DEFINITION: 219

INTERACTION RESULTS FOR MOVEMENT AND STRATEGY VARIABLES: 219

Speed 219
Turn Count 219
Hypothesis Switch Count 220

VOLATILITY SWITCH ERPE RESULTS..220

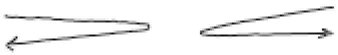
EFFECT OF REMOVING PARTICIPANTS WITH ADHD AND DEPRESSION:..........................220

FULL STATISTICAL MODELS:..221

Accuracy Mixed Model 221
Time Spent Moving Mixed Model 224
Speed Mixed Model 226
Acceleration Mixed Model 229
Jerk Mixed Model 233
Number of Turns Mixed Model 236
Number of Dominant Policy Turns Mixed Model 238
Number of Hypothesis Switches Mixed Model 241
Average Prediction Error Mixed Model: 244
Condition-wise Prediction Error Slope Mixed Model: Accuracy Not Included 248
Condition-wise Prediction Error Mixed Model: Accuracy Random Effect Included 250
Prediction Error Slope by Accuracy and Agency Mixed Model 253
Volatility ERPE Mixed Model 258
Hypothesis Switch ERPE Mixed Model 263
Incorrect Trials Only Hypothesis Switch ERPE Mixed Model 273
Incorrect Trials Only Hypothesis Switch ERPE Figure 278

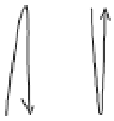
Policy Definitions:

Horizontal



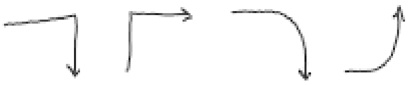
Any turn (surrounded on both sides by at least three frames of direction maintenance) in which the start and end directions are either east or west (right or left).

Vertical



Any turn (surrounded on both sides by at least three frames of direction maintenance) in which the start and end directions are either north or south (up or down).

Perpendicular-Cardinal

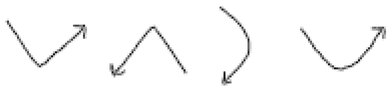


Any turn (surrounded on both sides by at least three frames of direction maintenance) in which the start and end directions are any of the four cardinal directions (right/left/up/down).

OR

Three successive* turns, whose directions move clockwise or anticlockwise but is too short to count as a circle, where the maintenance around the first and last turns are in cardinal directions. This is a *rounded corner* starting and ending in cardinal directions. These get counted as one turn.

Non-Cardinal

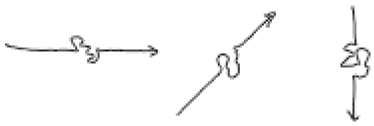


Any turn (surrounded on both sides by at least three frames of direction maintenance) in which the start and end directions are NOT any of the four cardinal directions (diagonals).

OR

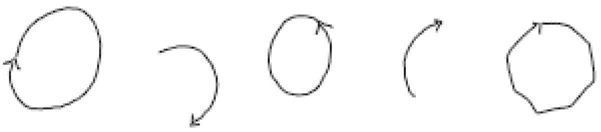
Three successive* turns, whose directions move clockwise or anticlockwise but is too short to count as a circle, where the maintenance around the first and last turns are NOT in cardinal directions. This is a *rounded corner* starting and ending in diagonals. These get counted as one turn.

Hesitant-Straight



Any time when the direction of travel changes for a little bit, but the directions of the maintenance around the 'hesitation' are in the same direction. Maintenance in any direction in the middle of these cannot be longer than 3 frames (otherwise the turns get counted separately). Can be found in the middle of what is otherwise defined as a rounded turn or a circle also. This counts as a 'turn', because it indicates a decision to maintain direction. Likely occurs frequently as a result of picking up and moving the laser mouse while not intending to change direction at all.

Circle



Four successive (not counting adjust-maintains in between) turns whose directions move clockwise or anticlockwise. This counts anything that is at least ¾ of a semi-circle in shape. Each full circle (or any qualifying part of a first circle) gets counted as one turn for this category. Circles are surrounded by at least 3 frames of maintenance in directions non-consistent with a circle, or in the opposite direction of rotation.

*where hesitant-straights in the middle get skipped for counting and labelling this

Biased-nearest-object Method for Hypothesis Definition:

For each frame, the shortest Euclidean distance between the eyes and any square was taken to be the square at which the participant was looking, which was taken as a proxy for their *hypothesis* about which square they controlled at that point in time. When two squares were close together (distance to eye positions is less than one squares' length different) and when the nearest square to each eye differed, the hypothesis was biased towards the square which was hypothesised immediately prior. If this was not within the scope of the nearest squares, the nearest square to the average location of the two eyes was chosen.

Interaction Results for Movement and Strategy Variables:

Where interactions are not reported here, there were no significant interactions.

Speed

There was a significant interaction between variability and volatility (F(1,4661)=16.36, p<0.001). Post-hoc analyses showed that in the low variability/high volatility condition, participants moved faster than the low variability/low volatility (z=3.89, p<0.001), the high variability/high volatility (z=7.18, p<0.001), and the high variability/low volatility conditions (z=5.31, p<0.001). Further, participants moved faster when both variability and volatility were low than when they were both high (z=3.22, p=0.008).

Turn Count

There was a significant interaction between variability and volatility (F(1,4661)=4.47, p=0.034). Post-hoc analysis showed the effect of variability was stronger in low volatility conditions (all pairwise contrasts between conditions were significant, z>11.74, p<0.001, except low/low vs low/high and high/low vs. high/high which were non-significant).

Hypothesis Switch Count

There was also a significant interaction between variability and volatility (**Figure 5**; F(1,4661)=6.45, p=0.011). Post-hoc tests showed that in low variability conditions only, participants switched hypotheses more often under high volatility than low volatility (z=2.79, p=0.031).

Volatility Switch ERPE Results

To look at whether changes to the variability distribution due to volatility lead to any differences in prediction error, we performed an MLM on the ERPE centered on volatility switches. In addition to the standard MLM, we included the fixed effect of time-bin, and an additional fixed effect of average prediction error as a covariate. Of the fixed effects of interest, only variability was marginally significant (F(1,662) = 3.91, p=0.05), such that around the time of volatility switches, high variability trials had greater prediction error than low variability trials (t(662)=1.98, p=0.049). The lack of interaction with time-bin suggests this difference was not temporally sensitive to the onset of a new variability distribution and similar patterns across uncertainty conditions and autism traits.

Effect of Removing Participants with ADHD and Depression:

All models were re-examined with these two participants removed. The following bullet-points report the fixed effects for which the significant effects were altered by removing these two participants.

- For the mixed model using the number of turns per trial as a dependent variable, removing these participants results in a new significant interaction between variability and AQ (F(1,4441)=4.58, p=0.032). This is down from insignificant p = 0.074 in reported sample. Simple effects showed that the difference in variability conditions decreased with increasing AQ score.
- The significant three-way interaction between variability, volatility and AQ for dominant policy use is lost (F(1,4441)=3.23, p=0.072). Up from p = 0.039 in reported sample.
- The marginally significant main effect of volatility for condition-wise prediction error slope is lost (F(1,108)=3.79, p=0.054).
- The marginally significant main effect of variability for the volatility centered ERPE is lost (F(1,630)=3.04, p=0.082).

Full Statistical Models:

For variability and volatility coding, 0 is low, 1 is high.

Accuracy Mixed Model

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	Accuracy ~ 1 + Variability + Volatility + AQ + Variability:Volatility + AQ:Volatility + AQ:Variability + AQ:Volatility:Variability+(1 | id)
AIC	4238.694
R-squared Marginal	0.019
R-squared Conditional	0.068

Model Results

Fixed Effect Omnibus tests

	F	**Num df**	**Den df**	**p**
Variability	85.068	1	4664.477	< .001
Volatility	0.130	1	4664.483	0.718
AQ	0.057	1	37.215	0.813
Variability ✻ Volatility	8.617	1	4664.578	0.003
Volatility ✻ AQ	0.041	1	4662.239	0.840
Variability ✻ AQ	0.585	1	4665.167	0.445
Variability ✻ Volatility ✻ AQ	0.682	1	4664.549	0.409

Note. Satterthwaite method for degrees of freedom

Fixed Effect Omnibus tests

	F	Num df	Den df	p

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	(Intercept)	0.817	0.015	0.789	0.846	37.382	56.026	< .001
Variability1	1 - 0	-0.101	0.011	-0.122	-0.079	4664.477	-9.223	< .001
Volatility1	1 - 0	-0.004	0.011	-0.025	0.017	4664.483	-0.361	0.718
AQ	AQ	5.958e-4	0.003	-0.004	0.006	37.215	0.238	0.813
Variability1 ✻ Volatility1	1 - 0 ✻ 1 - 0	-0.064	0.022	-0.107	-0.021	4664.578	-2.935	0.003
Volatility1 ✻ AQ	1 - 0 ✻ AQ	-3.758e−4	0.002	-0.004	0.003	4662.239	-0.202	0.840
Variability1 ✻ AQ	1 - 0 ✻ AQ	0.001	0.002	-0.002	0.005	4665.167	0.765	0.445
Variability1 ✻ Volatility1 ✻ AQ	1 - 0 ✻ 1 - 0 ✻ AQ	-0.003	0.004	-0.010	0.004	4664.549	-0.826	0.409

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	0.085	0.007	0.050
Residual		0.374	0.140	

Fixed Effect Omnibus tests

	F	Num df	Den df	p

Note. Number of Obs: 4707 , groups: id , 40

Post Hoc Tests

Post Hoc Comparisons - Variability ✻ Volatility

Comparison								
Variability	**Volatility**		**Variability**	**Volatility**	**Difference**	**SE**	**z**	**$p_{bonferroni}$**
0	0	-	0	1	-0.028	0.016	-1.811	0.421
0	0	-	1	0	0.069	0.016	4.414	< .001
0	0	-	1	1	0.105	0.015	6.785	< .001
0	1	-	1	1	0.133	0.015	8.661	< .001
1	0	-	0	1	-0.097	0.015	-6.260	< .001
1	0	-	1	1	0.036	0.015	2.343	0.115

