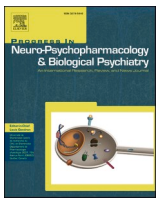
Progress in Neuropsychopharmacology & Biological Psychiatry 136 (2025) 111206

Contents lists available at ScienceDirect

Progress in Neuropsychopharmacology & Biological Psychiatry

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



How dopamine shapes trust beliefs

Bianca A. Schuster *, Claus Lamm

Department of Cognition, Emotion, and Methods in Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

ARTICLE INFO

Keywords
Dopamine
Trust
Social learning
Bayesian inference
Paranoia

ABSTRACT

Learning whom to trust is integral for healthy relationships and social cohesion, and atypicalities in trust learning are common across a range of clinical conditions, including schizophrenia spectrum disorders, Parkinson's disease, and depression. Persecutory delusions – rigid, unfounded beliefs that others are intending to harm oneself – significantly impact affected individuals' lives as they are associated with a range of negative health outcomes, including suicidal behaviour and relapse. Recent advances in computational modelling and psychopharmacology have significantly extended our understanding of the brain bases of dynamic trust learning, and the neuromodulator dopamine has been suggested to play a key role in this. However, the specifics of this role on a computational and neurobiological level remain to be fully established. The current review article provides a comprehensive summary of novel conceptual developments and empirical findings regarding the computational role of dopamine in social learning processes. Research findings strongly suggest a conceptual shift, from dopamine as a reward mechanism to a teaching signal indicating which information is relevant for learning, and shed light on the neurocomputational mechanisms via which antipsychotics may alleviate symptoms of aberrant social learning processes such as persecutory delusions. Knowledge gaps and inconsistencies in the extant literature are examined and the most pressing issues highlighted, laying the foundation for future research that will further advance our understanding of the neuromodulation of social belief updating processes.

1. Introduction

Forming accurate representations of other people's intentions not only helps us succesfully navigate social relationships, but is highly adaptive beyond promoting group cohesion and cooperation. For instance, establishing whether we can trust others is essential for survival, as this informs us about a potential direct threat (are others going to harm us or are their intentions towards us favourable?) and guides our learning under uncertainty (can we trust information provided by others when our own knowledge is limited, as is the case when following guidance by authorities like politicians or scientists?).

In social situations, individuals learn whom to trust based on a combination of others' reputations (i.e., prior knowledge about individual dispositions and traits) and their actions across time. However, our knowledge about other people's character and their observable behaviour are only noisy clues to others' true attitudes and intentions. Bayesian inference provides an optimal account of how individuals deal with noisy information to draw maximally accurate inferences about their environment, including their social environment(s). Within the Bayesian inference formulation of social learning, agents are thought to form and maintain beliefs about others by continuously updating internal models about others' unobservable inner states. These beliefs can be thought of, and are mathematically formalised as, probability distributions over these states, with the distribution mean representing the agent's expectation and uncertainty around this expectation being signalled by its *precision* (the inverse of its variance). Agents learn about the world by integrating their prior beliefs with new evidence, whilst the extent to which existing beliefs are updated with new information is determined by the deviation between the agent's predictions and the actual outcome, also called the *prediction error*. Crucially, in contrast to non-Bayesian formulations about belief updating (such as the classical Rescorla-Wagner reinforcement learning model (Rescorla and Wagner, 1972), under the Bayesian framework the precision ascribed to both prior beliefs and incoming evidence determines the degree to which prediction errors drive belief updating. Mathematically, the precision weights ascribed to both prior and current information add up to 1, thus a newly formed (i.e., *posterior*) belief can either be influenced equally by both components (both precision weigths at 0.5) or biased in one or the other direction. For example, if we are very certain (i.e., hold highly precise beliefs) that another person's intention is to act hostile towards

* Corresponding author.
E-mail address: biancaschuster05@gmail.com (B.A. Schuster).

https://doi.org/10.1016/j.pnpbp.2024.111206
Received 22 August 2024; Received in revised form 21 November 2024; Accepted 21 November 2024
Available online 23 November 2024

us, any fluctuations in that person's behaviour, such as an unexpected kind gesture, will likely not affect the representations we hold about this person much. However if we are less certain about their hostile intention, perhaps because we do not know this person very well, this same kind gesture may well cause us to change our mind about the other person's true hostility. How the precision of our existing beliefs evolves across time is described by a quantity termed *belief volatility*, where higher volatility implies lower precision of trial-by-trial belief estimates, and, therefore, high rates of belief shifts in response to new information. In contrast, lower belief volatility reflects more robust (or in extreme cases rigid) beliefs that are less affected by new evidence.

As one biologically plausible algorithmic explanation of how the brain implements Bayesian inference, 'predictive processing' accounts (Clark, 2013) situate belief updating processes at multiple levels of the cortical hierarchy, where predictions from higher cortical levels form priors for inferences from lower-level inputs, and in turn lower-level estimates (e.g., from sensory observations) serve as input for inferences at higher cortical levels. Computationally, this message passing is influenced by the relative uncertainties of beliefs held at each level of the hierarchy. Specifically, *perceptual uncertainty*, the noisiness of the observation itself, sits at the lowest level of the hierarchy (e.g., we see our friend pulling up the corners of their mouth – suggestive of their happy internal state (Wegrzyn et al., 2017) – but we are not wearing our glasses, which introduces noisiness to our sensory observation of their facial expression). At higher levels agents encode *expected uncertainty*, which describes the known stochasticity of cue-outcome relationships within a stable environment (e.g., experience has taught us that when we see our friend they are happy about 80 % of the time), and *unexpected uncertainty*, which relates to switches in the probabilistic relationships of the environment, challenging predictions made based on expected uncertainty (e.g., our friend has recently lost their job, so the contingency between seeing their face and them truly being happy has changed) (Yu and Dayan, 2005).

