# Why Do We Believe Things That Hurt Us?
# A predictive processing account of conspiracy theories and cognitive pathologies

Authored by: Alice Ferguson-O'Brien
Supervised by: Dr Mark Miller
September 5th, 2023

**Acknowledgements**

This research was conducted through the Laidlaw Scholars Program. It was through the generosity and robustness of the Laidlaw Program that I was able to travel, and meet with a number of the authors cited in this paper. I would like to thank Dr. Kryztof Dolega, Dr. Ines Hipolito, and Dr. Nina Poth for taking the time to meet with me, discuss their work, and help encourage and inspire me through my research. I would like to extend gratitude to everyone involved in the Workshops on Emergence at the Centre for Cognitive Science at the University of Sussex for allowing me to attend, and specifically to Ben White for welcoming me, sharing his knowledge, and encouraging my research. Finally, this work would have been impossible without the unwavering support of my supervisor Dr Mark Miller, who has contributed so much to this research since the beginning.

**Abstract**

Humans form beliefs in order to stay alive. We depend on beliefs to inform actions about what is safe, or optimal in an environment. Yet, we are still prone to believing things that are not only wrong, but harmful to ourselves and our communities. Survival is complex - beliefs are crucial to fulfilling basic physiological needs (maintaining a stable body temperature, blood acidity, etc), but there are also much more complex social, psychological, and existential needs. Finding a balance between belief accuracy, and belief utility is crucial to living a happy, healthy, and connected life. The predictive processing framework offers an explanation for how optimal agents may adopt suboptimal beliefs. At the core of this explanation is misinformation, which can lead predictive systems to form warped beliefs, and adapt warped updating strategies. Contemporary technologies, like social media, enable the sharing of misinformation about news, history, social dynamics, and the self. This paper investigates conspiracy theory belief, and the beliefs that are associated with cognitive pathologies under the predictive processing framework, unveiling similarities between different bad belief systems, and shedding light on things that may make people less susceptible to these bad belief systems.

## Introduction

We are surrounded, and threatened, by huge amounts of uncertainty in any given environment. Being able to reduce uncertainty by accurately predicting our internal and external environments is critical to surviving, and to living happy and healthy lives. It has recently been proposed by the predictive processing framework (PPF) that humans are able to effectively reduce uncertainty by generating internal models of the world that are regularly updated in the face of new, reliable information (Clark, 2013). By generating these optimal predictive systems, we are able to excel in chaotic environments, processing salient factors and changes, and ignoring the majority of the noise that surrounds us.

At their best, these predictive models are able to help us predict our environments, and tune themselves given evidence that contradicts the model - constantly adjusting to our ever changing environments. However, when exposed to misleading or incorrect information, predictive models are susceptible to maladaptive updating, and can lead to suboptimal outcomes such as maladaptive beliefs about our environment, or the development of suboptimal learning patterns - some of which can become pathological. These suboptimal results can also lead to non-pathological outcomes, such as conspiracy theory beliefs.

These bad belief systems are often dismissed as resulting from a lack of rationality or some problem with the cognition itself, but research across various fields has suggested that there are many rational, and even adaptive reasons to adopt a bad-belief system, specifically in moments of uncertainty (Poth & Dolega, 2023; Sato, 2023; Bortollotti, 2020; Williams, 2020). Misleading evidence, either presented to us overtly (in cases of fake news), subtly (in the case of social media), or simply misinterpreted information radically alters our belief system. There is an imperative to reduce error between the expected and experienced world (Clark, 2013; Miller et al., 2022). If our experienced world is riddled with misinformation, our expected world will adapt itself in order to better predict the misinformation. In order to regularly accommodate misinformation, systems must alter their updating practices, which often lead to the adoption of sub-optimal belief systems.

This paper will argue that conspiracy theory belief and cognitive pathologies may be partially the result of some shared cognitive tendencies within PPF, explaining some similarities between the two phenomenon, their large rates of comorbidity, and also offering some insights into how we can become less susceptible to both cognitive pathologies, and conspiracy theory belief. I begin by offering a brief explanation of PPF. By applying these principles to conspiracy theories, I then discuss some of the ways that conspiracy theory belief may be rooted in rational, adaptive strategies. I continue to unpack the PPF accounts of two

different cognitive pathologies, revealing two different ways that optimal belief systems can lead to bad beliefs, before applying this diametric model of bad-belief systems to conspiracy theory belief, highlighting possible similarities. Finally, I suggest theoretical and practical solutions for individuals and communities to lessen susceptibility to bad-belief systems.

This paper is not aiming to equate conspiracy theory belief with pathologies, or to suggest that those who subscribe to conspiracy theories are mentally ill. Rather, it aims to highlight that even optimal predictive brains can lead to suboptimal outcomes like cognitive pathologies and conspiracy theory belief in the face of bad evidence or maladaptive updating mechanisms, and to highlight strategies that may help individuals and communities become less susceptible to maladaptive beliefs.

## Predictive Processing, Error Dynamics, and Misinformation

Having a good predictive grip of our environments helps us act in ways that promote optimal states. According to PPF, central to having a good predictive grip is having an optimal predictive model that represents the world in ways that enable optimal action (Clark, 2013; Miller et al., 2022).

PPF argues that, as opposed to traditional models wherein the brain is a passive system that simply takes information in and processes it, the brain is an active system that generates predictions about how you will interact with the world (Clark, 2016; Metzinger & Wiese, 2017). These predictions are based on predictive models which synthesize our interactions with the world, and actually comprise most of our perceptual experiences (Nave et al., 2020, 3). The priors that form these predictive models are complex, and are most likely generated through a combination of past experiences, as well as genetic, evolutionary, and developmental factors (Gallagher et al., 2021). These models are organized hierarchically throughout cortical areas based on the temporal and spatial scale of the things they predict, and each predictive model is actively being tuned by the predictive models above and below it (Metzinger & Wiese, 2017). These models shape our experience of the world by tuning out irrelevant, predictable sensory information, saving considerable energy by focusing on important details, and by actively altering our sensory information to help it conform to our models.