Time Spent Moving Mixed Model

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	TimeSpentMoving ~ 1 + AQ + Volatility + Variability + TimetoMovement + Volatility:Variability + AQ:Variability + AQ:Volatility + AQ:Variability:Volatility+(1 | id)
AIC	13750.596
R-squared Marginal	0.127
R-squared Conditional	0.559

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
AQ	0.374	1	38.003	0.544
Volatility	0.251	1	4660.287	0.616
Variability	727.707	1	4660.277	< .001
TimetoMovement	431.902	1	4674.855	< .001
Volatility ✻ Variability	2.694	1	4660.280	0.101
AQ ✻ Variability	2.489	1	4660.325	0.115
AQ ✻ Volatility	1.702	1	4660.132	0.192
AQ ✻ Volatility ✻ Variability	11.367	1	4660.309	< .001

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**df**	**t**	**p**
(Intercept)	(Intercept)	13.721	0.160	13.408	14.035	38.013	85.780	< .001
AQ	AQ	-0.017	0.027	-0.071	0.037	38.003	-0.612	0.544
Volatility1	1 - 0	-0.015	0.030	-0.073	0.043	4660.287	-0.501	0.616
Variability1	1 - 0	0.801	0.030	0.743	0.860	4660.277	26.976	< .001
TimetoMovement	TimetoMovement	-0.940	0.045	-1.029	-0.852	4674.855	-20.782	< .001
Volatility1 ✻ Variability1	1 - 0 ✻ 1 - 0	-0.098	0.059	-0.214	0.019	4660.280	-1.641	0.101
AQ ✻ Variability1	AQ ✻ 1 - 0	0.008	0.005	-0.002	0.018	4660.325	1.578	0.115
AQ ✻ Volatility1	AQ ✻ 1 - 0	-0.007	0.005	-0.017	0.003	4660.132	-1.305	0.192
AQ ✻ Volatility1 ✻ Variability1	AQ ✻ 1 - 0 ✻ 1 - 0	-0.034	0.010	-0.054	-0.014	4660.309	-3.371	< .001

Random Components

Groups	**Name**	**SD**	**Variance**	**ICC**
id	(Intercept)	1.007	1.014	0.495
Residual		1.017	1.035	

Note. Number of Obs: 4707 , groups: id , 40

Simple Effects

Simple effects of Volatility : Omnibus Tests

Moderator levels				
Variability	**AQ**	**Χ²**	**df**	**p**
0	Mean-1·SD	0.215	1.000	0.643
	Mean	0.643	1.000	0.423
	Mean+1·SD	2.594	1.000	0.107
1	Mean-1·SD	1.617	1.000	0.204
	Mean	2.319	1.000	0.128
	Mean+1·SD	11.553	1.000	< .001

Simple effects of Volatility : Parameter estimates

Moderator levels					**95% Confidence Interval**			
Variability	**AQ**	**contrast**	**Estimate**	**SE**	**Lower**	**Upper**	**z**	**p**
0	Mean-1·SD	1 - 0	-0.028	0.060	-0.145	0.089	-0.463	0.643
	Mean	1 - 0	0.034	0.042	-0.049	0.117	0.802	0.423
	Mean+1·SD	1 - 0	0.095	0.059	-0.021	0.211	1.611	0.107
1	Mean-1·SD	1 - 0	0.075	0.059	-0.041	0.191	1.271	0.204
	Mean	1 - 0	-0.064	0.042	-0.146	0.018	-1.523	0.128
	Mean+1·SD	1 - 0	-0.203	0.060	-0.319	-0.086	-3.399	< .001

Note. Simple effects are estimated keeping constant other independent variable(s) in the model

Speed Mixed Model

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	MeanSpeed ~ 1 + Variability + Volatility + AQ + Variability:Volatility + Variability:AQ + Volatility:AQ + Variability:Volatility:AQ+(1 | id)
AIC	26650.527
R-squared Marginal	0.041
R-squared Conditional	0.641

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Variability	36.418	1	4661.161	< .001
Volatility	2.196	1	4661.164	0.138
AQ	2.443	1	38.004	0.126
Variability ✻ Volatility	16.359	1	4661.166	< .001
Variability ✻ AQ	0.823	1	4661.186	0.364
Volatility ✻ AQ	0.082	1	4661.079	0.775
Variability ✻ Volatility ✻ AQ	0.687	1	4661.165	0.407

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**df**	**t**	**p**
(Intercept)	(Intercept)	13.115	0.820	11.508	14.723	38.011	15.991	< .001
Variability1	1 - 0	-0.706	0.117	-0.935	-0.477	4661.161	-6.035	< .001
Volatility1	1 - 0	0.173	0.117	-0.056	0.403	4661.164	1.482	0.138
AQ	AQ	-0.220	0.141	-0.497	0.056	38.004	-1.563	0.126
Variability1 ✻ Volatility1	1 - 0 ✻ 1 - 0	-0.947	0.234	-1.405	-0.488	4661.166	-4.045	< .001
Variability1 ✻ AQ	1 - 0 ✻ AQ	-0.018	0.020	-0.057	0.021	4661.186	-0.907	0.364
Volatility1 ✻ AQ	1 - 0 ✻ AQ	-0.006	0.020	-0.045	0.033	4661.079	-0.286	0.775
Variability1 ✻ Volatility1 ✻ AQ	1 - 0 ✻ 1 - 0 ✻ AQ	0.033	0.040	-0.045	0.111	4661.165	0.829	0.407

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	5.173	26.758	0.625
Residual		4.007	16.054	

Note. Number of Obs: 4707 , groups: id , 40

Post Hoc Tests

Post Hoc Comparisons - Variability ✻ Volatility

Comparison								
Variability	**Volatility**		**Variability**	**Volatility**	**Difference**	**SE**	**z**	**$p_{bonferroni}$**
0	0	-	0	1	-0.647	0.166	-3.888	< .001

Post Hoc Comparisons - Variability ✻ Volatility

Comparison								
Variability	Volatility		Variability	Volatility	Difference	SE	z	$p_{bonferroni}$
0	0	-	1	0	0.233	0.167	1.397	0.975
0	0	-	1	1	0.533	0.165	3.223	0.008
0	1	-	1	1	1.179	0.164	7.180	< .001
1	0	-	0	1	-0.879	0.166	-5.309	< .001
1	0	-	1	1	0.300	0.165	1.822	0.411

Post Hoc Comparisons - Variability

Comparison						
Variability		Variability	Difference	SE	z	$p_{bonferroni}$
0	-	1	0.706	0.117	6.035	< .001

Acceleration Mixed Model

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	Acceleration ~ 1 + Variability + Volatility + AQ + Variability:Volatility + Variability:AQ + Volatility:AQ + Variability:Volatility:AQ+(1 | id)
AIC	-23077.864
R-squared Marginal	0.013
R-squared Conditional	0.077

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Variability	12.677	1	4664.138	< .001
Volatility	0.779	1	4664.152	0.377
AQ	5.729	1	37.799	0.022
Variability ✻ Volatility	1.056	1	4664.221	0.304
Variability ✻ AQ	0.007	1	4664.674	0.932
Volatility ✻ AQ	0.379	1	4662.388	0.538
Variability ✻ Volatility ✻ AQ	0.247	1	4664.197	0.619

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	(Intercept)	0.013	8.981e-4	0.011	0.014	37.935	14.088	< .001
Variability1	1 - 0	0.002	5.964e-4	9.545e-4	0.003	4664.138	3.560	< .001
Volatility1	1 - 0	-5.265e−4	5.963e-4	-0.002	6.423e-4	4664.152	-0.883	0.377
AQ	AQ	-3.691e−4	1.542e-4	-6.713e−4	-6.684e−5	37.799	-2.393	0.022
Variability1 ✻ Volatility1	1 - 0 ✻ 1 - 0	-0.001	0.001	-0.004	0.001	4664.221	-1.028	0.304
Variability1 ✻ AQ	1 - 0 ✻ AQ	8.737e-6	1.018e-4	-1.908e−4	2.083e-4	4664.674	0.086	0.932
Volatility1 ✻ AQ	1 - 0 ✻ AQ	6.260e-5	1.017e-4	-1.368e−4	2.620e-4	4662.388	0.615	0.538

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
Variability1 ✻ Volatility1 ✻ AQ	1 - 0 ✻ 1 - 0 ✻ AQ	-1.012e−4	2.036e-4	-5.003e−4	2.978e-4	4664.197	-0.497	0.619

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	0.005	2.867e-5	0.064
Residual		0.020	4.172e-4	

Note. Number of Obs: 4707 , groups: id , 40

Post Hoc Tests

Post Hoc Comparisons - Variability

Comparison						
Variability		Variability	Difference	SE	z	$p_{bonferroni}$
0	-	1	-0.002	5.964e-4	-3.560	< .001

Post Hoc Tests

Post Hoc Comparisons - Variability

Comparison						
Variability		Variability	Difference	SE	z	$p_{bonferroni}$
0	-	1	-0.0021233	5.9637e-4	-3.5604	< .001

Linear Regression

Model Fit Measures

Model	R	R²
1	0.104	0.011

Model Coefficients - Acceleration

Predictor	Estimate	SE	t	p
Intercept	0.021	0.001	17.713	< .001
AQ	-3.767e−4	5.248e-5	-7.178	< .001

Jerk Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	Jerk ~ 1 + Variability + Volatility + AQ + Variability:Volatility + Variability:AQ + Volatility:AQ + Variability:Volatility:AQ+(1 | id)
AIC	-32582
R-squared Marginal	8.5604e-4
R-squared Conditional	8.5604e-4

Note. Results may be uninterpretable or misleading. Try to refine your model.