2. Beyond reward signalling: the role of dopamine in belief updating

Growing evidence suggests that belief updating processes are modulated by various neurochemical processes, and the neuromodulator dopamine is thought to play a particularly pertinent role within the domain of precision and belief volatility. Until recently, dopamine has been predominantly discussed in the context of reinforcement learning, where the stimulus driven, *phasic* firing of dopamine neurons (short bursts of action potentials at more than 10 Hz, synchronised across multiple sites (Liu et al., 2021)) in the human midbrain has been found to track the magnitude of prediction errors relating to rewarding outcomes (Schultz et al., 1997; Schultz, 1998; Schultz, 2007), while sustained (*tonic*) dopamine activity (at 0.2–10 Hz) has been suggested to represent the precision associated with these prediction errors (Fiorillo et al., 2003; Friston et al., 2012; Preuschoff et al., 2006). In other words, increasing evidence suggests that dopamine mediates multiple aspects of belief updating at different timescales (Schultz, 2007; Preuschoff et al., 2006). More recently, research has accumulated which implicates a more general role for dopamine in learning that goes beyond merely signalling reward. This evidence integrates well into the Bayesian framework, wherein optimal behaviour is achieved through minimising (informational) surprise (i.e., unsigned prediction errors), rather than simply maximising reward (Diederen and Fletcher, 2020; Friston et al., 2013). Specifically, Bayesian accounts of the role of dopamine in learning and inference posit that dopamine activity tracks the precision of sensory predictions irrespective of reward (Fiorillo et al., 2003; Friston et al., 2012; Diederen and Fletcher, 2020; Adams et al., 2013; Haarsma et al., 2020; Cassidy et al., 2018; de Lafuente and Romo, 2011). It has been proposed that dopamine does this by modulating the synaptic gain (post-synaptic responsiveness) of neurons propagating prediction error signals, where increased post-synaptic gain is associated with increased precision on prediction errors, at the expense of precision attributed to prior beliefs (Adams et al., 2013). That is, it is suggested that dopamine mediates the neuronal excitability of prediction error units, determining the degree to which prediction errors are propagated further up the hierarchy, and thus how likely they are to elicit belief updates.

In summary, theoretical accounts based on synthetic data and limited empirical evidence suggest that dopamine modulates belief updating via signalling the uncertainty associated with new information. However, with little existing direct experimental evidence, our understanding of the precise mechanistic relationship between dopamine, uncertainty, and prediction error is still evolving. In this scoping selective review, we therefore provide a snapshot of the existing evidence on how dopamine modulates the formation and maintenance of beliefs, with a focus on beliefs about others' trustworthiness. Recent advances in theoretical understanding, and evidence from neuroimaging, pharmacological and clinical studies are reviewed and areas in need for future research identified.

2.1. Dopamine and the formation and maintenance of trust beliefs

Seminal functional magnetic resonance imaging (fMRI) studies showed that only prediction errors that are followed by shifts in beliefs (formulated as Bayesian surprise, i.e., the shift from prior to posterior belief) were encoded in dopamine-rich regions of the midbrain, such as the substantia nigra (SN) and ventral tegmental area (VTA) (Nour et al., 2018), while this was not the case for prediction errors that did not elicit belief shifts (characterised as purely information-theoretic, but not meaningful surprise). This finding was corroborated by a positron emission tomography (PET) study showing a negative association between the neural encoding of belief updates and dopamine release capacity following dopamine agonist administration in the striatum, highlighting a key role for dopamine in belief updating processes beyond reward learning and salience accounts (Nour et al., 2018). Specifically, amphetamine-induced dopamine release was interpreted as a proxy for increased spontaneous dopamine firing in the drug-free state, which should lead to a reduction in the signal-to-noise ratio of stimulus-dependent phasic dopamine bursts, and thus to decreased precision of prediction errors. In essence, these and other (Jeong et al., 2022) results illustrate that rather than tracking reward or surprising events, striatal dopamine may represent a teaching signal which explicitly highlights information that is relevant for learning, causing us to update our internal models of the world. Moreover, the latter study reported a relationship between striatal D2/D3 receptor availability and sensitivity to meaningful information in the form of an inverted-U function, suggesting that both relatively decreased and increased D2/D3 receptor function may be detrimental to information processing and learnig. Chronically or temporarily elevated levels of striatal dopamine may thus lead to an agent failing to filter out meaningful new information from noise, whereas depressed dopamine levels may result in a failure to recognise when new information warrants a belief update. This idea aligns with well-established evidence (Cools and D'Esposito, 2011) that highlights an inverted-U relationship between striatal dopamine and cognitive flexibility, where optimal levels of dopamine promote an adaptive balance of updating vs stabilising information.

This view is further supported by two recent psychopharmacological studies investigating beliefs related to the trustworthiness of others, which illustrate that dopamine *antagonism* results in increased belief volatility, and consequently in increased updating in response to novel information (Mikus et al., 2023a; Barnby et al., 2024). A classical paradigm used to investigate trust beliefs is the multi-player economic trust (or investment) game (Berg et al., 1995), where a participant is required to decide how much of an initial monetary endowment they will share with a trustee (see Fig. 1C). Any money is then multiplied before it is passed on to the trustee, who can either keep the entire sum or pass a portion of it back to the participant. Mikus and colleagues