According to PPF, the raw, noisy, and overwhelming sensory data collected by our eyes, ears, mouth, nose, and skin largely serve as a ‘proof-reading’ mechanism, setting off alarm bells when the prediction generated by the model is inconsistent with incoming sensory information (Clark, 2016; Friston, 2013; Miller et al., 2022; Nave et al., 2020). These alarm bells are known as prediction errors. The ‘weight’ or precision of a prediction error is adjusted depending on how reliable the sensory input upon which they

are based is deemed (Clark, 2016, 41; Andersen, 2022, 1656).When a specific predictive model is successful, it is more reliable and precise, and thus weighted more heavily than a model which is performing poorly (Metzinger & Wiese, 2017). Prediction errors that arise in noisy environments, with higher levels of ambiguity are also rated to be less precise - for instance a visual prediction error in a dark room would be less precise than one in a well lit room (Clark, 2016; Miller et al, 2022). A highly precise prediction error that arises in an imprecise model is thus weighted more than a prediction error that arises in a model that is performing well.

There are two strategies for agents to reduce prediction error - either by changing their predictive model in order to better reflect their sensory input, or through active inference, wherein the agent alters the available input in order to bring about the predicted sensory states (Andersen, 2022; Baddock et al., 2017; Clark, 2016; Hohwy, 2017). For most people this is relatively easy. For instance, if one lives in a home where they keep butter in the fridge, and then looks for butter at a friend's house, it would make sense, according to their predictive model, that they should first check the fridge. If there is no butter in the fridge, perhaps because the friend keeps their butter in a dish by the toaster, the visitor would experience increased prediction error about where butter is kept. They could then either update their beliefs about where butter is kept in order to accommodate the fact that butter can be kept in numerous places, or they could act on their environment and move the butter to the fridge in order to decrease prediction error.

It is maladaptive for an agent to update their generative model every time there is prediction error, as it would result in overfitting (Andersen, 2022). A model that is too specific is less effective at making generalizations that save cognitive energy, and help us interpret the most important details of our environment efficiently. It is equally maladaptive for an agent to update their model only in the face of very significant error, as the model would become undefined (Andersen, 2022). An under-fitted model that allows too much generalization fails to distinguish crucial changes and nuances in our environments.

To evaluate whether or not a predictive model should update in the face of prediction error, it does precision weighing. Precision weighing gauges the reliability of the prediction error as opposed to the generative model. If there is higher precision, or reliability, given to the prediction error, then the model updates, but if there is more precision given to the generative model, the input is altered to make it conform to the predictive model. Key then, to an optimal predictive model is having a dynamically adaptive model, which is able to do precision weighing, and update beliefs systems and alter perception based on accurate judgements of precision (Andersen, 2022). The implications of precision weighing become extremely relevant when our predictive model is being fed misleading, or incorrect information.

If we interpret misinformation to be true or from a reliable source, it may be given high precision, and our model may update. If exposed to further misinformation, our model will be predicting it well, leading the misinformed model to be rated as highly precise.

Take an optimal predictive agent, with a predictive model that is able to adapt its beliefs according to sufficient evidence information, and can in turn actively alter their perceptions so that they more accurately fit the generative model. If this predictive agent is exposed to misinformation, for example they are receiving misleading information about what people look like by regularly consuming edited photos and videos of people on social media, their belief about what a normal face or body looks like would, and should, begin to shift (Miller & White, 2021). This is how a predictive agent ought to react to new contradictory information. The predictive agent, with their new beliefs about what normal faces and bodies look like, may then experience increased prediction error when looking at their own face or body, as they have come to predict faces and bodies to look like the enhanced faces and bodies online (Miller & White, 2021). It would then be expected that the predictive agent may take steps to reduce this prediction error by making themselves look more like the faces and bodies they are expecting, perhaps by wearing makeup. Wearing makeup may help reduce prediction error when they see themselves in the mirror with makeup on, and it would further confirm their belief about what people look like, making the belief stronger, and in turn leading to even higher prediction error when they see themselves without their makeup (Miller & White, 2021).

This process is an example of an optimal predictive agent who has been exposed to misleading information, even though it is easy to imagine how this type of belief installation could lead to pathological outcomes, such as eating disorders. It plausibly follows that many pathological, or otherwise maladaptive beliefs do not result from a malfunctioning system, but are actually the result of a functional, and optimal system processing over bad information that then alters the predictive dynamics of the brain (Poth & Dolega, 2023). It is not necessarily a sick or irrational brain that leads to the installation of bad beliefs, and is perhaps an adaptive brain that is updating its beliefs in the face of ‘bad’ evidence.

Research further suggests that reducing prediction error is not only important to survival, but also feels good (Miller et al., 2022). Evidence suggests that an unexpected reduction of prediction error leads to feelings of ‘hedonic pleasure’ while unexpected increases in prediction error lead to feelings of stress, anxiety, and discomfort (Miller et al., 2022). In predictive processing, precision weighing is represented through dopaminergic discharges that increase dopamine levels when our predictive models are doing unexpectedly well (Adams et al., 2013; Fletcher & Firth, 2008; Miller et al., 2022). When one of our

models succeeds at anticipating the world, it gains precision, and makes it more powerful. If on the other hand a model that usually does well begins to fail, it feels bad, and renders the model less precise (Miller et al., 2020). As well as incentivizing belief systems that minimize uncertainty, this also makes it difficult to update belief systems that we do feel are successful in reducing error in any specific domain, even if that means increasing overall uncertainty (Miller et al., 2020).