Note. singular fit

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Variability	0.257546	1	4699.0	0.612
Volatility	0.122248	1	4699.0	0.727
AQ	0.282417	1	4699.0	0.595
Variability ✻ Volatility	0.987322	1	4699.0	0.320
Variability ✻ AQ	1.483438	1	4699.0	0.223
Volatility ✻ AQ	0.024998	1	4699.0	0.874
Variability ✻ Volatility ✻ AQ	0.869415	1	4699.0	0.351

Note. singular fit

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**df**	**t**	**p**
(Intercept)	(Intercept)	1.3243e-4	1.0928e-4	-8.1763e–5	3.4662e-4	4699.0	1.21179	0.226
Variability1	1 - 0	1.1092e-4	2.1856e-4	-3.1746e–4	5.3930e-4	4699.0	0.50749	0.612
Volatility1	1 - 0	7.6419e-5	2.1856e-4	-3.5196e–4	5.0480e-4	4699.0	0.34964	0.727
AQ	AQ	9.9132e-6	1.8654e-5	-2.6648e–5	4.6474e-5	4699.0	0.53143	0.595
Variability1 ✻ Volatility1	1 - 0 ✻ 1 - 0	4.3435e-4	4.3713e-4	-4.2241e–4	0.0012911	4699.0	0.99364	0.320
Variability1 ✻ AQ	1 - 0 ✻ AQ	4.5439e-5	3.7308e-5	-2.7682e–5	1.1856e-4	4699.0	1.21796	0.223
Volatility1 ✻ AQ	1 - 0 ✻ AQ	5.8986e-6	3.7308e-5	-6.7223e–5	7.9020e-5	4699.0	0.15811	0.874
Variability1 ✻ Volatility1 ✻ AQ	1 - 0 ✻ 1 - 0 ✻ AQ	-6.9573e–5	7.4615e-5	-2.1582e–4	7.6670e-5	4699.0	-0.93242	0.351

Random Components

Groups	**Name**	**SD**	**Variance**	**ICC**
id	(Intercept)	0.0000000	0.0000	0.0000
Residual		0.0074961	5.6191e-5	

Note. Number of Obs: 4707 , groups: id , 40

Number of Turns Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	nTurns ~ 1 + Variability + Volatility + AQ + Variability:Volatility + Variability:AQ + Volatility:AQ + Variability:Volatility:AQ+(1 | id)
AIC	35054.526405
R-squared Marginal	0.036379
R-squared Conditional	0.532821

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Variability	346.2197567	1	4661.184	< .001
Volatility	0.0086040	1	4661.188	0.926
AQ	0.0675365	1	37.939	0.796
Variability ✻ Volatility	4.4744804	1	4661.192	0.034
Variability ✻ AQ	3.1868060	1	4661.224	0.074
Volatility ✻ AQ	0.5633213	1	4661.055	0.453
Variability ✻ Volatility ✻ AQ	3.7371077	1	4661.190	0.053

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	(Intercept)	35.435365	1.606677	32.2863359	38.5843951	37.949	22.055061	< .001

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
Variability 1	1 - 0	-5.333633	0.286647	-5.8954505	-4.7718154	4661.184	-18.606981	< .001
Volatility 1	1 - 0	-0.026587	0.286632	-0.5883752	0.5352006	4661.188	-0.092758	0.926
AQ	AQ	-0.071753	0.276104	-0.6129075	0.4694008	37.939	-0.259878	0.796
Variability 1 ✻ Volatility 1	1 - 0 ✻ 1 - 0	1.212699	0.573300	0.0890523	2.3363465	4661.192	2.115297	0.034
Variability 1 ✻ AQ	1 - 0 ✻ AQ	0.087372	0.048944	-0.0085554	0.1832997	4661.224	1.785163	0.074
Volatility 1 ✻ AQ	1 - 0 ✻ AQ	0.036691	0.048886	-0.0591237	0.1325065	4661.055	0.750547	0.453
Variability 1 ✻ Volatility 1 ✻ AQ	1 - 0 ✻ 1 - 0 ✻ AQ	-0.189177	0.097859	-0.3809762	0.0026230	4661.190	-1.933160	0.053

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	10.1187	102.388	0.51518
Residual		9.8160	96.353	

Post Hoc Tests

Post Hoc Comparisons - Variability ✻ Volatility

Comparison								
Variability	**Volatility**		**Variability**	**Volatility**	**Difference**	**SE**	**z**	$p_{bonferroni}$
0	0	-	0	1	0.63294	0.40745	1.5534	0.722
0	0	-	1	0	5.93998	0.40833	14.5471	< .001
0	0	-	1	1	5.36022	0.40492	13.2377	< .001
0	1	-	1	1	4.72728	0.40241	11.7473	< .001
1	0	-	0	1	-5.30705	0.40582	-13.0774	< .001
1	0	-	1	1	-0.57976	0.40328	-1.4376	0.903

Post Hoc Comparisons - Variability

Comparison						
Variability		**Variability**	**Difference**	**SE**	**z**	$p_{bonferroni}$
0	-	1	5.3336	0.28665	18.607	< .001

Number of Dominant Policy Turns Mixed Model

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	nDomPol ~ 1 + nTurns + AQ + Variability + Volatility + Variability:Volatility + Variability:AQ + Volatility:AQ + Variability:Volatility:AQ+(1 | id)
AIC	29580.463
R-squared Marginal	0.576
R-squared Conditional	0.633

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
nTurns	3839.669	1	3842.454	< .001
AQ	0.039	1	36.935	0.845
Variability	0.704	1	4684.260	0.401
Volatility	1.807	1	4660.627	0.179
Variability ✻ Volatility	0.574	1	4660.975	0.449
AQ ✻ Variability	1.853	1	4661.195	0.174
AQ ✻ Volatility	19.166	1	4659.774	< .001
AQ ✻ Variability ✻ Volatility	4.273	1	4661.029	0.039

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	(Intercept)	14.082	0.355	13.386	14.778	36.998	39.655	< .001
nTurns	nTurns	0.497	0.008	0.482	0.513	3842.454	61.965	< .001
AQ	AQ	0.012	0.061	-0.108	0.132	36.935	0.196	0.845
Variability1	1 - 0	0.140	0.167	-0.187	0.467	4684.260	0.839	0.401
Volatility1	1 - 0	-0.217	0.161	-0.533	0.099	4660.627	-1.344	0.179

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
Variability1 ✻ Volatility1	1 - 0 ✻ 1 - 0	-0.244	0.323	-0.876	0.388	4660.975	-0.758	0.449
AQ ✻ Variability1	AQ ✻ 1 - 0	0.037	0.028	-0.016	0.091	4661.195	1.361	0.174
AQ ✻ Volatility1	AQ ✻ 1 - 0	-0.120	0.027	-0.174	-0.066	4659.774	-4.378	< .001
AQ ✻ Variability1 ✻ Volatility1	AQ ✻ 1 - 0 ✻ 1 - 0	0.114	0.055	0.006	0.222	4661.029	2.067	0.039

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	2.186	4.780	0.136
Residual		5.520	30.470	

Note. Number of Obs: 4707 , groups: id , 40

Simple Effects

Simple effects of Volatility : Omnibus Tests

Moderator levels			
AQ	X²	df	p
Mean-1·SD	4.597	1.000	0.032
Mean	1.807	1.000	0.179
Mean+1·SD	16.367	1.000	< .001

Simple effects of Volatility : Omnibus Tests

Moderator levels			
AQ	**X²**	**df**	**p**

Simple effects of Volatility : Parameter estimates

Moderator levels				**95% Confidence Interval**			
AQ	**contrast**	**Estimate**	**SE**	**Lower**	**Upper**	**z**	**p**
Mean-1·SD	1 - 0	0.489	0.228	0.042	0.935	2.144	0.032
Mean	1 - 0	-0.217	0.161	-0.533	0.099	-1.344	0.179
Mean+1·SD	1 - 0	-0.922	0.228	-1.369	-0.475	-4.046	< .001

Note. Simple effects are estimated keeping constant other independent variable(s) in the model

Number of Hypothesis Switches Mixed Model

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	nHypSwitches ~ 1 + Variability + Volatility + AQ + Variability:Volatility + AQ:Volatility + AQ:Variability + AQ:Volatility:Variability+(1 | id)
AIC	36948.652
R-squared Marginal	0.020
R-squared Conditional	0.578

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Variability	195.913	1	4661.203	< .001
Volatility	2.045	1	4661.206	0.153
AQ	0.116	1	38.005	0.736
Variability ✻ Volatility	6.446	1	4661.209	0.011
Volatility ✻ AQ	0.143	1	4661.099	0.705
Variability ✻ AQ	0.280	1	4661.234	0.597
Variability ✻ Volatility ✻ AQ	0.508	1	4661.207	0.476

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	(Intercept)	42.218	2.187	37.932	46.505	38.013	19.303	< .001
Variability1	1 - 0	-4.904	0.350	-5.590	-4.217	4661.203	-13.997	< .001
Volatility1	1 - 0	0.501	0.350	-0.186	1.188	4661.206	1.430	0.153
AQ	AQ	-0.128	0.376	-0.865	0.609	38.005	-0.340	0.736
Variability1 ✻ Volatility1	1 - 0 ✻ 1 - 0	-1.779	0.701	-3.152	-0.406	4661.209	-2.539	0.011
Volatility1 ✻ AQ	1 - 0 ✻ AQ	-0.023	0.060	-0.140	0.094	4661.099	-0.379	0.705
Variability1 ✻ AQ	1 - 0 ✻ AQ	-0.032	0.060	-0.149	0.086	4661.234	-0.529	0.597
Variability1 ✻ Volatility1 ✻ AQ	1 - 0 ✻ 1 - 0 ✻ AQ	0.085	0.120	-0.149	0.320	4661.207	0.712	0.476

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	13.785	190.038	0.569
Residual		11.997	143.936	

Note. Number of Obs: 4707 , groups: id , 40

Post Hoc Tests

Post Hoc Comparisons - Variability

Comparison						
Variability		Variability	Difference	SE	z	$p_{bonferroni}$
0	-	1	4.904	0.350	13.997	< .001

Post Hoc Comparisons - Variability ✻ Volatility

Comparison								
Variability	Volatility		Variability	Volatility	Difference	SE	z	$p_{bonferroni}$
0	0	-	0	1	-1.390	0.498	-2.792	0.031
0	0	-	1	0	4.014	0.499	8.044	< .001
0	0	-	1	1	4.403	0.495	8.896	< .001
0	1	-	1	1	5.793	0.492	11.779	< .001
1	0	-	0	1	-5.405	0.496	-10.897	< .001
1	0	-	1	1	0.388	0.493	0.788	1.000

Average Prediction Error Mixed Model:

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	AvPE ~ 1 + Variability + Volatility + AQ + Accuracy + Variability:Volatility + AQ:Volatility + AQ:Variability + Variability:Accuracy + Volatility:Accuracy + AQ:Accuracy + AQ:Volatility:Variability + Variability:Volatility:Accuracy + AQ:Accuracy:Variability + AQ:Accuracy:Volatility:Variability+(1 | id)
AIC	26741.610
R-squared Marginal	0.102
R-squared Conditional	0.634

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Variability	284.0514	1	4653.2	< .001
Volatility	0.5022	1	4653.2	0.479
AQ	1.7937	1	38.4	0.188
Accuracy	172.5723	1	4655.9	< .001
Variability ✻ Volatility	0.2649	1	4653.3	0.607
Volatility ✻ AQ	1.2145	1	4653.2	0.271
Variability ✻ AQ	10.5796	1	4653.2	0.001
Variability ✻ Accuracy	14.1596	1	4653.4	< .001
Volatility ✻ Accuracy	0.0374	1	4653.3	0.847
AQ ✻ Accuracy	0.5401	1	4655.4	0.462
Variability ✻ Volatility ✻ AQ	3.50e-5	1	4653.4	0.995

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Variability ✻ Volatility ✻ Accuracy	6.6103	1	4653.5	0.010
Variability ✻ AQ ✻ Accuracy	0.3085	1	4653.4	0.579
Variability ✻ Volatility ✻ AQ ✻ Accuracy	1.0043	2	4653.3	0.366

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	(Intercept)	11.1392	0.7766	9.6170	12.6614	38.4	14.343	< .001
Variability1	1 - 0	2.6838	0.1592	2.3717	2.9959	4653.2	16.854	< .001
Volatility1	1 - 0	0.1128	0.1592	-0.1991	0.4247	4653.2	0.709	0.479
AQ	AQ	-0.1788	0.1335	-0.4405	0.0829	38.4	-1.339	0.188
Accuracy1	1 - 0	-2.1419	0.1630	-2.4614	-1.8223	4655.9	-13.137	< .001
Variability1 ✻ Volatility1	1 - 0 ✻ 1 - 0	-0.1641	0.3188	-0.7889	0.4607	4653.3	-0.515	0.607
Volatility1 ✻ AQ	1 - 0 ✻ AQ	-0.0310	0.0281	-0.0861	0.0241	4653.2	-1.102	0.271
Variability1 ✻ AQ	1 - 0 ✻ AQ	-0.1043	0.0281	-0.1594	-0.0492	4653.2	-3.708	< .001
Variability1 ✻ Accuracy1	1 - 0 ✻ 1 - 0	1.2002	0.3190	0.5751	1.8253	4653.4	3.763	< .001
Volatility1 ✻ Accuracy1	1 - 0 ✻ 1 - 0	-0.0616	0.3186	-0.6861	0.5629	4653.3	-0.193	0.847

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
AQ ✻ Accuracy1	AQ ✻ 1 - 0	0.0211	0.0287	-0.0351	0.0773	4655.4	0.735	0.462
Variability1 ✻ Volatility1 ✻ AQ	1 - 0 ✻ 1 - 0 ✻ AQ	0.0270	0.0563	-0.0834	0.1374	4653.4	0.479	0.632
Variability1 ✻ Volatility1 ✻ Accuracy1	1 - 0 ✻ 1 - 0 ✻ 1 - 0	-1.6411	0.6383	-2.8921	-0.3900	4653.5	-2.571	0.010
Variability1 ✻ AQ ✻ Accuracy1	1 - 0 ✻ AQ ✻ 1 - 0	-0.0183	0.0563	-0.1286	0.0919	4653.2	-0.326	0.744
Variability0 ✻ Volatility1 ✻ AQ ✻ Accuracy1	Variability0 ✻ 1 - 0 ✻ AQ ✻ 1 - 0	0.0405	0.0887	-0.1333	0.2143	4653.4	0.457	0.648
Variability1 ✻ Volatility1 ✻ AQ ✻ Accuracy1	Variability1 ✻ 1 - 0 ✻ AQ ✻ 1 - 0	0.0932	0.0695	-0.0429	0.2294	4653.2	1.342	0.180

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	4.88	23.9	0.593
Residual		4.05	16.4	

Note. Number of Obs: 4707 , groups: id , 40

Post Hoc Tests

Post Hoc Comparisons - Variability

Comparison						
Variability		Variability	Difference	SE	z	$p_{bonferroni}$
0	-	1	-2.68	0.159	-16.9	< .001

Simple Effects

Simple effects of Variability : Omnibus Tests

Moderator levels			
AQ	X^2	df	p
Mean-1·SD	208.5	1.00	< .001
Mean	284.1	1.00	< .001
Mean+1·SD	81.2	1.00	< .001

Simple effects of Variability : Parameter estimates

Moderator levels				95% Confidence Interval			
AQ	contrast	Estimate	SE	Lower	Upper	z	p
Mean-1·SD	1 - 0	3.29	0.228	2.85	3.74	14.44	< .001
Mean	1 - 0	2.68	0.159	2.37	3.00	16.85	< .001
Mean+1·SD	1 - 0	2.07	0.230	1.62	2.52	9.01	< .001

Note. Simple effects are estimated setting higher order moderator (if any) in covariates to zero and averaging across moderating factors levels (if any)

Condition-wise Prediction Error Slope Mixed Model: Accuracy Not Included

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	ConditionAvPEGradient ~ 1 + Variability + Volatility + AQ + Variability:Volatility + Variability:AQ + Volatility:AQ + Variability:Volatility:AQ+(1 | id)
AIC	-1131.354
R-squared Marginal	0.104
R-squared Conditional	0.749

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Variability	58.154	1	114.000	< .001
Volatility	3.959	1	114.000	0.049
AQ	0.260	1	38.000	0.613
Variability ✻ Volatility	0.217	1	114.000	0.643
Variability ✻ AQ	0.779	1	114.000	0.379
Volatility ✻ AQ	0.052	1	114.000	0.819
Variability ✻ Volatility ✻ AQ	0.020	1	114.000	0.887

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	(Intercept)	4.350e-4	9.956e-4	-0.002	0.002	38.000	0.437	0.665
Variability1	1 - 0	0.005	5.923e-4	0.003	0.006	114.000	7.626	< .001
Volatility1	1 - 0	-0.001	5.923e-4	-0.002	-1.762e−5	114.000	-1.990	0.049
AQ	AQ	8.719e-5	1.711e-4	-2.482e−4	4.226e-4	38.000	0.509	0.613
Variability1 ✻ Volatility1	1 - 0 ✻ 1 - 0	5.514e-4	0.001	-0.002	0.003	114.000	0.465	0.643
Variability1 ✻ AQ	1 - 0 ✻ AQ	-8.987e−5	1.018e-4	-2.894e−4	1.097e-4	114.000	-0.883	0.379
Volatility1 ✻ AQ	1 - 0 ✻ AQ	-2.329e−5	1.018e-4	-2.228e−4	1.763e-4	114.000	-0.229	0.819
Variability1 ✻ Volatility1 ✻ AQ	1 - 0 ✻ 1 - 0 ✻ AQ	-2.900e−5	2.036e-4	-4.281e−4	3.701e-4	114.000	-0.142	0.887

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	0.006	3.614e-5	0.720
Residual		0.004	1.403e-5	

Note. Number of Obs: 160 , groups: id , 40

Post Hoc Tests

Post Hoc Comparisons - Variability

Comparison							
Variability		Variability	Difference	SE	t	df	$p_{bonferroni}$
0	-	1	-0.005	5.923e-4	-7.626	114.000	< .001

Post Hoc Comparisons - Volatility

Comparison							
Volatility		Volatility	Difference	SE	t	df	$p_{bonferroni}$
0	-	1	0.001	5.923e-4	1.990	114.000	0.049

Condition-wise Prediction Error Mixed Model: Accuracy Random Effect Included

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	ConditionAvPEGradient ~ 1 + Variability + Volatility + AQ + Variability:Volatility + Variability:AQ + Volatility:AQ + Variability:Volatility:AQ+(1 | id)+(1 | Accuracy)
AIC	-1214.8500
R-squared Marginal	0.0936
R-squared Conditional	0.7798

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Variability	1.2907	1	2.19e-9	1.000
Volatility	4.2748	1	113.1	0.041
AQ	0.2722	1	38.0	0.605
Variability ✻ Volatility	0.3029	1	113.1	0.583
Variability ✻ AQ	0.8424	1	113.0	0.361
Volatility ✻ AQ	0.0367	1	113.0	0.848
Variability ✻ Volatility ✻ AQ	0.0320	1	113.0	0.858

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	(Intercept)	4.10e-4	0.00222	-0.00395	0.00477	3.27e-9	0.185	1.000
Variability1	1 - 0	0.00457	0.00402	-0.00331	0.01245	2.19e-9	1.136	1.000
Volatility1	1 - 0	-0.00123	5.95e-4	-0.00240	-6.41e−5	113.1	-2.068	0.041
AQ	AQ	8.91e-5	1.71e-4	-2.46e−4	4.24e-4	38.0	0.522	0.605
Variability1 ✻ Volatility1	1 - 0 ✻ 1 - 0	6.55e-4	0.00119	-0.00168	0.00299	113.1	0.550	0.583
Variability1 ✻ AQ	1 - 0 ✻ AQ	-9.36e−5	1.02e-4	-2.94e−4	1.06e-4	113.0	-0.918	0.361
Volatility1 ✻ AQ	1 - 0 ✻ AQ	-1.95e−5	1.02e-4	-2.19e−4	1.80e-4	113.0	-0.192	0.848

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**df**	**t**	**p**
Variability1 ✻ Volatility1 ✻ AQ	1 - 0 ✻ 1 - 0 ✻ AQ	-3.65e−5	2.04e-4	-4.36e−4	3.63e-4	113.0	-0.179	0.858

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	0.00599	3.59e-5	0.719
Accuracy	(Intercept)	0.00281	7.91e-6	0.360
Residual		0.00375	1.41e-5	

Random Effect LRT

Test	N. par	AIC	LRT	df	p
(1 | id)	10	-1003	100	1.00	< .001
(1 | Accuracy)	10	-1103	0	1.00	1.000