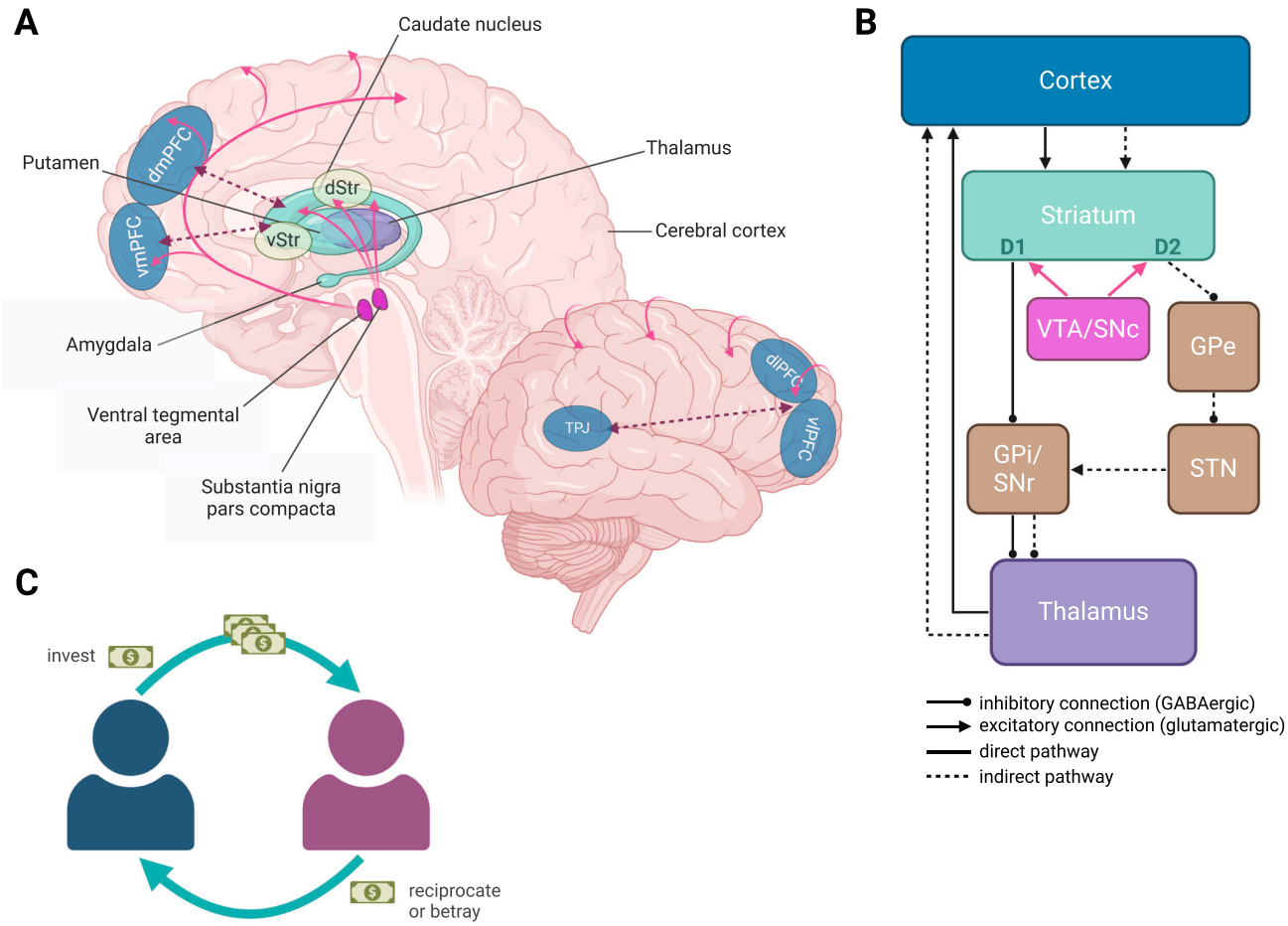


Fig. 1. A) Schematic display of nigrostriatal and mesocortical dopamine pathways (indicated by solid pink arrows) relevant for trust learning. The nigrostriatal pathway involves dopaminergic projections from the substantia nigra pars compacta (SNc) to the dorsal striatum (dStr); the mesocortical pathway involves projections from the ventral tegmental area (VTA) to cortical regions, including the ventromedial- and dorsomedial prefrontal cortices (vmPFC, dmPFC). Dashed purple arrows represent direct and indirect connections between cortex and striatum, and between cortical regions such as the prefrontal cortex and the temporo-parietal junction (TPJ). **B)** Fronto-striatal communication depends on two circuits, described by the direct and indirect pathway model of basal ganglia, and involving D1- and D2 receptor expressing neurons, respectively. Elevated dopamine increases the signal-to-noise ratio in the direct pathway via D1-, and decreases activity in the indirect pathway via D2 receptors (for a review see Frank et al (Frank, 2005).). GPe/i = globus pallidus external/internal; STN = subthalamic nucleus; SNr = substantia nigra pars reticulata. **C)** Schematic depiction of the trust game. A participant receives an initial sum of money and can then decide to share some proportion of it with a trustee. The shared amount is tripled before it reaches the trustee, who can choose to either keep all the money (betray) or return some of it to the participant (reciprocate). Figure created with BioRender (https://www.biorender.com/). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Mikus et al., 2023a) showed that dopamine D2/D3 antagonism using sulpiride increased healhy participants' belief volatility by reducing the precision of prior beliefs and, in turn, increasing precision-weights on prediction errors, allowing individuals to more flexibly adjust their beliefs about a trustee's trustworthiness. This is in line with Barnby et al.'s findings (Barnby et al., 2024) where blocking D2/D3 dopamine receptors using haloperidol resulted in a similar overall increase in belief volatility. Intriguingly, the latter study also showed that haloperidol reduced the precision of beliefs that an agent's actions were driven by harmful intent towards the participant, while having no effects on participants' beliefs that the agent acted out of self interest. This finding was interpreted as the dopamine antagonist haloperidol reducing the perceived relevance of others' actions for oneself, with decisions about whether others' actions are directed towards oneself, or not, being an easily accessible heuristic helping us to quickly adjust our beliefs about others in situations of high social uncertainty.

Importantly, the latter results are in agreement with other evidence suggesting not all beliefs are updated equally: Higher-order beliefs about statistical associations between character traits and volatility of others' behaviour (i.e., stability of a certain trait or behavioural disposition across time) influence belief volatility on lower levels of the hierarchy. For example, it has been shown that people more readily revise their beliefs about others when they are initially deemed to be morally bad, with this updating asymmetry potentially reflecting an evolutionary adaptive mechanism that cautiously favours social cooperation (Siegel et al., 2018) (though note that this is in contrast to earlier accounts which posit that impressions about unfavourable traits are easier to form and harder to lose (Rothbart and Park, 1986)). Further in line with the Bayesian perspective, reducing uncertainty through access to prior social information about interaction partners affects participants' belief volatility. Specifically, in a study employing the trust game, prior information regarding the moral 'goodness' of an interaction partner, relative to neutral information, led to a dampening of the neural response in the caudate nucleus to the outcome valence (positive vs. negative feedback) (Delgado et al., 2005). In other words, when there was less uncertainty regarding the trustworthiness of a partner due to task-relevant prior information, neural activity in reward-related structures was less attuned to changes in the outcome value, signalling

a lower tendency to learn from feedback. Findings from another study suggest that this reduced updating in the existence of prior information even occurs when prior knowledge is contradicted by novel experiences (when a previously cooperative partner violates trust), and may arise from increased top-down influence on the caudate exerted by prefrontal regions (Fouragnan et al., 2013). Notably, the relatively high uncertainty inherent to social interactions may lead us to strongly rely on prior information about others, and thus social priors may exert a uniquely strong top-down influence on the dynamics of social learning processes: Studies investigating both social learning unrelated to trust (Devaine et al., 2014) and those examining trust beliefs (Lamba et al., 2020) suggest that humans are particularly attuned to social information and that social priors, incorporating models about the self and others, may sit at the highest level of the cognitive hierarchy (Frith and Frith, 2023; FeldmanHall and Shenhav, 2019). Put differently, in order to successfully navigate the social world, we have to be able to rely on the conviction that others' behavioural dispositions are relatively stable across time, and these higher-level beliefs strongly affect how we interpret the individual actions of people around us.