For agents that seek to minimize uncertainty in their environments, moments of heightened uncertainty offer unique opportunities. They are often stressful, and can lead to predictive agents feeling out of control as prediction error rises, but they also offer a chance to update generative models in order to make them more robust and resilient (Clark, 2017; Kiverstein et al., 2019). This is what drives humans to be curious and brave, to try new and sometimes risky things in hopes that it may offer new information to the generative model (Sato, 2023). There is a critical balance between exploration and exploitation in predictive processing - comparing the epistemic value of learning new things with the utility value of doing something known to reliably decrease uncertainty. However, increased uncertainty, driven by both internal and external factors, can lead to suboptimal belief formation if the generative model misinterprets the novel information.

## Belief Formation and Conspiracy Theory Belief

PPF accounts of conspiracy theory belief have demonstrated that conspiracy theory belief formation is functionally quite similar to typical belief formation (Dentith, 2016; Poth & Dolega, 2023). People make use of prior beliefs to judge incoming novel evidence. If the novel evidence is more probable, or precise than the top-down belief from the model, the model is altered to accommodate the evidence, or the evidence is altered to match the model.

There have been many attempts to define conspiracy theory, and conspiracy theory belief. For the sake of this paper, a neutral definition will be utilized, as many definitions surrounding conspiracy theories include their irrationality, or epistemic failings in the very definition (Dentith, 2016). I will borrow a definition from Dentith (2016) who states that a conspiracy theory is “an explanation for an event that cites a conspiracy as a salient cause of said event,” (Dentith, 2016, 2).

There are many reasons that people seem to subscribe to conspiracy theories, most of them rational, and even evolutionarily optimal (Dentith, 2016). People may turn to conspiracy theories for epistemic reasons (Douglas et al., 2017). Conspiracy theories often give a causal explanation for phenomena that we do not understand or that is otherwise unexplainable. They help us understand complex parts of our

environments that are often characterized by high uncertainty such as surprising disasters, political polarization, war, and so on. Even if these explanations are not perfect, they may become a reliable way to reduce uncertainty.

There are also existential reasons to subscribe to conspiracy theories, especially theories surrounding otherwise complex, or unexplainable phenomena (Douglas et al., 2017). Conspiracy theories can offer a sense of control over phenomena that are otherwise presented as uncontrollable - like natural disasters, or a pandemic (Douglas et al., 2017). Conspiracy theories can offer hope in otherwise hopeless, and anxiety inducing situations (Bortolotti, 2020). Further, rejecting the official account and instead believing in a conspiracy theory grants people a sense of self-importance or worth that may make them feel more empowered,or unique (Douglas et al., 2019; Douglas et al., 2017).

Williams (2021) points out that our beliefs are not simply epistemically motivated. They have impacts on who we surround ourselves with, how we feel about ourselves, and how much we believe our actions in the world matter. There are social reasons to believe conspiracy theories (Douglas et al., 2017). The success of adopting a socially adaptive belief is dependent on the ratio between the benefits gained from adopting a belief and the harm caused by adopting the belief (Williams, 2021). Conspiracy theories offer people a strong sense of community, through online and in person networks of believers (Douglas et al., 2017; Dentith, 2016; Douglas et al., 2019). Conspiracy theory belief also allows people to protect their self-image, either as an individual or part of a distinct social group by shifting blame towards secret conspirators. In contrast, there is often low risk associated with believing conspiracy theories (Bortolotti & Ichino, 2020).

Research surrounding those who believe in conspiracy theories has indicated that loneliness, anomia and disempowerment, perceived lack of belonging, and low intelligence are all associated with conspiracy theory belief (Goreis & Voracek, 2019). These factors are also highly associated with increased uncertainty. Being isolated from our community, unable to change or have any control of things that are happening to us in our political, public, or private lives, or feeling threatened by others with more intelligence or power are all things that lead to increased uncertainty.

It is healthy and adaptive to adopt our beliefs for epistemic, existential, or social reasons. Conspiracy theories offer beliefs that help people understand complex parts of the world, reduce uncertainty, connect with others, maintain positive self-image, and feel in control. For a predictive human agent, these factors are as crucial to sustaining healthy and happy lives (Williams, 2021). Having said that, conspiracy theory

belief is often harmful to individuals and communities, and at some point they do become maladaptive. This shift from adaptive to maladaptive plausibly occurs when belief in the conspiracy theory becomes too rigid, leading to a poorly updated model, or when belief in the conspiracy theory demands radical updating of other beliefs in order to accommodate it.

## PPF and Cognitive Pathologies

Models of pathology within PPF tend to be defined not as malfunctioning brains, but rather optimal brains processing over various suboptimal broken belief systems (Schwartenbeck et al. 2015). People with cognitive pathologies still make predictions based on their generative model, and use precision weighing in order to modify their generative model, but their belief system has been warped by misinformation that alters the way prediction error is handled. In other words, the pathology arises when the predictive model is being poorly updated, and the agent is acting based on a model that reflects the world in unhelpful ways. PPF is a normative model through which different organisms are able to thrive to different extents (Sato, 2023). The inferences people with cognitive pathologies are making may be the best inferences in a bad belief system.

The drastic differences between generative models, including differences that in some cases lead to pathological outcomes, can be explained by a number of factors. Developmental factors seem to be a plausible example of factors that may have a significant impact on the constitution of a generative model (Schwartenbeck et al. 2015), but so do genetic factors (Garland et al., 2010), and environmental factors like chronic stress (Miller et al., 2020; Baddock et al., 2017).

Two cognitive pathologies that represent how an optimal system can present suboptimal results are depression, and psychotic disorders.

### Depression

It has been suggested that depression is related to maladaptive error dynamics within the predictive model (Kube et al., 2020). More specifically, evidence suggests that depression is related to a more rigid belief system that is less able to update in the face of novel evidence (Baddock et al., 2017). Instead of updating the generative model when faced with overwhelming prediction error, people with depression tend to interpret novel information in ways that sustain their prior beliefs about the world (Kube et al., 2020; Fabry, 2019; Baddock et al., 2017).