Post Hoc Tests

Post Hoc Comparisons - Volatility

Comparison							
Volatility		**Volatility**	**Difference**	**SE**	**t**	**df**	**$p_{bonferroni}$**
0	-	1	0.00123	5.95e-4	2.07	113	0.041

Prediction Error Slope by Accuracy and Agency Mixed Model

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	AgAccAvPEGradient ~ 1 + JudgedAgency + Accuracy + AQ + JudgedAgency:Accuracy + Accuracy:AQ + JudgedAgency:AQ + Accuracy:JudgedAgency:AQ+(1 | id)
AIC	-1083.147
R-squared Marginal	0.168
R-squared Conditional	0.746

Model Results

Fixed Effect Omnibus tests

	F	**Num df**	**Den df**	**p**
JudgedAgency	82.886	1	113.120	< .001
Accuracy	1.279	1	113.120	0.260
AQ	0.001	1	38.007	0.971
JudgedAgency ✻ Accuracy	12.785	1	113.120	< .001
Accuracy ✻ AQ	0.512	1	113.071	0.476

Fixed Effect Omnibus tests

	F	Num df	Den df	p
JudgedAgency ✻ AQ	0.398	1	113.071	0.529
JudgedAgency ✻ Accuracy ✻ AQ	5.684	1	113.071	0.019

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	(Intercept)	0.004	0.001	0.001	0.006	38.057	3.285	0.002
JudgedAgency1	1 - 0	-0.006	6.892e-4	-0.008	-0.005	113.120	-9.104	< .001
Accuracy1	1 - 0	-7.794e−4	6.892e-4	-0.002	5.714e-4	113.120	-1.131	0.260
AQ	AQ	-6.890e−6	1.873e-4	-3.740e−4	3.602e-4	38.007	-0.037	0.971
JudgedAgency1 ✻ Accuracy1	1 - 0 ✻ 1 - 0	-0.005	0.001	-0.008	-0.002	113.120	-3.576	< .001
Accuracy1 ✻ AQ	1 - 0 ✻ AQ	-8.447e−5	1.181e-4	-3.159e−4	1.469e-4	113.071	-0.715	0.476
JudgedAgency1 ✻ AQ	1 - 0 ✻ AQ	7.449e-5	1.181e-4	-1.569e−4	3.059e-4	113.071	0.631	0.529
JudgedAgency1 ✻ Accuracy1 ✻ AQ	1 - 0 ✻ 1 - 0 ✻ AQ	5.630e-4	2.361e-4	1.002e-4	0.001	113.071	2.384	0.019

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	0.007	4.277e-5	0.694

Random Components

Groups	Name	SD	Variance	ICC
Residual		0.004	1.884e-5	

Note. Number of Obs: 159 , groups: id , 40

Post Hoc Tests

Post Hoc Comparisons - JudgedAgency ✻ Accuracy

Comparison									
JudgedAgency	**Accuracy**		**JudgedAgency**	**Accuracy**	**Difference**	**SE**	**t**	**df**	$p_{bonferroni}$
0	0	-	0	1	-0.002	9.707e-4	-1.736	113.001	0.512
0	0	-	1	0	0.004	9.787e-4	3.893	113.119	0.001
0	0	-	1	1	0.007	9.707e-4	7.267	113.001	< .001
0	1	-	1	1	0.009	9.707e-4	9.003	113.001	< .001
1	0	-	0	1	-0.005	9.787e-4	-5.615	113.119	< .001
1	0	-	1	1	0.003	9.787e-4	3.314	113.119	0.007

Post Hoc Comparisons - JudgedAgency

Comparison							
JudgedAgency		**JudgedAgency**	**Difference**	**SE**	**t**	**df**	$p_{bonferroni}$
0	-	1	0.006	6.892e-4	9.104	113.061	< .001

Simple Effects

Simple effects of Accuracy : Omnibus Tests

Moderator levels					
JudgedAgency	AQ	F	Num df	Den df	p
0	Mean-1·SD	7.744	1.000	113.060	0.006
	Mean	3.013	1.000	113.060	0.085
	Mean+1·SD	0.110	1.000	113.060	0.741
1	Mean-1·SD	10.058	1.000	113.180	0.002
	Mean	10.985	1.000	113.180	0.001
	Mean+1·SD	2.294	1.000	113.080	0.133

Simple effects of Accuracy : Parameter estimates

Moderator levels					95% Confidence Interval				
JudgedAgency	AQ	contrast	Estimate	SE	Lower	Upper	df	t	p
0	Mean-1·SD	1 - 0	0.004	0.001	0.001	0.007	113.060	2.783	0.006
	Mean	1 - 0	0.002	9.707e-4	-2.382e−4	0.004	113.060	1.736	0.085
	Mean+1·SD	1 - 0	-4.563e−4	0.001	-0.003	0.002	113.060	-0.331	0.741
1	Mean-1·SD	1 - 0	-0.004	0.001	-0.007	-0.002	113.179	-3.171	0.002
	Mean	1 - 0	-0.003	9.787e-4	-0.005	-0.001	113.178	-3.314	0.001

Simple effects of Accuracy : Omnibus Tests

Moderator levels									
JudgedAgency	AQ	F	Num df	Den df	p				
	Mean+ 1·SD	1 - 0	-0.002	0.001	-0.005	6.439e-4	113.080	-1.515	0.133

Estimated Marginal Means

JudgedAgency

				95% Confidence Interval	
JudgedAgency	Mean	SE	df	Lower	Upper
0	0.007	0.001	45.757	0.004	0.009
1	4.434e-4	0.001	46.012	-0.002	0.003

Note. Estimated means are estimated averaging across interacting variables

Accuracy

				95% Confidence Interval	
Accuracy	Mean	SE	df	Lower	Upper
0	0.004	0.001	46.012	0.002	0.006
1	0.003	0.001	45.757	8.915e-4	0.005

Note. Estimated means are estimated averaging across interacting variables

Accuracy:JudgedAgency

	95% Confidence Interval

JudgedAgency

				95% Confidence Interval		
JudgedAgency	**Mean**	**SE**	**df**	**Lower**	**Upper**	
Accuracy	**JudgedAgency**	**Mean**	**SE**	**df**	**Lower**	**Upper**
0	0	0.006	0.001	62.125	0.003	0.008
1	0	0.008	0.001	62.125	0.005	0.010
0	1	0.002	0.001	63.180	-4.272e−4	0.005
1	1	-0.001	0.001	62.125	-0.004	0.001

Note. Estimated means are estimated keeping constant other independent variable(s) in the model to the mean

Volatility ERPE Mixed Model

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	VolSwitchERPE ~ 1 + avPE + Volatility + Variability + AQ + TimeBin + Volatility:Variability + Volatility:TimeBin + Variability:TimeBin + AQ:TimeBin + AQ:Variability + AQ:Volatility + Volatility:Variability:TimeBin + AQ:TimeBin:Variability + AQ:TimeBin:Volatility + AQ:Variability:Volatility + AQ:TimeBin:Variability:Volatility+(1 | id)

Model Info

Info	
AIC	2504.522
R-squared Marginal	0.962
R-squared Conditional	0.964

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
avPE	10716.298	1	51.346	< .001
Volatility	0.520	1	722.053	0.471
Variability	3.909	1	662.430	0.048
AQ	0.286	1	37.521	0.596
TimeBin	0.155	4	721.814	0.961
Volatility ✻ Variability	0.406	1	722.935	0.524
Volatility ✻ TimeBin	1.000	4	721.814	0.407
Variability ✻ TimeBin	0.293	4	721.814	0.883
AQ ✻ TimeBin	0.067	4	721.814	0.992
Variability ✻ AQ	0.516	1	729.816	0.473
Volatility ✻ AQ	2.289	1	721.995	0.131
Volatility ✻ Variability ✻ TimeBin	0.572	4	721.814	0.683
Variability ✻ AQ ✻ TimeBin	0.533	4	721.814	0.712
Volatility ✻ AQ ✻ TimeBin	0.360	4	721.814	0.837
Volatility ✻ Variability ✻ AQ	0.157	1	722.043	0.692
Volatility ✻ Variability ✻ AQ ✻ TimeBin	0.095	4	721.814	0.984

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	(Intercept)	10.726	0.051	10.626	10.826	37.115	210.753	< .001
avPE	avPE	1.010	0.010	0.991	1.029	51.346	103.520	< .001
Volatility1	2 - 1	0.054	0.076	-0.094	0.203	722.053	0.721	0.471
Variability1	2 - 1	0.162	0.082	0.001	0.324	662.430	1.977	0.048
AQ	AQ	-0.005	0.009	-0.022	0.013	37.521	-0.535	0.596
TimeBin1	2 - 1	0.025	0.119	-0.209	0.259	721.814	0.209	0.834
TimeBin2	3 - 1	0.017	0.119	-0.217	0.251	721.814	0.142	0.887
TimeBin3	4 - 1	-0.046	0.119	-0.280	0.188	721.814	-0.387	0.699
TimeBin4	5 - 1	0.041	0.119	-0.194	0.275	721.814	0.340	0.734
Volatility1 ✻ Variability1	2 - 1 ✻ 2 - 1	-0.096	0.151	-0.393	0.200	722.935	-0.637	0.524
Volatility1 ✻ TimeBin1	2 - 1 ✻ 2 - 1	0.028	0.239	-0.440	0.496	721.814	0.116	0.907
Volatility1 ✻ TimeBin2	2 - 1 ✻ 3 - 1	-0.265	0.239	-0.734	0.203	721.814	-1.110	0.267
Volatility1 ✻ TimeBin3	2 - 1 ✻ 4 - 1	-0.348	0.239	-0.816	0.120	721.814	-1.456	0.146
Volatility1 ✻ TimeBin4	2 - 1 ✻ 5 - 1	-0.048	0.239	-0.517	0.420	721.814	-0.202	0.840
Variability1 ✻ TimeBin1	2 - 1 ✻ 2 - 1	-0.013	0.239	-0.481	0.455	721.814	-0.054	0.957
Variability1 ✻ TimeBin2	2 - 1 ✻ 3 - 1	0.203	0.239	-0.265	0.671	721.814	0.850	0.395
Variability1 ✻ TimeBin3	2 - 1 ✻ 4 - 1	0.103	0.239	-0.366	0.571	721.814	0.430	0.667
Variability1 ✻ TimeBin4	2 - 1 ✻ 5 - 1	0.013	0.239	-0.455	0.482	721.814	0.056	0.956