While the reviewed research indicates that dopamine regulates prediction error processing at lower hierarchical levels, different neuromodulatory systems may be responsible for tracking belief updates at higher levels. In an fMRI study, participants who learned about the winning probability of two cues from an adviser with fluctuating intentions (to help vs. to mislead) represented lower-level prediction errors about the validity of the adviser's clues in the dopaminergic midbrain and theory of mind-related cortical regions receiving dopamine projections, such as the temporo-parietal junction (TPJ) and the dorsomedial prefrontal cortex (dmPFC) (Diaconescu et al., 2017). In contrast, higher-level prediction errors about the fluctuation of advisers' intentions were associated with the modulation of neural responses in the cholinergic basal forebrain and the dorsal and middle anterior cingulate cortex (ACC), converging with other evidence (Iglesias et al., 2013) to suggest a role for the neurotransmitter acetylcholine in signalling (social) uncertainty at higher levels.

2.2. Aberrant dopamine and alterations in (trust) belief updating: the case of paranoia

Optimal functioning requires maintaining a delicate balance of belief volatility, where too much or too little precision on the prediction error is proposed to underly multiple psychopathologies. Computational psychiatry is a nascent field that has reframed aetiologies for conditions such as psychosis, depression, anxiety, and many more as aberrant belief updating processes (Huys et al., 2016; Adams et al., 2016). Paranoia, or persecutory delusions (PD), is perhaps the clinical phenomenon most intuitively relevant to trust beliefs. Characterised by persistent unfounded beliefs that others (single agents or entities) intend to harm oneself, PD are one of the core features of schizophrenia spectrum disorders (American Psychiatric, 2013) and delusional disorders (American Psychiatric, 2013), but also occur in multiple other conditions including depression (Freeman, 2007) and Parkinson's disease (Warren et al., 2018), and to a lower extent in the general population (Freeman, 2007). Evidence indicates a key role for altered presynaptic dopamine in the pathophysiology of PD in patients (McCutcheon et al., 2018). While the literature is yet far from providing a complete picture of dopaminergic alterations in schizophrenia and PD, clinical studies converge to show elevated dopamine synthesis capacity (Howes et al., 2012; Cheng et al., 2020) and increased dopamine release (Howes et al., 2015) in the striatum of patients, with more recent data highlighting a specific role for the dorsal striatum, a structure linked to associative learning and habit formation which receives projections from frontal areas (McCutcheon et al., 2019). For instance, dopaminergic dysfunction in the caudate nucleus of the dorsal striatum of patients with psychosis has been closely linked to paranoia scores and reduced baseline trust (Gromann et al., 2013). In particular, in the latter study all patients initially showed a reduced tendency to invest in a partner compared to controls. Upon receiving cooperative repayments by the partner, patients with higher persecutory ideation showed reduced signal change in the caudate, indicating a dampened neural response to actions which should have elicited updates to their initially pessimistic trust beliefs. While this finding may have resulted from elevated baseline dopamine levels in patients relative to controls, the latter study did not test this. Elevated dopaminergic activity in the striatum in schizophrenia has been attributed to disinhibition of glutamatergic modulation from the frontal cortex (Howes et al., 2024; Weinberger, 2022). Similarly, fronto-striatal abnormalities have been linked to elevated paranoia in healthy controls (Corlett and Fletcher, 2012; Sabaroedin et al., 2019), although the full underlying mechanisms and origins of these cortico-striatal alterations are not yet fully established.

Computationally, although PD have been associated with a dysregulation of dopamine-mediated prediction error signalling, existing data is currently inconclusive with respect to where, and at what timepoint in the belief-forming process things go awry (Howes et al., 2020; Nassar et al., 2021): Some studies propose pathologically high precision of low-level (e.g., sensory) prediction errors, causing patients to attribute high salience to irrelevant internal or external information and consequently to update their beliefs when they should not, thus resulting in irrational explanations for their unusual sensory experiences (Feeney et al., 2017; Kapur, 2003; Fromm et al., 2023). Others have argued that enhanced precision of prior beliefs (Teufel et al., 2015; Schmack et al., 2013; Baker et al., 2019), or heightened sensitivity to changes in environmental contingencies (Kaplan et al., 2016; Mikus et al., 2023b; Reed et al., 2020; Hauke et al., 2024) (unexpected uncertainty) may underlie delusional symptoms. It is important to note that in PD and other delusions, different computational alterations may be present at multiple levels of the processing hierarchy, which may reconcile the apparent discrepancy of increased prior belief precision on the one, and precision attributed to the prediction error on the other hand (Petrovic and Sterzer, 2023). More precisely, it has been suggested that in delusions, increased precision may be assigned to higher-level priors as a compensatory mechanism to 'explain away' excessively high sensory prediction errors, which in turn are associated with reduced precision of lower-level (perceptual) priors (Adams et al., 2013; Petrovic and Sterzer, 2023; Stuke et al., 2019). Accordingly, increased sensory precision may result in patients registering a relatively greater proportion of others' actions as behaviourally relevant and they in turn make sense of this experience by forming a higher level belief that can incorporate a large variety of observed actions (e.g., rather than integrating a neighbour's hand waving with the prior belief that this kind of greeting among neighbours is common and as such expected, a person suffering from PD may lack this contextual integration and instead form the belief that the hand waving signals the neighbour's malintent). Analogously, the higher-level belief that one's environment is excessively volatile may be a direct result from consistently high sensory precision (Hauke et al., 2024). Once formed, the high ambiguity inherent to social contexts may provide perfect conditions for these compensatory beliefs to become increasingly rigid (Ashinoff et al., 2022; Harding et al., 2024), and the social nature of persecutory delusions (i.e., models relating others' actions to oneself) (Bell et al., 2020) may additionally render them impenetrable to contradictory new information. This is supported by the observation that social priors are hard to overthrow by disproving evidence even in healthy populations (Fouragnan et al., 2013), which also suggests that delusional ideation may lie on a continuum, where at the extreme end pathologically rigid social beliefs serve to alleviate pathologically heightened uncertainty at lower levels. On a neurobiological level, this increased lower-level uncertainty may be reflected by relatively increased dopamine transmission in the striatum and initially lead to hyperconnectivity between the striatum and cortical regions representing social beliefs (such as the TPJ and dmPFC (Diaconescu et al., 2017)), signalling the need to infer the causes of current information by either forming new beliefs or new causal associations between novel

information and existing beliefs (Nour et al., 2018; Diaconescu et al., 2019; Pasupathy and Miller, 2005). At later stages beliefs may become fixed and learning biased towards belief-consistent information through stronger top-down modulation from prefrontal areas (Fouragnan et al., 2013) (though other routes are possible, see (Diaconescu et al., 2019); for an overview of the proposed cortico-striatal circuitry involved see Fig.1 A & B).