Someone with depression may begin with a vague negative belief, perhaps something like ‘I am failing at most things in my life,’ ‘very few people like me,’ or ‘there is not much point in life.’ This belief could arise following evidence like getting a bad grade, getting fired, or an argument with a friend. This belief becomes incorporated into their generative model, as it is able to reliably explain some experiences. Once part of the model this belief begins to shape the ways that novel evidence is interpreted. Future evidence that supports this belief increases the precision of the model, while evidence that disconfirms this belief lowers the precision of the model. As explained above, and by Miller et al. (2020), this precision weighing impacts higher levels of our generative model. When people are successfully reducing prediction, higher-level beliefs about the success of the model are positive. This encourages people to hold on to beliefs that help them feel like they have a grip, and encourages people to act in ways that bring about results that confirm their beliefs.

What appears to go wrong in cases of pathology is inflexible rigidness. We all have some belief systems that are rigid due to their high precision, but in cases of depression there appears to be an inflexibly high precision given to generative models, making them unlikely to update. For example, if someone begins to adopt a general belief, like ‘I am failing in my life’ after experiencing a couple failures, they may start to predict, or expect, failure. In non-pathological cases, this belief should be easily disconfirmed by input that disproves failure. Any input that indicates success should raise prediction errors, and the model should update. In cases of depression, people seem to protect their beliefs from updating by altering disconfirming input so that it continues to reinforce their top-down belief (Miller et al., 2020; Kube et al., 2020). There is a reluctance to update any belief that seems to reduce uncertainty, and there is a strong incentive to protect it. This leads people with depression to not only expect negative outcomes, but to alter their environments to bring about negative outcomes (Miller et al., 2020; Kube et al., 2020). This may offer insight into some symptoms of depression like reduced interest in life, rumination about the self and relationship, and an overall negative bias. These symptoms are attempts to make the environment resemble their model (Fabry, 2019). As this core belief is fed more evidence that supports it, the core belief is given a higher precision weighing, making it even stronger, and further limiting the ability of counter-evidence to alter the belief.

## Psychotic Disorders

Psychotic disorders are a class of mental disorders characterized by hallucinations, delusions, disorganized speech and thinking, disturbances to social functioning, including schizophrenia, schizoaffective disorder, brief psychotic disorder, and substance induced psychotic disorder (American Psychiatric Association, 2013). According to PPF, people experiencing psychotic disorders often have

overly malleable belief systems that accommodate inconsistent evidence into their predictive models (Adams et al., 2013; Fletcher & Firth, 2008, Firth & Friston, 2013).

People with psychotic disorders have higher precision weighing of bottom-up sensory signals (Adams et al., 2013; Fletcher & Firth, 2008; Firth & Friston, 2013). This interferes with the agent's ability to filter out irrelevant stimuli (Fletcher & Firth, 2008). When the top-down beliefs fail to accurately predict and explain the world, including all of the stimuli that most people ignore, there is an increase in prediction error that leads the models to update and accommodate the contradictions between the world and the model (Hohwy, 2013). Inferences must be made in the generative model in order to accommodate bizarre phenomena into a belief (Hohwy, 2013). As these lower-level models adapt to incoming sensory information, they become overfitted and less effective at making predictions (Andersen, 2022). This leads to a dependance on the higher-level models, which have to accommodate incoherence at the lower levels.

This model of psychotic disorders can account for many symptoms and experiences associated with psychotic disorders. Hallucinations for instance, which are false perceptions of the world may occur as a result of overfitting higher-order models in order to accommodate prediction errors. Take our predictive models that interpret vision for instance. There are complex evolutionary and developmental models that predict sight. One of the many phenomena these models are able to predict well are movement (Firth & Friston, 2013). Specifically, these models are able to distinguish between whether a visual target is moving, or if our own body or eyes are moving. This is why things that we are looking at appear as still when we shake our heads back and forth. The bottom-up sensory input of shaking your head while looking at a photo should indicate that the photo is moving, but top-down beliefs about what to expect when we shake our heads, and about how unlikely it is for a painting to move inform the prediction that the painting is not moving, so we do not perceive it to move (Firth & Friston, 2013). For someone with a psychotic disorder the sensory input of the painting moving may have a higher precision weighing then the top-down beliefs, which may lead to a hallucination of the painting moving (Firth & Friston, 2013). Their model would update in order to accommodate the belief that paintings move. This would then inform their expectations about paintings, and may cause them to perceive movement in future paintings.

Delusions, which are false beliefs about the world, occur when lower level models, like those that perceive raw sensory data, become incoherent with higher level beliefs. For instance, someone with a psychotic disorder who has adapted a belief that allows them to perceive movement in paintings would then have to answer to a higher level belief about how and why the paintings are moving. In order to reduce uncertainty, it is important that our higher level beliefs can reliably explain our lower level

perceptions (Firth & Friston, 2013). This may lead people to adopt beliefs like ‘there is someone inside the painting,’ or ‘the painting is actually a portal.’ People adopt strange beliefs in the face of bizarre information.

This proposed model of psychotic disorders converges well with empirical evidence surrounding tendencies of people with psychotic disorders. People with psychotic disorders generally do not perform significantly worse on assessments requiring logic based reasoning than controls, but they do perform worse tests of probabilistic reasoning, plausibly indicating that precision weighing - or identifying the probability of a model being able to correctly predict the world - may be significantly impacted in cases of psychotic disorders (Adams et al., 2013). Further, phenomenological reports from people who live with or have lived with psychotic disorders often report enhanced sensory experiences, where colors are brighter, noises louder, and so on (Fletcher & Firth, 2008). This is plausibly the result of the breakdown of low level models that forces people to take in more raw sensory data instead of using reliable models to predict sensory information.

It feels good to be able to reliably reduce prediction error, so good that we are regularly willing to sacrifice getting a true grip on reality and adopt untrue beliefs in order to feel like we are reducing uncertainty. In cases of cognitive pathologies, inaccurate precision weighing can lead people to form and rely on maladaptive beliefs in order to reduce uncertainty, even though these beliefs have pathological outcomes, and often lead to heightened uncertainty in other domains.