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
AQ ✻ TimeBin1	AQ ✻ 2 - 1	-0.007	0.021	-0.048	0.033	721.814	-0.362	0.718
AQ ✻ TimeBin2	AQ ✻ 3 - 1	-0.008	0.021	-0.049	0.032	721.814	-0.408	0.683
AQ ✻ TimeBin3	AQ ✻ 4 - 1	-0.005	0.021	-0.045	0.036	721.814	-0.223	0.824
AQ ✻ TimeBin4	AQ ✻ 5 - 1	-0.009	0.021	-0.049	0.031	721.814	-0.450	0.653
Variability1 ✻ AQ	2 - 1 ✻ AQ	-0.009	0.013	-0.035	0.016	729.816	-0.718	0.473
Volatility1 ✻ AQ	2 - 1 ✻ AQ	0.020	0.013	-0.006	0.045	721.995	1.513	0.131
Volatility1 ✻ Variability1 ✻ TimeBin1	2 - 1 ✻ 2 - 1 ✻ 2 - 1	0.356	0.478	-0.581	1.293	721.814	0.745	0.456
Volatility1 ✻ Variability1 ✻ TimeBin2	2 - 1 ✻ 2 - 1 ✻ 3 - 1	0.134	0.478	-0.802	1.071	721.814	0.281	0.779
Volatility1 ✻ Variability1 ✻ TimeBin3	2 - 1 ✻ 2 - 1 ✻ 4 - 1	0.075	0.478	-0.861	1.012	721.814	0.158	0.875
Volatility1 ✻ Variability1 ✻ TimeBin4	2 - 1 ✻ 2 - 1 ✻ 5 - 1	0.630	0.478	-0.307	1.566	721.814	1.318	0.188
Variability1 ✻ AQ ✻ TimeBin1	2 - 1 ✻ AQ ✻ 2 - 1	-0.044	0.041	-0.124	0.036	721.814	-1.071	0.284
Variability1 ✻ AQ ✻ TimeBin2	2 - 1 ✻ AQ ✻ 3 - 1	-0.005	0.041	-0.085	0.076	721.814	-0.114	0.909
Variability1 ✻ AQ ✻ TimeBin3	2 - 1 ✻ AQ ✻ 4 - 1	0.012	0.041	-0.069	0.092	721.814	0.291	0.771
Variability1 ✻ AQ ✻ TimeBin4	2 - 1 ✻ AQ ✻ 5 - 1	-0.016	0.041	-0.096	0.065	721.814	-0.381	0.703
Volatility1 ✻ AQ ✻ TimeBin1	2 - 1 ✻ AQ ✻ 2 - 1	0.007	0.041	-0.073	0.088	721.814	0.171	0.865
Volatility1 ✻ AQ ✻ TimeBin2	2 - 1 ✻ AQ ✻ 3 - 1	0.031	0.041	-0.050	0.111	721.814	0.753	0.452

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**df**	**t**	**p**
Volatility1 ✻ AQ ✻ TimeBin3	2 - 1 ✻ AQ ✻ 4 - 1	0.040	0.041	-0.040	0.121	721.814	0.979	0.328
Volatility1 ✻ AQ ✻ TimeBin4	2 - 1 ✻ AQ ✻ 5 - 1	0.032	0.041	-0.049	0.112	721.814	0.771	0.441
Volatility1 ✻ Variability1 ✻ AQ	2 - 1 ✻ 2 - 1 ✻ AQ	0.010	0.026	-0.041	0.061	722.043	0.396	0.692
Volatility1 ✻ Variability1 ✻ AQ ✻ TimeBin1	2 - 1 ✻ 2 - 1 ✻ AQ ✻ 2 - 1	0.034	0.082	-0.127	0.195	721.814	0.418	0.676
Volatility1 ✻ Variability1 ✻ AQ ✻ TimeBin2	2 - 1 ✻ 2 - 1 ✻ AQ ✻ 3 - 1	0.003	0.082	-0.158	0.164	721.814	0.037	0.970
Volatility1 ✻ Variability1 ✻ AQ ✻ TimeBin3	2 - 1 ✻ 2 - 1 ✻ AQ ✻ 4 - 1	-0.012	0.082	-0.173	0.149	721.814	-0.149	0.882
Volatility1 ✻ Variability1 ✻ AQ ✻ TimeBin4	2 - 1 ✻ 2 - 1 ✻ AQ ✻ 5 - 1	0.017	0.082	-0.144	0.178	721.814	0.212	0.832

Random Components

Groups	**Name**	**SD**	**Variance**	**ICC**
id	(Intercept)	0.216	0.047	0.039
Residual		1.068	1.142	

Post Hoc Tests

Post Hoc Comparisons - Variability

Comparison							
Variability		Variability	Difference	SE	t	df	$p_{bonferroni}$
1	-	2	-0.162	0.082	-1.975	662.208	0.049

Hypothesis Switch ERPE Mixed Model

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	HypSwitchERPE ~ 1 + Variability + Volatility + TimeBin + AQ + avPE + nHypSwitch + Variability:Volatility + Variability:TimeBin + Volatility:TimeBin + Variability:AQ + Volatility:AQ + TimeBin:AQ + Variability:Volatility:TimeBin + Variability:Volatility:AQ + Variability:TimeBin:AQ + Volatility:TimeBin:AQ + Variability:Volatility:TimeBin:AQ+(1 | id)
AIC	2893.939
R-squared Marginal	0.920
R-squared Conditional	0.964

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Variability	125.092	1	511.660	< .001
Volatility	6.140	1	719.046	0.013
TimeBin	252.294	4	718.606	< .001
AQ	0.234	1	36.500	0.632
avPE	1139.785	1	486.130	< .001

Fixed Effect Omnibus tests

	F	Num df	Den df	p
nHypSwitch	8.626	1	271.145	0.004
Variability ✻ Volatility	10.758	1	728.613	0.001
Variability ✻ TimeBin	17.938	4	718.606	< .001
Volatility ✻ TimeBin	0.060	4	718.606	0.993
Variability ✻ AQ	3.443	1	734.027	0.064
Volatility ✻ AQ	7.034e-6	1	718.910	0.998
TimeBin ✻ AQ	12.162	4	718.606	< .001
Variability ✻ Volatility ✻ TimeBin	0.445	4	718.606	0.776
Variability ✻ Volatility ✻ AQ	0.005	1	718.957	0.943
Variability ✻ TimeBin ✻ AQ	0.138	4	718.606	0.968
Volatility ✻ TimeBin ✻ AQ	0.032	4	718.606	0.998
Variability ✻ Volatility ✻ TimeBin ✻ AQ	0.426	4	718.606	0.790

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**df**	**t**	**p**
(Intercept)	(Intercept)	12.384	0.232	11.930	12.838	35.472	53.485	< .001
Variability1	2 - 1	-2.386	0.213	-2.804	-1.968	511.660	-11.184	< .001
Volatility1	2 - 1	0.226	0.091	0.047	0.405	719.046	2.478	0.013
TimeBin1	2 - 1	0.747	0.144	0.465	1.029	718.606	5.184	< .001
TimeBin2	3 - 1	3.655	0.144	3.372	3.937	718.606	25.360	< .001
TimeBin3	4 - 1	-0.006	0.144	-0.289	0.276	718.606	-0.043	0.966
TimeBin4	5 - 1	-0.221	0.144	-0.504	0.061	718.606	-1.537	0.125
AQ	AQ	-0.019	0.040	-0.098	0.059	36.500	-0.484	0.632

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**df**	**t**	**p**
avPE	avPE	1.318	0.039	1.241	1.395	486.130	33.761	< .001
nHypSwitch	nHypSwitch	-0.049	0.017	-0.081	-0.016	271.145	-2.937	0.004
Variability1 ✻ Volatility1	2 - 1 ✻ 2 - 1	-0.603	0.184	-0.963	-0.243	728.613	-3.280	0.001
Variability1 ✻ TimeBin1	2 - 1 ✻ 2 - 1	-0.467	0.288	-1.032	0.098	718.606	-1.620	0.106
Variability1 ✻ TimeBin2	2 - 1 ✻ 3 - 1	-2.158	0.288	-2.723	-1.593	718.606	-7.488	< .001
Variability1 ✻ TimeBin3	2 - 1 ✻ 4 - 1	-0.433	0.288	-0.997	0.132	718.606	-1.501	0.134
Variability1 ✻ TimeBin4	2 - 1 ✻ 5 - 1	-0.200	0.288	-0.765	0.364	718.606	-0.695	0.487
Volatility1 ✻ TimeBin1	2 - 1 ✻ 2 - 1	0.044	0.288	-0.521	0.609	718.606	0.153	0.879
Volatility1 ✻ TimeBin2	2 - 1 ✻ 3 - 1	0.133	0.288	-0.432	0.697	718.606	0.460	0.646
Volatility1 ✻ TimeBin3	2 - 1 ✻ 4 - 1	0.052	0.288	-0.513	0.617	718.606	0.181	0.857
Volatility1 ✻ TimeBin4	2 - 1 ✻ 5 - 1	0.024	0.288	-0.541	0.589	718.606	0.083	0.934
Variability1 ✻ AQ	2 - 1 ✻ AQ	0.030	0.016	-0.002	0.062	734.027	1.855	0.064
Volatility1 ✻ AQ	2 - 1 ✻ AQ	-4.157e−5	0.016	-0.031	0.031	718.910	-0.003	0.998
TimeBin1 ✻ AQ	2 - 1 ✻ AQ	-0.031	0.025	-0.080	0.017	718.606	-1.268	0.205
TimeBin2 ✻ AQ	3 - 1 ✻ AQ	-0.144	0.025	-0.192	-0.095	718.606	-5.804	< .001
TimeBin3 ✻ AQ	4 - 1 ✻ AQ	-0.009	0.025	-0.058	0.039	718.606	-0.381	0.703