In sum, clinical evidence implicates striatal presynaptic dopamine and faulty precision weighting in aberrant belief updating. However, we currently do not have a complete and consistent picture of the specific neuro-computational processes involved. Besides confounding variability in medication state and disease stage across patients, gaining insights about belief updating processes from clinical studies is complicated by the fact that (persecutory) delusions are primarily studied in the context of psychosis and therefore seldom examined in isolation. Thus, we cannot be sure whether differences in behaviour are specific to characteristics of delusions, co-occurring symptoms such as hallucinations (which may relate to entirely distinct, partially shared, or fully shared underlying pathophysiology) or more global neurocognitive deficits associated with psychosis (Ashinoff et al., 2022). This and other, for instance, methodological, inconsistencies mean that at present there is no convergence across studies on where in the processing hierarchy which kinds of alterations of belief updating and/or -maintenance occur (Katthagen et al., 2022). Hierarchical accounts offer a promising path to the computational phenotyping of transdiagnostic phenomena, with prior work showing distinct hierarchical alterations in delusions and hallucinations (Wengler et al., 2020). Yet, only a small number of studies has explicitly modelled belief updating processes in delusions within the same hierarchical framework, presenting an important direction for future research.

3. Summary & conclusions

In this brief review we have provided an overview of the current knowledge on how dopamine shapes how we form and maintain trust beliefs. We highlight how recent experimental, clinical, and psychopharmacological work has advanced our knowledge on the neuro-computational bases of belief formation and updating, illustrating a role for the neuromodulator dopamine in processes that can be conceptualized in a Bayesian framework: precision weighted, error-dependent, and hierarchical belief updating. Specifically, the reviewed evidence suggests that both phasic and tonic dopamine signals play distinct roles in regulating the delicate balance of weighting prior information against incoming evidence. Studies involving healthy volunteers highlight the D2/D3 dopamine system as a key factor in regulating belief volatility, as blocking D2/D3 receptors reduces the precision of prior beliefs about others' trustworthiness. This, in turn, allows individuals to flexibly revise those beliefs upon on incoming new evidence from others' actions. Such findings provide novel insights into the potential mechanisms via which antipsychotics, all of which currently target D2/D3 receptors, effectively reduce symptoms in conditions such as PD. (Pre-) clinical work adds to this evidence by highlighting cortico-striatal circuits wherein the modulation of prior- relative to likelihood precision may be regulated. However, many key questions are still unanswered by the current literature, with the following being among the most pressing:

First, while dopamine clearly has a role in belief updating processes, the specifics of this role on a neurochemical and computational level are as of yet unclear. For instance, while recent data (Mikus et al., 2023a; Barnby et al., 2024) suggests that in healthy people, D2/D3 dopamine antagonism leads to increased precision weights on the prediction error and consequently increased blief volatility, there are also findings which contradict this (Haarsma et al., 2020; Jocham et al., 2014; Rybicki et al., 2022). Integrating findings from existing psychopharmacological studies is complicated by well-established (Cools and D'Esposito, 2011; Frank and O'Reilly, 2006) interactions between individual baseline dopamine function and drug responsivity and -dosage, leading to different, and potentially opposing, drug effects in different subgroups of individuals. This is for example illustrated in Mikus et al (Mikus et al., 2023a),. who observed increased belief volatility only in a genetic subgroup of the population that was specifically selected to have higher striatal presynaptic dopamine availability (Taq1a polymorphism). In addition to genetically conferred variations in striatal baseline dopamine activity, individual working memory capacity has been found to interact with effects of dopaminergic drugs (Frank and O'Reilly, 2006; Gibbs and D'Esposito, 2005) and dopamine synthesis capacity (Chen et al., 2023) on learning performance. Future work should carefully control for these potential sources of interindividual variability by modelling either participants' baseline dopamine function (assessed via genetic proxies or directly using PET) and/or behavioural mediators such as working memory capacity, while using dedicated pharmacological designs in combination with computational modelling.

Furthermore, extensive evidence indicates that dopamine likely works in complex interaction with other neuromodulatory systems to regulate belief updating processes. While computational work suggests that dopamine signals the precision associated with incoming information, other types of uncertainty may be tracked by different neuromodulators, such as acetylcholine, serotonin or noradrenaline (Yu and Dayan, 2005; Diederen and Fletcher, 2020; Marshall et al., 2016). Thus, although limited data suggests dopamine antagonism successfully alleviates delusions in a proportion of patients (Muñoz-Negro et al., 2020), other systems, including cholinergic or serotonergic receptors, might present additional suitable targets for the pharmacological treatment of delusions and related disorders (Caton et al., 2020). Future work teasing apart the specific computational roles played by different neuromodulatory systems may offer new avenues for treating specific computational phenotypes of aberrant belief updating rather than groups of disorders or syndromes.

Second, although a growing body of experimental work suggests similar, domain-general dopamine-mediated mechanims underpinning social and non-social belief updatig processes (Reed et al., 2020; Rybicki et al., 2022; Behrens et al., 2008), there is also evidence that suggests individuals employ different strategies (Lamba et al., 2020), use more sophisticated models (Devaine et al., 2014), and recruit specialised neural structures (Stanley, 2016) when social beliefs are involved. Yet, most studies investigating social learning processes lack non-social control conditions, making it difficult to evaluate whether observed computations are specific to the social domain or rather reflect more general associative learning processes. To dissect social and non-social components of belief updating, studies need experimental designs which directly contrast social and non-social conditions, where a particular challenge may be the matching of salience and/or complexity across social and non-social stimuli (Yon and Heyes, 2024).

Third, most paradigms investigating trust beliefs – such as the much-used trust game – rely on monetary incentives to investigate social belief formation and updating, which irrevocably conflates participants' trust choices with reward. It is thus not surprising that structures and neuromodulatory systems predominantly associated with reward processing, such as the striatum and dopamine, are implicated in these studies. Even though some research has shown that explicit social rewards are processed in structures also linked to non-social rewards (e.g., the ventral striatum (Ruff and Fehr, 2014)), often the kind of inferences involved in social interactions are not directly associated with either type of reward. Consistent with this idea, an fMRI study examining social learning while explicitly forgoing rewards found no prediction-error dependent responses in the classical reward-related structures, and instead revealed activations in higher-level areas associated with model-based learning, such as the dorsolateral prefrontal cortex (Stanley, 2016). Further research is needed that advances our understanding of the types of social learning processes that are not incentivised by explicit rewards.