## PPF and Conspiracy Theory Belief

The highly rigid and highly malleable learning patterns that are related to depression and psychotic disorders respectively may also be related to conspiracy theory belief. This section will explore the ways that both highly rigid, and highly malleable predictive models can become maladaptive, and demonstrate the ways in which misinformation can exacerbate this process.

Poth and Dolega (2023) and Sato (2023) have both suggested that there are two distinct methods of belief updating that are associated with conspiracy theory belief. While Poth and Dolega refer to them as monological as opposed to self sealing beliefs, and Sato as lower and higher level distrust beliefs, they are both highlighting similar phenomena. I wish to further draw attention to the similarities between these belief patterns, and the cognitive pathologies that I have mentioned above. This is not an effort to argue that conspiracy theory believers are experiencing something pathological, but rather to shed light on the types of belief systems that make us prone to bad belief, while also possibly revealing some of the

possible reasons for the high comorbidity between cognitive pathologies and conspiracy theory belief, and finally by suggesting some possible strategies for helping individuals and communities become less prone to bad belief systems, weather pathological or not.

## Highly Rigid

Many aspects of conspiracy theory belief, including the manner in which the belief system is formed and updated, as well as many of the factors that are associated with conspiracy theory belief, seem to be linked to the rigid belief systems that can lead to cognitive pathologies such as depression.

According to PPF one ought to update their belief only if the probability of contradictory evidence being correct is higher than the probability of the prior belief being wrong. It would be irrational for a belief system to fail to update given new evidence only if that new evidence had a higher probability than the probability of the belief itself (Poth and Dolega, 2022). A rigid belief system does not meet this definition of irrational, as it fails to update because the prior belief is given a very high precision weighing. The probability of the prior belief being correct is very high. This is what happens in depression, making it very difficult for depressed people to change their beliefs about themselves or the world. This is also what appears to happen to many people who subscribe to conspiracy theory belief.

Rigid belief systems make conspiracy theory belief seem completely insulated. An insulated belief is one that is completely immune to evidence, and never updates. Although conspiracy theory beliefs may feel impervious to evidence from an outside perspective, PPF encourages a hypothesis that there is a slow and complex system of belief updating that takes place as conspiracy theory believers interact with evidence that either supports or debunks their beliefs (Poth and Dolega, 2023).

Most of our beliefs do not exist alone, but rather in the complex hierarchy. When there is high uncertainty in one model, we turn to other adjacent, and higher level beliefs in order to reduce the uncertainty. Similarly, if one belief is disproven, it may impact the reliability of other beliefs. Conspiracy theory beliefs are equally intertwined (Poth and Dolega, 2023). The core belief of a conspiracy theory may be something like ‘COVID-19 was a man-made attempt to control the population,’ and is comparable to the core belief of a depressed person which may be ‘I am a failure.’ The core belief may be supported by auxiliary beliefs like ‘the media is not trustworthy.’ The core belief could encounter disconfirming evidence like a reputable report demonstrating that COVID-19 was the result of natural causes. While it may seem that this evidence should alter the probability of the core belief, given the auxiliary beliefs that are held in conjunction with the core belief, it becomes less clear which belief the evidence truly targets

(Poth and Dolega, 2023). Should an article proving that COVID-19 was not man-made really disprove the core belief given the adjacent belief that media is unreliable? Counter evidence may lead people to update their auxiliary hypothesis without changing their core belief significantly (Poth and Dolega, 2023). In fact, a piece of evidence that appears to be disconfirming the core belief may actually reinforce it by confirming an auxiliary hypothesis (Poth and Dolega, 2023). Showing someone who does not trust the media a reputable article about the origins of COVID-19 may interpret that article as further evidence of the media manipulating them, thus strengthening their belief about the media, and thereby reinforcing the core belief about COVID-19 being man-made.

A rigid belief is not inherently irrational (Sato, 2023). For instance, many of us subscribe to the belief that the earth is round despite the fact that most of our first hand experiences of the world, including our visual perception of the horizon, disconfirm the belief that the earth is round. Many of us continue to subscribe to this belief because of a combination of auxiliary beliefs about science, who are trusted sources, and so on. The combined probability of all of these beliefs, which support our core belief that the earth is round, is higher than the probability that the evidence provided by our sensory input is accurate. The higher precision of the top-down belief, or core belief, encourages altering prediction error signals in order to make bottom-up signals conform to the top-down belief. Having said that, it seems that systems that are *inflexibly* rigid, or have rigidity that spans across beliefs in many domains, is maladaptive, as it can lead to harmful outcomes like depression, and conspiracy theory belief.

### Highly Malleable

There are some common aspects of conspiracy theory belief and the belief patterns that seem related to psychotic disorders. Instead of the predictive models being impervious to updating in the face of novel evidence, highly malleable belief systems assign high precision to low-level evidence (like sensory information), forcing the predictive model to constantly adapt to accommodate new information.

Highly malleable belief systems are related to the hierarchical organization of predictive models proposed by PPF. Low-level models receive and predict novel information from the environment in the moment, like sensory information, while high-level models are associated with predictions over more long-term states of our environment (Metzinger & Wiese, 2017). When low-level models are unable to resolve uncertainty, higher-level models decrease the precision of the lower-level models. As low-level models become less and less precise, the dependance on higher-level predictions increases. The higher-level models have to reduce uncertainty across many domains with only the imprecise input of low-level

models (Howhy, 2013). This can lead to ‘gooey’ high-level models that, in an attempt to reduce uncertainty, are often distorted.