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
TimeBin4 ✻ AQ	5 - 1 ✻ AQ	-2.090e−5	0.025	-0.049	0.049	718.606	-8.436e−4	0.999
Variability1 ✻ Volatility1 ✻ TimeBin1	2 - 1 ✻ 2 - 1 ✻ 2 - 1	-0.217	0.576	-1.347	0.912	718.606	-0.377	0.706
Variability1 ✻ Volatility1 ✻ TimeBin2	2 - 1 ✻ 2 - 1 ✻ 3 - 1	-0.707	0.576	-1.837	0.422	718.606	-1.227	0.220
Variability1 ✻ Volatility1 ✻ TimeBin3	2 - 1 ✻ 2 - 1 ✻ 4 - 1	-0.420	0.576	-1.549	0.710	718.606	-0.728	0.467
Variability1 ✻ Volatility1 ✻ TimeBin4	2 - 1 ✻ 2 - 1 ✻ 5 - 1	-0.164	0.576	-1.294	0.966	718.606	-0.284	0.776
Variability1 ✻ Volatility1 ✻ AQ	2 - 1 ✻ 2 - 1 ✻ AQ	0.002	0.031	-0.059	0.064	718.957	0.071	0.943
Variability1 ✻ TimeBin1 ✻ AQ	2 - 1 ✻ 2 - 1 ✻ AQ	-0.017	0.050	-0.114	0.080	718.606	-0.347	0.728
Variability1 ✻ TimeBin2 ✻ AQ	2 - 1 ✻ 3 - 1 ✻ AQ	0.019	0.050	-0.079	0.116	718.606	0.374	0.708
Variability1 ✻ TimeBin3 ✻ AQ	2 - 1 ✻ 4 - 1 ✻ AQ	0.002	0.050	-0.095	0.099	718.606	0.033	0.974
Variability1 ✻ TimeBin4 ✻ AQ	2 - 1 ✻ 5 - 1 ✻ AQ	-0.006	0.050	-0.103	0.091	718.606	-0.120	0.905
Volatility1 ✻ TimeBin1 ✻ AQ	2 - 1 ✻ 2 - 1 ✻ AQ	-0.007	0.050	-0.104	0.090	718.606	-0.147	0.883
Volatility1 ✻ TimeBin2 ✻ AQ	2 - 1 ✻ 3 - 1 ✻ AQ	-0.011	0.050	-0.109	0.086	718.606	-0.232	0.817

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**df**	**t**	**p**
Volatility1 ✻ TimeBin3 ✻ AQ	2 - 1 ✻ 4 - 1 ✻ AQ	-0.017	0.050	-0.114	0.080	718.606	-0.343	0.732
Volatility1 ✻ TimeBin4 ✻ AQ	2 - 1 ✻ 5 - 1 ✻ AQ	-0.012	0.050	-0.109	0.086	718.606	-0.233	0.816
Variability1 ✻ Volatility1 ✻ TimeBin1 ✻ AQ	2 - 1 ✻ 2 - 1 ✻ 2 - 1 ✻ AQ	0.064	0.099	-0.130	0.258	718.606	0.643	0.520
Variability1 ✻ Volatility1 ✻ TimeBin2 ✻ AQ	2 - 1 ✻ 2 - 1 ✻ 3 - 1 ✻ AQ	0.126	0.099	-0.068	0.320	718.606	1.271	0.204
Variability1 ✻ Volatility1 ✻ TimeBin3 ✻ AQ	2 - 1 ✻ 2 - 1 ✻ 4 - 1 ✻ AQ	0.071	0.099	-0.123	0.265	718.606	0.716	0.474
Variability1 ✻ Volatility1 ✻ TimeBin4 ✻ AQ	2 - 1 ✻ 2 - 1 ✻ 5 - 1 ✻ AQ	0.043	0.099	-0.151	0.237	718.606	0.436	0.663

Random Components

Groups	**Name**	**SD**	**Variance**	**ICC**
id	(Intercept)	1.436	2.062	0.554
Residual		1.289	1.661	

Post Hoc Tests

Post Hoc Comparisons - Variability

Comparison							
Variability		Variability	Difference	SE	t	df	$p_{bonferroni}$
1	-	2	2.386	0.215	11.094	518.881	< .001

Post Hoc Comparisons - Volatility

Comparison							
Volatility		Volatility	Difference	SE	t	df	$p_{bonferroni}$
1	-	2	-0.226	0.091	-2.478	720.631	0.013

Post Hoc Comparisons - TimeBin

Comparison							
TimeBin		TimeBin	Difference	SE	t	df	$p_{bonferroni}$
2	-	3	-2.908	0.144	-20.176	720.208	< .001
2	-	4	0.753	0.144	5.227	720.208	< .001
2	-	5	0.968	0.144	6.721	720.208	< .001
1	-	2	-0.747	0.144	-5.184	720.208	< .001
1	-	3	-3.655	0.144	-25.360	720.208	< .001
1	-	4	0.006	0.144	0.043	720.208	1.000
1	-	5	0.221	0.144	1.537	720.208	1.000
3	-	4	3.661	0.144	25.403	720.208	< .001
3	-	5	3.876	0.144	26.897	720.208	< .001
4	-	5	0.215	0.144	1.494	720.208	1.000

Post Hoc Comparisons - Variability ✻ Volatility

Comparison									
Variability	**Volatility**		**Variability**	**Volatility**	**Difference**	**SE**	**t**	**df**	**$p_{bonferroni}$**
2	1	-	2	2	0.076	0.129	0.584	723.906	1.000
2	1	-	1	2	-2.612	0.233	-11.218	584.968	< .001
1	2	-	2	2	2.687	0.239	11.223	565.458	< .001
1	1	-	2	2	2.160	0.234	9.217	589.917	< .001
1	1	-	2	1	2.084	0.228	9.135	608.443	< .001
1	1	-	1	2	-0.527	0.130	-4.068	726.869	< .001

Post Hoc Comparisons - Variability ✻ TimeBin

Comparison									
Variability	**TimeBin**		**Variability**	**TimeBin**	**Difference**	**SE**	**t**	**df**	**$p_{bonferroni}$**
2	2	-	2	3	-2.062	0.204	-10.117	720.208	< .001
2	2	-	2	4	0.736	0.204	3.612	720.208	0.015
2	2	-	2	5	0.835	0.204	4.099	720.208	0.002
2	2	-	1	3	-5.954	0.282	-21.121	697.056	< .001
2	2	-	1	4	-1.431	0.282	-5.075	697.056	< .001
2	2	-	1	5	-1.099	0.282	-3.900	697.056	0.005
2	1	-	2	2	-0.514	0.204	-2.520	720.208	0.537
2	1	-	2	3	-2.575	0.204	-12.638	720.208	< .001
2	1	-	2	4	0.222	0.204	1.091	720.208	1.000
2	1	-	2	5	0.322	0.204	1.578	720.208	1.000
2	1	-	1	2	-2.715	0.282	-9.630	697.056	< .001
2	1	-	1	3	-6.468	0.282	-22.943	697.056	< .001
2	1	-	1	4	-1.944	0.282	-6.897	697.056	< .001

Post Hoc Comparisons - Variability ✻ TimeBin

Comparison									
Variability	**TimeBin**		**Variability**	**TimeBin**	**Difference**	**SE**	**t**	**df**	**$p_{bonferroni}$**
2	1	-	1	5	-1.613	0.282	-5.722	697.056	< .001
2	3	-	2	4	2.798	0.204	13.729	720.208	< .001
2	3	-	2	5	2.897	0.204	14.216	720.208	< .001
2	3	-	1	4	0.631	0.282	2.239	697.056	1.000
2	3	-	1	5	0.962	0.282	3.414	697.056	0.030
2	4	-	2	5	0.099	0.204	0.487	720.208	1.000
2	4	-	1	5	-1.835	0.282	-6.511	697.056	< .001
1	2	-	2	2	2.201	0.282	7.808	697.056	< .001
1	2	-	2	3	0.139	0.282	0.494	697.056	1.000
1	2	-	2	4	2.937	0.282	10.419	697.056	< .001
1	2	-	2	5	3.036	0.282	10.771	697.056	< .001
1	2	-	1	3	-3.753	0.204	-18.416	720.208	< .001
1	2	-	1	4	0.770	0.204	3.780	720.208	0.008
1	2	-	1	5	1.102	0.204	5.406	720.208	< .001
1	1	-	2	2	1.221	0.282	4.330	697.056	< .001
1	1	-	2	1	1.734	0.282	6.152	697.056	< .001
1	1	-	2	3	-0.841	0.282	-2.984	697.056	0.132
1	1	-	2	4	1.957	0.282	6.941	697.056	< .001
1	1	-	2	5	2.056	0.282	7.293	697.056	< .001
1	1	-	1	2	-0.980	0.204	-4.811	720.208	< .001
1	1	-	1	3	-4.734	0.204	-23.227	720.208	< .001
1	1	-	1	4	-0.210	0.204	-1.031	720.208	1.000
1	1	-	1	5	0.121	0.204	0.595	720.208	1.000
1	3	-	2	3	3.892	0.282	13.807	697.056	< .001
1	3	-	2	4	6.690	0.282	23.732	697.056	< .001

Post Hoc Comparisons - Variability ✻ TimeBin

Comparison									
Variability	TimeBin		Variability	TimeBin	Difference	SE	t	df	$p_{bonferroni}$
1	3	-	2	5	6.789	0.282	24.084	697.056	< .001
1	3	-	1	4	4.523	0.204	22.196	720.208	< .001
1	3	-	1	5	4.855	0.204	23.822	720.208	< .001
1	4	-	2	4	2.167	0.282	7.686	697.056	< .001
1	4	-	2	5	2.266	0.282	8.038	697.056	< .001
1	4	-	1	5	0.331	0.204	1.626	720.208	1.000
1	5	-	2	5	1.935	0.282	6.863	697.056	< .001

Simple Effects

Simple effects of AQ : Omnibus Tests

Moderator levels				
TimeBin	F	Num df	Den df	p
1	0.165	1.000	50.450	0.686
2	0.104	1.000	50.450	0.748
3	8.576	1.000	50.450	0.005
4	0.035	1.000	50.450	0.852
5	0.165	1.000	50.450	0.687

Simple effects of AQ : Parameter estimates

Moderator levels			95% Confidence Interval				
TimeBin	Estimate	SE	Lower	Upper	df	t	p
1	0.018	0.043	-0.069	0.104	50.450	0.406	0.686
2	-0.014	0.043	-0.100	0.073	50.450	-0.322	0.748