Understanding how humans update their beliefs about whom around them they should trust is not only crucial on an individual level,

affecting a vast array of psychiatric disorders. It also has additional societal implications in the current age of misinformation (as for example, the healthy population may increasingly develop beliefs that their environment is highly volatile, with downstream effects on their epistemic trust updates (Schulz et al., 2023)). While there are still many unanswered questions, cognitive neuroscience has come a far way in enhancing our understanding of how neuromodulators like dopamine regulate how humans form and maintain social beliefs, and has provided rich experimental frameworks (e.g., (Barnby et al., 2024; Diaconescu et al., 2019)) allowing us to further probe how they dynamically evolve. In particular, the computational approach allows for precise, directly testable, mechanistic hypotheses and is a promising tool for discovering subtle differences and changes in belief updating behaviour between and within individuals that might be missed by summary statistics. As a rapidly advancing field, computational psychiatry will bring about highly useful insights and tools that will advance diagnosis, prognosis and the development of novel interventions, benefitting individuals at risk for and affected by disorders of aberrant belief processing.

CRediT authorship contribution statement

Bianca A. Schuster: Writing – review & editing, Writing – original draft, Funding acquisition, Conceptualization. **Claus Lamm:** Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Bianca A. Schuster reports financial support and article publishing charges were provided by Austrian Science Fund. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Nace Mikuš for his valuable insights and helpful comments on the manuscript. This work was supported by the Austrian Science Fund (ESP339, P32686, PAT 1936023).

References

Adams, R., Stephan, K., Brown, H., Frith, C., Friston, K., 2013. The computational anatomy of psychosis. Front. Psychol. 4. https://doi.org/10.3389/fpsyt.2013.00047.
Adams, R.A., Huys, Q.J.M., Roiser, J.P., 2016. Computational psychiatry: towards a mathematically informed understanding of mental illness. J. Neurol. Neurosurg. &amp. Psychiat. 87, 53. https://doi.org/10.1136/jnnp-2015-310737.
American Psychiatric, A., 2013. Diagnostic and Statistical Manual of Mental Disorders: DSM-5™, 5th ed. American Psychiatric Publishing, Inc.
Ashinoff, B.K., Singletary, N.M., Baker, S.C., Horga, G., 2022. Rethinking delusions: a selective review of delusion research through a computational lens. Schizophr. Res. 245, 23–41. https://doi.org/10.1016/j.schres.2021.01.023.
Baker, S.C., Konova, A.B., Daw, N.D., Horga, G., 2019. A distinct inferential mechanism for delusions in schizophrenia. Brain 142, 1797–1812. https://doi.org/10.1093/brain/awz051.
Barnby, J.M., Bell, V., Deeley, Q., Mehta, M.A., Moutoussis, M., 2024. D2/D3 dopamine supports the precision of mental state inferences and self-relevance of joint social outcomes. Nat. Ment. Health. https://doi.org/10.1038/s44220-024-00220-6.
Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., Rushworth, M.F.S., 2008. Associative learning of social value. Nature 456, 245–249. https://doi.org/10.1038/nature07538.
Bell, V., Raihani, N., Wilkinson, S., 2020. Derationalizing delusions. Clin. Psychol. Sci. 9, 24–37. https://doi.org/10.1177/2167702620951553.
Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games Econom. Behav. 10, 122–142. https://doi.org/10.1006/game.1995.1027.
Cassidy, C.M., et al., 2018. A perceptual inference mechanism for hallucinations linked to striatal dopamine. Curr. Biol. 28, 503–514.e504. https://doi.org/10.1016/j.cub.2017.12.059.
Caton, M., Ochoa, E.L.M., Barrantes, F.J., 2020. The role of nicotinic cholinergic neurotransmission in delusional thinking. NPJ Schizophr. 6 (16). https://doi.org/10.1038/s41537-020-0105-9.
Chen, P., et al., 2023. Effect of striatal dopamine on Pavlovian bias. A large [18F]-DOPA PET study. Behav. Neurosci. 137, 184–195. https://doi.org/10.1037/bne0000547.
Cheng, P.W.C., et al., 2020. The role of dopamine dysregulation and evidence for the transdiagnostic nature of elevated dopamine synthesis in psychosis: a positron emission tomography (PET) study comparing schizophrenia, delusional disorder, and other psychotic disorders. Neuropsychopharmacology 45, 1870–1876. https://doi.org/10.1038/s41386-020-0740-x.
Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36, 181–204. https://doi.org/10.1017/S0140525X12000477.
Cools, R., D'Esposito, M., 2011. Inverted-U–shaped dopamine actions on human working memory and cognitive control. Biol. Psychiatry 69, e113–e125. https://doi.org/10.1016/j.biopsych.2011.03.028.