For example, raw sensory information may send us a signal about body temperature. If body temperature is lower or higher than predicted, prediction error arises. This sensory data may then inform a handful of beliefs about our safety, our comfort, or the heating system in our house. In order to resolve prediction errors we may act upon our environment by doing things that our models believe will reliably reduce the prediction error, like turning on the heat, or putting on a sweater. If after turning on the heat or putting on the sweater, sensory signals still indicate that you are cold, there is an unexpected increase in prediction error. This precise model that should be a reliable tool to reduce error prediction still failed. Suddenly, the precision of the model, as well as related models is called into question. Higher-level beliefs have to change to explain why low-level beliefs about how to warm yourself failed. Maybe you call a heat contractor, suspecting that there is something wrong with your heating system, but they confirm that nothing is wrong. In order to make sense of the contradictions within the belief system, a higher order belief must accommodate the contradictions. Many things could explain why you are feeling cold; sickness, a failure to acclimate after a tropical holiday. Perhaps it is something more malicious though - maybe the contractor is lying to you, or maybe someone is going around opening windows when you are in the other room. While these beliefs seem far-fetched, and even irrational, they are accommodating all the information and prior beliefs in order to make an inference under the best available evidence.

It is not uncommon for conspiracy theory believers to believe in contradictory conspiracy theories. For instance, people are more likely to believe that Osama Bin Laden was both alive *and* dead when US forces arrived if they believe that there was a cover up (Wood et al., 2012). These inconsistent beliefs are held together by a higher-order belief that makes sense of both, in this case, a cover up. If evidence contradicts their belief, a new belief can be adopted in order to incorporate the evidence into their beliefs.

It has been observed that conspiracy theories often become more entrenched the more that they are challenged (Bortolotti et al., 2021; Poth and Dolega, 2023; Sato, 2023). This may be the result of how a malleable belief system reacts to new evidence. When a healthy model encounters disconfirming evidence, beliefs are updated, or the evidence is ignored. In highly malleable models, the model is able to adapt so that almost any evidence can be reconstructed to support the model. The higher-level beliefs are able to accommodate so that there is very little evidence that can be used to counter them (Poth and Dolega, 2023; Sato, 2023). For instance, showing anti-vaxxer government studies about the importance of vaccines may lead their model to adopt the belief that governments are part of the conspiracy that is trying

to harm us with vaccines. Explaining to an anti-vaxxer that you yourself are vaccinated, and have not experienced any health effects may lead them to adopt the belief that you are brainwashed, or have been replaced by an alien, or anything else. The model updates in the face of new evidence, but the evidence is never disconfirming.

This belief pattern, which is strikingly similar to the belief pattern that is proposedly related to psychotic disorders, seems to be aligned with the formation, maintenance, and updating of conspiracy theory believers. In moments of increased uncertainty, when reliable models for predicting the world begin to fail, conspiracy theories offer beliefs that can reduce uncertainty across the belief system.

When the world starts becoming more unpredictable, we latch onto beliefs that help us reduce prediction error, even if those beliefs sacrifice accuracy, credibility, or wellness. There is a true imperative to make sense of the world, forcing us sometimes to go to extreme lengths in order to protect our sense of grip. In some cases, this leads to rigidness, in order to protect beliefs that may reduce uncertainty. In others, malleability gives the impression that there is very little prediction error. When adopted inflexibly, neither are truly able to reduce uncertainty across the entire system, and tend to lead to increased uncertainty. This increased uncertainty may actually make it more difficult to change belief systems once they have formed. Poth and Dolega, 2023; Sato, 2023

## Developing Practices for Better Belief Formation

The implications of the belief formation and updating traits shared by conspiracy theory belief and cognitive pathologies may unveil a number of things about what it means to have a healthy belief system. It appears that both overly rigid, and overly malleable belief systems are related to the formation of sub-optimal belief systems, yet having aspects of both rigidity and malleability also appears to be crucial to healthy belief systems.

This converges well with a predicted processing account of well-being outlined by Miller et al. (2022). They found that predictive agents must find a balance between exploiting their known environment, or exploring new, uncertain environments. There is an optimal amount of uncertainty in any model, a point at which there is room for learning and growing the model, but which is manageable and not overwhelming. Being at this point is conducive to the unexpected decrease of prediction error that ignites feelings of hedonic pleasure by creating an opportunity for the formation of new beliefs that help predict the world. Balancing around this point of optimal uncertainty across many domains is crucial to overall well-being. Exploring one domain only may result in hedonic pleasure as there is plenty of error reduction

in that one field, but ultimately cannot lead to overall well-being, as other domains are neglected. As well as finding the optimal point of uncertainty, it is also important to be able to shift from that point in any given domain. There are moments where being at the point is optimal, but in moments of high uncertainty, it is often adaptive to be able to shift back to depending on strong models with high precision, and in the face of problems that require novel solutions, it is important to be able to shift to a more explorative dynamic (Miller et al., 2022).

This model of well-being further confirms an independent hypothesis surrounding the importance of cognitive flexibility to well-being (Safron et al., 2022). High network flexibility is positively associated with rate of learning, working memory, relational reasoning and planning, and more. However overly flexible cognition is not always good, and can also be associated with reduced success of attention networks, and seems to sometimes hinder the rate of learning, especially in infants (Safron et al., 2022). Research in cognitive flexibility converges with Miller et al. (2022) insofar as it suggests that the most adaptive agents are able to find a balance between rigidity and malleability that enable growth, change, and adaptation without becoming overwhelming.

There are a number of suggestions for how individuals and communities are likely to find this sacred balance. Unsurprisingly, being socially engaged in a community appears to be critical to finding this balance (Miller et al., 2022; Garland et al., 2010). Being connected with others helps us engage with our beliefs in a critical manner by giving us insight into their belief systems - helping us extend our own models without having to actively engage in uncertainty (Clark, 1998). This converges with research suggesting that the onset of both cognitive pathologies and conspiracy theory belief are positively correlated with feelings of loneliness and isolation (Baddock et al., 2017; Douglas et al., 2017; Fabry, 2019). Our current practices of institutionalizing those with cognitive pathologies, and isolating those who subscribe to conspiracy theories may actually increase uncertainty and encourage the adoption of sub-optimal belief. One of the most simple, yet effective ways to combat sub-optimal belief systems is by radically including people. It is frustrating when someone persistently disregards counter-evidence disconfirming a belief. However, knowing that conspiracy theory belief is often supported by the same cognitive systems that underlie depression may encourage a shift towards compassion as opposed to frustration. It would be surprising and uncompassionate to expect someone with depression to completely shift their cognitive model after one conversation, no matter how much evidence they were shown disconfirming their negative beliefs. On a social and institutional level we have mostly stopped accusing depressed people of being crazy or stubborn for remaining depressed, instead moving towards a more compassionate approach. I suggest that we need to continue this approach in the face of cognitive

pathologies, and adopt a similar approach in the case of conspiracy theory belief. There are real reasons that it is hard to change beliefs, and perhaps we need to accept that conspiracy theory believers, like people with cognitive pathologies, may need many months or years, and support teams, and compassion in order to dispel sub-optimal beliefs.