Simple effects of AQ : Parameter estimates

Moderator levels			95% Confidence Interval				
TimeBin	Estimate	SE	Lower	Upper	df	t	p
3	-0.126	0.043	-0.213	-0.040	50.450	-2.928	0.005
4	0.008	0.043	-0.079	0.095	50.450	0.187	0.852
5	0.017	0.043	-0.069	0.104	50.450	0.406	0.687

Note. Simple effects are estimated setting higher order moderator (if any) in covariates to zero and averaging across moderating factors levels (if any)

Correlation Matrix

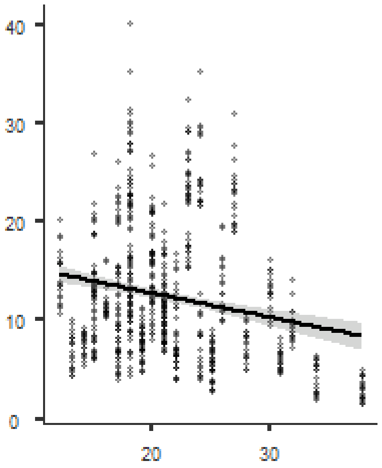
Correlation Matrix

		AQ	HypSwitchERPE
AQ	Pearson's r	—	
	p-value	—	
HypSwitchERPE	Pearson's r	-0.211	—
	p-value	< .001	—

Plot

AQ HypSwitchERPE

AQ



HypSwitchERPE

Incorrect Trials Only Hypothesis Switch ERPE Mixed Model

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	HypSwitchIncorrectERPE ~ 1 + Variability + Volatility + AQ + TimeBin + nHypSwitch + avPE + Variability:Volatility + Variability:TimeBin + Volatility:TimeBin + Variability:Volatility:TimeBin+(1 | id)
AIC	4222.870
R-squared Marginal	0.697
R-squared Conditional	0.800

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Variability	44.1261	1	300.6	< .001
Volatility	12.2004	1	721.3	< .001
AQ	0.8675	1	37.2	0.358
TimeBin	22.8229	4	718.8	< .001
nHypSwitch	2.1231	1	132.1	0.147
avPE	213.9828	1	236.7	< .001
Variability ✻ Volatility	0.0233	1	729.1	0.879
Variability ✻ TimeBin	1.8091	4	718.8	0.125
Volatility ✻ TimeBin	0.1829	4	718.8	0.947
Variability ✻ Volatility ✻ TimeBin	0.1215	4	718.8	0.975

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	(Intercept)	13.6946	0.3987	12.913	14.4761	35.9	34.346	< .001
Variability1	2 - 1	-3.2973	0.4964	-4.270	-2.3244	300.6	-6.643	< .001
Volatility1	2 - 1	0.8430	0.2413	0.370	1.3160	721.3	3.493	< .001
AQ	AQ	-0.0647	0.0695	-0.201	0.0715	37.2	-0.931	0.358
TimeBin1	2 - 1	0.5932	0.3797	-0.151	1.3374	718.8	1.562	0.119
TimeBin2	3 - 1	2.8187	0.3797	2.074	3.5629	718.8	7.423	< .001
TimeBin3	4 - 1	-0.1177	0.3797	-0.862	0.6265	718.8	-0.310	0.757

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**df**	**t**	**p**
TimeBin4	5 - 1	-0.2853	0.3797	-1.030	0.4589	718.8	-0.751	0.453
nHypSwitch	nHypSwitch	-0.0529	0.0363	-0.124	0.0182	132.1	-1.457	0.147
avPE	avPE	1.2958	0.0886	1.122	1.4694	236.7	14.628	< .001
Variability1 ✻ Volatility1	2 - 1 ✻ 2 - 1	-0.0740	0.4854	-1.025	0.8773	729.1	-0.153	0.879
Variability1 ✻ TimeBin1	2 - 1 ✻ 2 - 1	-0.5160	0.7594	-2.004	0.9725	718.8	-0.679	0.497
Variability1 ✻ TimeBin2	2 - 1 ✻ 3 - 1	-1.8104	0.7594	-3.299	-0.3219	718.8	-2.384	0.017
Variability1 ✻ TimeBin3	2 - 1 ✻ 4 - 1	-0.3250	0.7594	-1.813	1.1635	718.8	-0.428	0.669
Variability1 ✻ TimeBin4	2 - 1 ✻ 5 - 1	-0.1695	0.7594	-1.658	1.3190	718.8	-0.223	0.823
Volatility1 ✻ TimeBin1	2 - 1 ✻ 2 - 1	0.3876	0.7594	-1.101	1.8761	718.8	0.510	0.610
Volatility1 ✻ TimeBin2	2 - 1 ✻ 3 - 1	0.6374	0.7594	-0.851	2.1259	718.8	0.839	0.402
Volatility1 ✻ TimeBin3	2 - 1 ✻ 4 - 1	0.3769	0.7594	-1.112	1.8654	718.8	0.496	0.620
Volatility1 ✻ TimeBin4	2 - 1 ✻ 5 - 1	0.4192	0.7594	-1.069	1.9076	718.8	0.552	0.581
Variability1 ✻ Volatility1 ✻ TimeBin1	2 - 1 ✻ 2 - 1 ✻ 2 - 1	-0.4618	1.5189	-3.439	2.5151	718.8	-0.304	0.761
Variability1 ✻ Volatility1 ✻ TimeBin2	2 - 1 ✻ 2 - 1 ✻ 3 - 1	-0.8884	1.5189	-3.865	2.0886	718.8	-0.585	0.559
Variability1 ✻ Volatility1 ✻ TimeBin3	2 - 1 ✻ 2 - 1 ✻ 4 - 1	-0.7765	1.5189	-3.753	2.2004	718.8	-0.511	0.609

Fixed Effects Parameter Estimates

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	t	p
Variability1 ✻ Volatility1 ✻ TimeBin4	2 - 1 ✻ 2 - 1 ✻ 5 - 1	-0.8593	1.5189	-3.836	2.1177	718.8	-0.566	0.572

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	2.40	5.77	0.339
Residual		3.35	11.23	

Post Hoc Tests

Post Hoc Comparisons - TimeBin

Comparison							
TimeBin		TimeBin	Difference	SE	t	df	$p_{bonferroni}$
2	-	3	-2.225	0.380	-5.861	720	< .001
2	-	4	0.711	0.380	1.872	720	0.616
2	-	5	0.879	0.380	2.314	720	0.210
1	-	2	-0.593	0.380	-1.562	720	1.000
1	-	3	-2.819	0.380	-7.423	720	< .001
1	-	4	0.118	0.380	0.310	720	1.000
1	-	5	0.285	0.380	0.751	720	1.000
3	-	4	2.936	0.380	7.733	720	< .001

Post Hoc Comparisons - TimeBin

Comparison							
TimeBin		**TimeBin**	**Difference**	**SE**	**t**	**df**	**$p_{bonferroni}$**
3	-	5	3.104	0.380	8.174	720	< .001
4	-	5	0.168	0.380	0.441	720	1.000

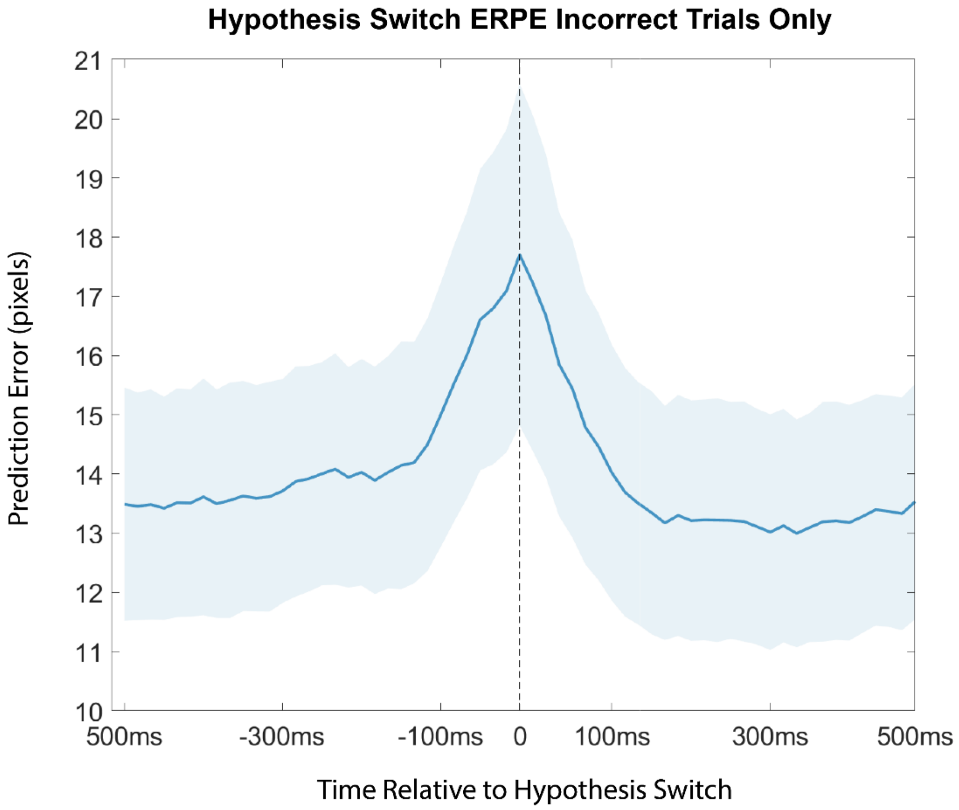
Post Hoc Comparisons - Volatility

Comparison							
Volatility		**Volatility**	**Difference**	**SE**	**t**	**df**	**$p_{bonferroni}$**
1	-	2	-0.843	0.241	-3.49	722	< .001

Post Hoc Comparisons - Variability

Comparison							
Variability		**Variability**	**Difference**	**SE**	**t**	**df**	**$p_{bonferroni}$**
1	-	2	3.30	0.501	6.58	304	< .001

Incorrect Trials Only Hypothesis Switch ERPE Figure



The grand average (blue line) prediction error across participants in a one second epoch centered on hypothesis switches for incorrect trials only. Shading represents a 95% CI.