Corlett, P.R., Fletcher, P.C., 2012. The neurobiology of schizotypy: Fronto-striatal prediction error signal correlates with delusion-like beliefs in healthy people. Neuropsychologia 50, 3612–3620. https://doi.org/10.1016/j.neuropsychologia.2012.09.045.
de Lafuente, V., Romo, R., 2011. Dopamine neurons code subjective sensory experience and uncertainty of perceptual decisions. Proc. Natl. Acad. Sci. 108, 19767–19771. https://doi.org/10.1073/pnas.1117636108.
Delgado, M.R., Frank, R.H., Phelps, E.A., 2005. Perceptions of moral character modulate the neural systems of reward during the trust game. Nat. Neurosci. 8, 1611–1618. https://doi.org/10.1038/nn1575.
Devaine, M., Hollard, G., Daunizeau, J., 2014. The social Bayesian brain: does Mentalizing make a difference when we learn? PLoS Comput. Biol. 10, e1003992. https://doi.org/10.1371/journal.pcbi.1003992.
Diaconescu, A.O., et al., 2017. Hierarchical prediction errors in midbrain and septum during social learning. Soc. Cogn. Affect. Neurosci. 12, 618–634. https://doi.org/10.1093/scan/nsw171.
Diaconescu, A.O., Hauke, D.J., Borgwardt, S., 2019. Models of persecutory delusions: a mechanistic insight into the early stages of psychosis. Mol. Psychiatry 24, 1258–1267. https://doi.org/10.1038/s41380-019-0427-z.
Diederen, K.M.J., Fletcher, P.C., 2020. Dopamine, prediction error and beyond. Neuroscientist 27, 30–46. https://doi.org/10.1177/1073858420907591.
Feeney, E.J., Groman, S.M., Taylor, J.R., Corlett, P.R., 2017. Explaining delusions: reducing uncertainty through basic and computational neuroscience. Schizophr. Bull. 43, 263–272. https://doi.org/10.1093/schbul/sbw194.
FeldmanHall, O., Shenhav, A., 2019. Resolving uncertainty in a social world. Nat. Hum. Behav. 3, 426–435. https://doi.org/10.1038/s41562-019-0590-x.
Fiorillo, C.D., Tobler, P.N., Schultz, W., 2003. Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299, 1898–1902. https://doi.org/10.1126/science.1077349.
Fouragnan, E., et al., 2013. Reputational priors magnify striatal responses to violations of trust. J. Neurosci. 33, 3602–3611. https://doi.org/10.1523/jneurosci.3086-12.2013.
Frank, M.J., 2005. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism. J. Cogn. Neurosci. 17, 51–72. https://doi.org/10.1162/0898929052880093.
Frank, M.J., O'Reilly, R.C., 2006. A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. Behav. Neurosci. 120, 497–517. https://doi.org/10.1037/0735-7044.120.3.497.
Freeman, D., 2007. Suspicious minds: the psychology of persecutory delusions. Clin. Psychol. Rev. 27, 425–457. https://doi.org/10.1016/j.cpr.2006.10.004.
Friston, K.J., et al., 2012. Dopamine, affordance and active inference. PLoS Comput. Biol. 8, e1002327. https://doi.org/10.1371/journal.pcbi.1002327.
Friston, K., et al., 2013. The anatomy of choice: active inference and agency. Front. Hum. Neurosci. 7, 598.
Frith, C., Frith, U., 2023. What Makes us Social? MIT Press.
Fromm, S., et al., 2023. Belief updating in subclinical and clinical delusions. Schizophrenia Bull. Open 4. https://doi.org/10.1093/schizbullopen/sgac074 sgac074.
Gibbs, S.E., D'Esposito, M., 2005. Individual capacity differences predict working memory performance and prefrontal activity following dopamine receptor stimulation. Cogn. Affect. Behav. Neurosci. 5, 212–221. https://doi.org/10.3758/cabn.5.2.212.
Gromann, P.M., et al., 2013. Trust versus paranoia: abnormal response to social reward in psychotic illness. Brain 136, 1968–1975. https://doi.org/10.1093/brain/awt076.
Haarsma, J., et al., 2020. Precision weighting of cortical unsigned prediction error signals benefits learning, is mediated by dopamine, and is impaired in psychosis. Mol. Psychiatry. https://doi.org/10.1038/s41380-020-0803-8.
Harding, J.N., et al., 2024. A new predictive coding model for a more comprehensive account of delusions. Lancet Psychiatry 11, 295–302. https://doi.org/10.1016/S2215-0366(23)00411-X.
Hauke, D.J., et al., 2024. Altered perception of environmental volatility during social learning in emerging psychosis. Computati. Psychiatr. https://doi.org/10.5334/cpsy.95.
Howes, O.D., et al., 2012. The nature of dopamine dysfunction in schizophrenia and what this means for treatment: Meta-analysis of imaging studies. Arch. Gen. Psychiatry 69, 776–786. https://doi.org/10.1001/archgenpsychiatry.2012.169.
Howes, O., McCutcheon, R., Stone, J., 2015. Glutamate and dopamine in schizophrenia: an update for the 21st century. J. Psychopharmacol. 29, 97–115. https://doi.org/10.1177/0269881114563634.
Howes, O.D., Hird, E.J., Adams, R.A., Corlett, P.R., McGuire, P., 2020. Aberrant salience, information processing, and dopaminergic signaling in people at clinical high risk for psychosis. Biol. Psychiatry 88, 304–314. https://doi.org/10.1016/j.biopsych.2020.03.012.