Focusing on non-zero-sum games, as opposed to zero-sum games, is also associated with optimal belief systems (Miller et al., 2022). Research shows that people who focus their goals on non-zero-sum games, like self-improvement and altruism, have stronger social bonds and skills, and higher trust in their communities. People who focus on zero-sum games, like increased wealth, on the other hand are much less trusting of others, and less fulfilled generally (Miller et al., 2022). As well as including people, it is crucial that we shift social and institutional rhetoric to encourage non-zero-sum games. Shifting to a more collaborative, communitarian mindset in both personal and political contexts may be crucial to encouraging non-zero-sum mindsets. On a personal level things like volunteering, participating in non-competitive team sports or group activities, and regularly trying new things, especially things that we do not naturally excel at may encourage non-zero sum games. On a political level social programs like welfare, universal basic income, and encouraging political engagement are potential examples of ways to encourage non-zero sum games.

Luckily, just like sub-optimal belief systems can be learned, so can optimal ones. Research suggests that there are strategies that can help people learn to generate healthier belief systems, and learn to find balance between exploration and exploitation. Further, once learned, predictive processing encourages people to stay at this optimal point where they are likely to reduce uncertainty (Garland et al., 2010). Optimal belief systems encourage optimal belief systems. Mindfulness, mediation, and other forms of metacognitive training appear to be especially useful in helping people adopt optimal flexibility (Garland et al., 2010).

Mindfulness is the act of paying attention to current moment to moment states through self-regulation of attentional and metacognitive states (Garland et al., 2010). Mindfulness encourages a shift of perspective away from generative models, and can help challenge cognitive biases (Laukkonen & Slagter, 2021). Research shows that mindfulness leads to many positive outcomes, including increased attention, emotional regulation, better responses to stressful or negative events, and increased neuroplasticity (Garland et al., 2010). Learning, and becoming accustomed to challenging top-down beliefs encourages more adaptive belief patterns.

Other types of meditation and mindfulness that focus on different things, like loving-kindness meditation which focus on generating warm, caring feelings towards the self and others, have also shown to have positive impacts on our relationships (Garland et al., 2010). By verbalizing and visualizing positive things, like feelings of comfort, safety, and health, people may be actively incorporating these goals into their generative model, shifting the way that novel evidence is perceived. Further, like mindfulness meditation, loving-kindness meditation encourages a focus on others as well as the self. Again, positive relationships within communities, including the relationship with the self are crucial to optimal belief systems.

Radical inclusion and recentering of attention may be at the core of making us less prone to adopting sub-optimal belief systems, and dispelling sub-optimal belief systems that are in place. Societal and personal shifts in behavior, like focusing on non-zero-sum games, and encouraging meta-cogntiive training, including various types of meditation are some of the theoretical and practical steps that we can, and must, begin to take in order to avoid pathological and nonpathological belief systems that are harmful to individuals, and their communities.

## Conclusion

Conspiracy theory belief and cognitive pathologies are examples of optimal belief systems that have adopted sub-optimal beliefs that are harmful to individuals and their communities. It is critical that we address these suboptimal beliefs meaningfully. This investigation into the similarities between belief formation, maintenance, and updating in conspiracy theory belief and cognitive pathologies demonstrates the impact of misinformation on belief networks. The importance of finding a balance between rigidity and malleability can be demonstrated by the impacts of adopting belief systems that are overly rigid or overly malleable. While having both rigid and malleable beliefs are adaptive, as they allow us to maintain or change beliefs in the face of novel evidence, having a dynamic belief system that is able to flexibly shift between rigid and malleable seems to be key to maintaining healthy predictive models.

In light of the similarities between conspiracy theory belief and cognitive pathologies, we may need to reevaluate the ways that we treat both phenomena. While some people are more prone to bad-belief networks, it seems unlikely that anyone is completely immune. By practicing radical inclusion, focusing on non-zero-sum games, exercising mindfulness and loving-kindness meditation, and changing our social and political systems to encourage healthy belief networks, we may be able to combat both the mental health crisis, and the misinformation crisis that are plaguing us.

**References**


Adams, R. A., Stephans, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013, May 30). The Computational Anatomy of PSychosis. *Frontiers in Psychiatry*, *4*. https://doi.org/10.3389/fpsyt.2013.00047

Andersem, B. (2022, July 11). Autistic-like Traits and Positive Schizotypy as Diametric Specializations of the Predictive Mind. *Perspectives on Psychological Science*, *17*(6). https://journals.sagepub.com/doi/abs/10.1177/17456916221075252?journalCode=ppsa#:~:text=https%3A//doi.org/10.1177/17456916221075252

Baddock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017, March). The Depressed Brain: An Evolutionary Systems Theory. *Trends in Cognitive Science*, *21*(3), 182-194. Science Direct. https://doi.org/10.1016/j.tics.2017.01.005

Bortolotti, L., & Ichino, A. (2020, December 9). *Conspiracy theories may seem irrational – but they fulfill a basic human need*. The Conversation. Retrieved May 28, 2023, from https://theconversation.com/conspiracy-theories-may-seem-irrational-but-they-fulfill-a-basic-human-need-151324

Bortolotti, L., Ichino, A., & Mameli, M. (2021, December 22). Conspiracy theories and delusions. *Reti, saperi, linguaggi: Italian Journal of Cognitive Sciences*, *8*(2), 183-200. https://doi.org/10.12832/102760

Clark, A., & Chalmers, D. (1998). The Extended Mind. Analysis, 58(1), 7–19. http://www.jstor.org/stable/3328150

Clark, A. (2017, September 27). A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenology and the Cognitive Sciences*, *17*, 521–534. https://doi.org/10.1007/s11097-017-9525-z

Clark, J. E., Watson, S., & Friston, K. (2016, February 26). What is the mood? A computational perspective. *Psychological Medicine*, *48*(14). 10.1017/S0033291718000430

Dentith, M. R.X. (2016, May 02). When inferring a conspiracy might be the best explanation. *Social Epistemology*, *30*(5-6), 572-591. Taylor & Francis Online. https://doi.org/10.1080/02691728.2016.1172362

Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017, December). The Psychology of Conspiracy Theories. *Current Directions in Psychological Science*, *26*(6), 538-542. Sage Journals. https://doi.org/10.1177/0963721417718261

Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Siang Aang, C., & Deravi, F. (2019, March 20). Understanding Conspiracy Theories. *Political Psychology*, *40*(S1), 3-35. https://doi.org/10.1111/pops.12568

Fabry, R. E. (2019, August 30). Into the dark room: a predictive processing account of major depressive disorder. *Phenomenology and Cognitive Sciences*, *19*, 685–704. https://doi.org/10.1007/s11097-019-09635-4

Firth, C. D., & Friston, K. J. (2013, January 1). False perceptions & false beliefs: Understanding schizophrenia. *Neurosciences and the Human Person: New Perspectives on Human Activities*, *121*, 1-15. https://www.researchgate.net/publication/285696518_False_perceptions_false_beliefs_Understanding_schizophrenia

Fletcher, P. C., & Firth, C. D. (2008, December 3). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*, 48-58. https://doi.org/10.1038/nrn2536

Friston, K. (2013). Active Inference and Free Energy. *Behavioral and Brain Sciences*, *36*(2), 212-213.

Gallagher, S., Hutto, D., & Hipólito, I. (2021, November 24). Predictive Processing and Some Disillusions about Illusions. *Review of Philosophy and Psychology*, *13*, 999-1017. Springer. https://doi.org/10.1007/s13164-021-00588-9

Garland, E. L., Fredrickson, B., King, A. M., Johnson, D. P., Meyer, P. S., & Penn, D. L. (2010, November). Upward spirals of positive emotions counter downward spirals of negativity: Insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. *Clinical Psychology Review*, *30*(7), 849-864. https://doi.org/10.1016/j.cpr.2010.03.002

Goreis, A., & Voracek, M. (2019, February 11). A Systematic Review and Meta-Analysis of Psychological Research on Conspiracy Beliefs: Field Characteristics, Measurement Instruments, and Associations With Personality Traits. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00205

Hohwy, J. (2013). Delusions, illusions, and inference under uncertainty. *Mind & Language*, *28*, 57-71. 10.1111/mila.12008

Hohwy, J. (2017, January). Priors in perception: Top-down modulation, Bayesian perceptual learning rate, and prediction error minimization. *Conscious Cognition*, *47*, 75-85. 10.1016/j.concog.2016.09.004

Kiverstein, J., Miller, M., & Rietveld, E. (2019). The feeling of grip: novelty, error dynamics, and the predictive brain. *Synthese*, *196*, 2847–2869. https://doi.org/10.1007/s11229-017-1583-9

Kube, T., Schwarting, R., Rosenkrantz, L., Glombiewski, J. A., & Rief, W. (2020, March 1). Distorted Cognitive Processes in Major Depression: A Predictive Processing Perspective. *Biological Psychiatry*, *87*(5), 388-398. https://doi.org/10.1016/j.biopsych.2019.07.017

Metzinger, T., & Wiese, W. (2017, March). Vanilla PP for Philosophers: A Primer on Predictive Processing. *Philosophy and Predictive Processing*, *1*. https://predictive-mind.net/DOI?isbn=9783958573024

Miller, M., Kiverstein, J., & Rietveld, E. (2020, February). Embodying addiction: A predictive processing account. *Brain and Cognition*, *138*. https://doi.org/10.1016/j.bandc.2019.105495

Miller, M., Kiverstein, J., & Rietveld, E. (2022, January). The Predictive Dynamics of Happiness and Well-Being. *Emotion Review*, *14*(1), 15-30. https://doi.org/10.1177/17540739211063851

Miller, M., & White, B. (2021, May 25). *Social media and the neuroscience of predictive processing*. Aeon. Retrieved August 22, 2023, from https://aeon.co/essays/social-media-and-the-neuroscience-of-predictive-processing

Poth, N., & Dolega, K. (2023, January 31). Bayesian Belief Protection: A Study of Belief in Conspiracy Theories. *Philosophical Psychology*, *6*, 1182-1207. https://doi.org/10.1080/09515089.2023.2168881

Saffron, A., Klimaj, V., & Hipolito, I. (2022, January 21). On the Importance of Being Flexible: Dynamic Brain Networks and Their Potential Functional Significances. *Frontiers in Systems Neuroscience*, *5*. https://doi.org/10.3389/fnsys.2021.688424

Smith, R., Varshney, L. R., Nagayama, S., Kazama, M., Kitagawa, T., Managi, S., & Ishikawa, Y. (2022, October 31). A computational neuroscience perspective on subjective wellbeing within the active

inference framework. *International Journal of Wellbeing*, *12*(4), 102-131. https://doi.org/10.5502/ijw.v12i4.2659

Weatherall, J. O., & O'Connor, C. (2019). *The Misinformation Age: How False Beliefs Spread*. Yale University Press.

Williams, D. (2021, June). Socially Adaptive Belief. *Mind and Language*, *36*(3), 332-354. https://doi.org/10.1111/mila.12294open_in_new