Howes, O.D., Bukala, B.R., Beck, K., 2024. Schizophrenia: from neurochemistry to circuits, symptoms and treatments. Nat. Rev. Neurol. 20, 22–35. https://doi.org/10.1038/s41582-023-00904-0.
Huys, Q.J., Maia, T.V., Frank, M.J., 2016. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat. Neurosci. 19, 404–413. https://doi.org/10.1038/nn.4238.
Iglesias, S., et al., 2013. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron 80, 519–530. https://doi.org/10.1016/j.neuron.2013.09.009.
Jeong, H., et al., 2022. Mesolimbic dopamine release conveys causal associations. Science 378, eabq6740. https://doi.org/10.1126/science.abq6740.
Jocham, G., Klein, T.A., Ullsperger, M., 2014. Differential modulation of reinforcement learning by D2 dopamine and NMDA glutamate receptor antagonism. J. Neurosci. 34, 13151–13162. https://doi.org/10.1523/jneurosci.0757-14.2014.
Kaplan, C.M., et al., 2016. Estimating changing contexts in schizophrenia. Brain 139, 2082–2095. https://doi.org/10.1093/brain/aww095.
Kapur, S., 2003. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am. J. Psychiatry 160, 13–23. https://doi.org/10.1176/appi.ajp.160.1.13.
Katthagen, T., Fromm, S., Wieland, L., Schlagenhauf, F., 2022. Models of dynamic belief updating in psychosis—a review across different computational approaches. Front. Psychol. 13. https://doi.org/10.3389/fpsyt.2022.814111.
Lamba, A., Frank, M.J., FeldmanHall, O., 2020. Anxiety impedes adaptive social learning under uncertainty. Psychol. Sci. 31, 592–603. https://doi.org/10.1177/0956797620910993.
Liu, C., Goel, P., Kaeser, P.S., 2021. Spatial and temporal scales of dopamine transmission. Nat. Rev. Neurosci. 22, 345–358. https://doi.org/10.1038/s41583-021-00455-7.
Marshall, L., et al., 2016. Pharmacological fingerprints of contextual uncertainty. PLoS Biol. 14, e1002575. https://doi.org/10.1371/journal.pbio.1002575.
McCutcheon, R., Beck, K., Jauhar, S., Howes, O.D., 2018. Defining the locus of dopaminergic dysfunction in schizophrenia: a Meta-analysis and test of the mesolimbic hypothesis. Schizophr. Bull. 44, 1301–1311. https://doi.org/10.1093/schbul/sbx180.
McCutcheon, R.A., Abi-Dargham, A., Howes, O.D., 2019. Schizophrenia, dopamine and the striatum: from biology to symptoms. Trends Neurosci. 42, 205–220. https://doi.org/10.1016/j.tins.2018.12.004.
Mikus, N., et al., 2023a. Blocking D2/D3 dopamine receptors in male participants increases volatility of beliefs when learning to trust others. Nat. Commun. 14, 4049. https://doi.org/10.1038/s41467-023-39823-5.
Mikus, N., Lamm, C., Mathys, C., 2023b. Computational phenotyping of aberrant belief updating in individuals with schizotypal traits and schizophrenia. medRxiv. https://doi.org/10.1101/2023.11.27.23299069, 2023.2011.2027.23299069.
Muñoz-Negro, J.E., Gómez-Sierra, F.J., Peralta, V., González-Rodríguez, A., Cervilla, J. A., 2020. A systematic review of studies with clinician-rated scales on the pharmacological treatment of delusional disorder. Int. Clin. Psychopharmacol. 35, 129–136. https://doi.org/10.1097/yic.0000000000000306.
Nassar, M.R., Waltz, J.A., Albrecht, M.A., Gold, J.M., Frank, M.J., 2021. All or nothing belief updating in patients with schizophrenia reduces precision and flexibility of beliefs. Brain 144, 1013–1029. https://doi.org/10.1093/brain/awaa453.
Nour, M.M., et al., 2018. Dopaminergic basis for signaling belief updates, but not surprise, and the link to paranoia. Proc. Natl. Acad. Sci. USA 115, E10167–e10176. https://doi.org/10.1073/pnas.1809298115.
Pasupathy, A., Miller, E.K., 2005. Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature 433, 873–876. https://doi.org/10.1038/nature03287.
Petrovic, P., Sterzer, P., 2023. Resolving the delusion paradox. Schizophr. Bull. 49, 1425–1436. https://doi.org/10.1093/schbul/sbad084.
Preuschoff, K., Bossaerts, P., Quartz, S.R., 2006. Neural differentiation of expected reward and risk in human subcortical structures. Neuron 51, 381–390. https://doi.org/10.1016/j.neuron.2006.06.024.
Reed, E.J., et al., 2020. Paranoia as a deficit in non-social belief updating. eLife 9, e56345. https://doi.org/10.7554/eLife.56345.
Rescorla, R., Wagner, A., 1972. A Theory of Pavlovian Conditioning: The Effectiveness of Reinforcement and Non-reinforcement. Classical Conditioning: Current Research and Theory.
Rothbart, M., Park, B., 1986. On the confirmability and disconfirmability of trait concepts. J. Pers. Soc. Psychol. 50, 131.
Ruff, C.C., Fehr, E., 2014. The neurobiology of rewards and values in social decision making. Nat. Rev. Neurosci. 15, 549–562. https://doi.org/10.1038/nrn3776.
Rybicki, A.J., Sowden, S.L., Schuster, B., Cook, J.L., 2022. Dopaminergic challenge dissociates learning from primary versus secondary sources of information. eLife 11, e74893. https://doi.org/10.7554/eLife.74893.
Sabaroedin, K., et al., 2019. Functional connectivity of Corticostriatal circuitry and psychosis-like experiences in the general community. Biol. Psychiatry 86, 16–24. https://doi.org/10.1016/j.biopsych.2019.02.013.
Schmack, K., et al., 2013. Delusions and the role of beliefs in perceptual inference. J. Neurosci. 33, 13701–13712. https://doi.org/10.1523/jneurosci.1778-13.2013.
Schultz, W., 1998. Predictive reward signal of dopamine neurons. J. Neurophysiol. 80, 1–27. https://doi.org/10.1152/jn.1998.80.1.1.
Schultz, W., 2007. Multiple dopamine functions at different time courses. Annu. Rev. Neurosci. 30, 259–288. https://doi.org/10.1146/annurev.neuro.28.061604.135722.
Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275, 1593–1599. https://doi.org/10.1126/science.275.5306.1593.
Schulz, L., Schulz, E., Bhui, R., Dayan, P., 2023. Mechanisms of Mistrust: A Bayesian Account of Misinformation Learning.
Siegel, J.Z., Mathys, C., Rutledge, R.B., Crockett, M.J., 2018. Beliefs about bad people are volatile. Nat. Hum. Behav. 2, 750–756. https://doi.org/10.1038/s41562-018-0425-1.
Stanley, D.A., 2016. Getting to know you: general and specific neural computations for learning about people. Soc. Cogn. Affect. Neurosci. 11, 525–536. https://doi.org/10.1093/scan/nsv145.
Stuke, H., Weilnhammer, V.A., Sterzer, P., Schmack, K., 2019. Delusion proneness is linked to a reduced usage of prior beliefs in perceptual decisions. Schizophr. Bull. 45, 80–86. https://doi.org/10.1093/schbul/sbx189.
Teufel, C., et al., 2015. Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proc. Natl. Acad. Sci. 112, 13401–13406. https://doi.org/10.1073/pnas.1503916112.
Warren, N., O'Gorman, C., Hume, Z., Kisely, S., Siskind, D., 2018. Delusions in Parkinson's disease: a systematic review of published cases. Neuropsychol. Rev. 28, 310–316. https://doi.org/10.1007/s11065-018-9379-3.
Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., Kissler, J., 2017. Mapping the emotional face. How individual face parts contribute to successful emotion recognition. PLoS One 12, e0177239. https://doi.org/10.1371/journal.pone.0177239.
Weinberger, D.R., 2022. It's dopamine and schizophrenia all over again. Biol. Psychiatry 92, 757–759. https://doi.org/10.1016/j.biopsych.2022.08.001.
Wengler, K., Goldberg, A.T., Chahine, G., Horga, G., 2020. Distinct hierarchical alterations of intrinsic neural timescales account for different manifestations of psychosis. eLife 9, e56151. https://doi.org/10.7554/eLife.56151.
Yon, D., Heyes, C., 2024. Social learning in models and minds. Synthese 203, 199. https://doi.org/10.1007/s11229-024-04632-w.
Yu, A.J., Dayan, P., 2005. Uncertainty, neuromodulation, and attention. Neuron 46, 681–692. https://doi.org/10.1016/j.neuron.2005.04.026.